European Community on Computational Methods in Applied Sciences

ECCOMAS Congress 2016

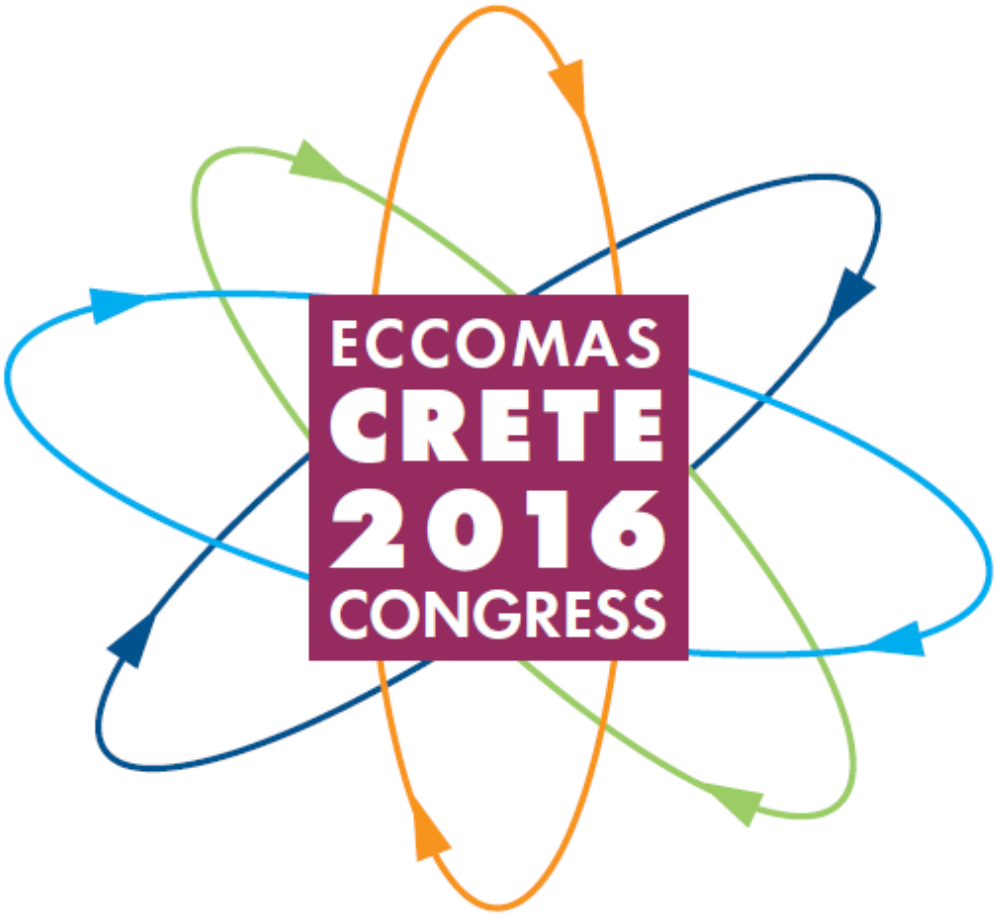
VII European Congress

on Computational Methods in Applied Sciences and Engineering

PROCEEDINGS

Volume I

M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (Eds)



ECCOMAS Congress 2016

Proceedings of the VII European Congress on Computational Methods
in Applied Sciences and Engineering
Held in Crete, Greece
5-10 June 2016

Edited by:

M. Papadrakakis
National Technical University of Athens, Greece

V. Papadopoulos
National Technical University of Athens, Greece

G. Stefanou
Aristotle University of Thessaloniki, Greece

V. Plevris
Oslo and Akershus University College of Applied Sciences, Oslo, Norway

A publication of:

Institute of Structural Analysis and Antiseismic Research
School of Civil Engineering
National Technical University of Athens (NTUA)
Greece

ECCOMAS Congress 2016
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (Eds)

First Edition, September 2016

© The authors

ISBN: **978-618-82844-0-1**

PREFACE

This volume contains the full-length papers presented at the **VII European Congress on Computational Methods in Applied Sciences and Engineering** (ECCOMAS Congress 2016) that was held on June 5-10, 2016 on the Crete Island, Greece.

The main objective of the quadrennial ECCOMAS Congresses is to provide a forum for presentation and discussion of state-of-the-art advances in computational methods in applied sciences and engineering, including basic methodologies, scientific developments and industrial applications and to serve as a platform for establishing links between research groups of academia and industry with common as well as complementary activities. About 2,200 papers were presented at the ECCOMAS Congress by authors from 53 countries around the world. This volume, consists of 667 full length accepted papers which will be indexed by SCOPUS database with access to the pdf file of the paper.

The ECCOMAS Congress 2016 is organized by the Institute of Structural Analysis and Antiseismic Research of the National Technical University of Athens under the support of the Greek Association for Computational Mechanics (GRACM), the Institute of Research and Development for Computational Methods in Engineering Sciences (ICMES) and the Computer Applications and Education in Engineering Sciences (CAEES).

The editor of this volume would like to thank all authors for their contributions. Special thanks go to the colleagues who contributed to the organization of the Minisymposia and to the reviewers who, with their work, contributed to the scientific quality of this e-book.

M. Papadrakakis
National Technical University of Athens, Greece

V. Papadopoulos
National Technical University of Athens, Greece

G. Stefanou
Aristotle University of Thessaloniki, Greece

V. Plevris
Oslo and Akershus University College of Applied Sciences, Oslo, Norway

ACKNOWLEDGEMENTS

The conference organizers acknowledge the support towards the organization of the 7th European Congress on Computational Methods in Applied Sciences and Engineering, to the following organizations:

- Greek Association for Computational Mechanics (GRACM)
- Institute of Structural Analysis and Antiseismic Research, National Technical University of Athens
- John Argyris Foundation
- Institute of Research and Development for Computational Methods in Engineering Sciences (ICMES)
- Computer Applications and Education in Engineering Sciences (CAEES)
- Region of Crete
- Municipality of Hersonissos

Plenary Speakers

We would also like to thank the Plenary and Semi-Plenary Speakers for their help in the setting up of a high standard Scientific Programme.

Plenary Speakers: Alexandre Ern, Charbel Farhat, Gerhard A. Holzapfel, Antonio Huerta, Thomas J.R. Hughes, Alfio Quarteroni, Ole Sigmund, Theodosios P. Tasios

Semi-Plenary Speakers: Assyr Abdulle, Olivier Allix, Lourenco Beirao da Veiga, Andreas G. Boudouvis, Annalisa Buffa, Ramon Codina, Laura De Lorenzis, Bernard Geurts, Kyriakos Giannakoglou, Antonio J. Gil, Dan Givoli, Anthony Gravouil, George Karniadakis, Mats G. Larson, Kim Meow Liew, Wing Kam Liu, Federico Negri, Shinji Nishiwaki, Ursula Rasthofer, Ernst Rank, Cord-Christian Rossow, Giancarlo Sangalli, Jörg Schröder, Bert Sluys, Christian Soize, Roberto Verzicco, Barbara Wohlmuth

SUMMARY

Preface..... iii

Acknowledgements..... iv

Contents..... xvii

VOLUME I

SEMI-PLENARY LECTURES 1

MINISYMPOSIA

MS 102 - COMPUTATIONAL MODELS IN BIOMECHANICS AND MECHANOBIOLOGY 27
Organized by Estefania Peña, Renato Natal Jorge, Miguel A. Martínez, Pedro S. Martins

MS 103 - MECHANICS OF BIOLOGICAL TISSUES 55
Organized by Markus Böl, Gerhard A. Holzapfel

MS 105 - SIMULATION OF CARDIOVASCULAR PROCEDURES AND DEVICES 81
Organized by Ferdinando Auricchio, Michele Conti, Simone Morganti, Alessandro Reali, Alessandro Veneziani

MS 106 - DIRECT AND INVERSE METHODS FOR CARDIOVASCULAR AND PULMONARY BIOMECHANICS 98
Organized by Wolfgang A. Wall, C. Alberto Figueroa, Marek Behr

MS 108 - NUMERICAL METHODS FOR COUPLED PROBLEMS IN BIOMEDICAL APPLICATIONS 105
Organized by Martina Bukac, Annalisa Quaini

MS 111 - POPULATION BALANCE MODELING: CURRENT STATUS, FUTURE PROSPECTS AND NOVEL APPLICATIONS FROM NANOPARTICLES' SYNTHESIS TO (LUNG) CANCER 120
Organized by Georgios Lolas, Georgios Bourantas, Panagiotis Gavriliadis, Konstantinos Syrigos

MS 112 - ANEURYSMS: SOLID MECHANICS, FLUID MECHANICS, AND MECHANOBIOLOGY 137
Organized by Christian J. Cyron, Sven Hirsch, Philippe Bijlenga, Roland C. Aydin, Anne M. Robertson, Gerhard A. Holzapfel

MS 113 - MATHEMATICAL AND NUMERICAL MODELING OF THE HEART 155
Organized by Luca Dede', Luca Pavarino, Alfio Quarteroni

MS 115 - TUMOR GROWTH MODELING AND THE MECHANICAL ASPECTS OF CANCER 178
Organized by Hector Gomez, Assad Oberai, Krishna Garikipati, Kristen Mills, Thomas J.R. Hughes

MS 201 - MICROSTRUCTURE-DRIVEN DEFORMATION AND FAILURE IN CRYSTALLINE MATERIALS 188
Organized by Pilar Ariza, Lucia Nicola, Angelo Simone

MS 202 - CIVIL ENGINEERING MATERIALS AND STRUCTURES UNDER EXTREME LOADINGS 216
Organized by Fabrice Gatuingt, Frédéric Dufour, Panagiotis Kotronis

MS 203 - COMPUTATIONAL METHODS FOR MODELLING INSTABILITIES IN SOLIDS & STRUCTURES 254
Organized by Spyros A. Karamanos

MS 204 - IMPACT AND CRASH MECHANICS 273
Organized by Manfred Bischoff, Fabian Duddeck

MS 205 - PROBABILISTIC APPROACH TO NUMERICAL SIMULATION OF FRACTURE 323
Organized by Alexander V. Gerasimov

MS 301 - METHODS FOR CUT AND COMPOSITE MESHES: THEORY, ALGORITHMS AND APPLICATIONS 373
Organized by Mats G. Larson, André Massing

MS 302 - MESH GENERATION AND ADAPTION 384
Organized by Josep Sarrate, Xevi Roca, Rafael Montenegro, Eloi Ruiz

MS 303 - CURVED MESH GENERATION FOR HIGH-ORDER METHODS 417
Organized by Xevi Roca, Josep Sarrate

MS 304 - COMPUTATIONAL MODELLING OF ADDITIVE PRODUCTION PROCESSES 469
Organized by Dirk Hartmann, Stefan Kollmannsberger, Ernst Rank, Utz Wever

MS 305 - ADVANCED MESHING METHODS FOR INDUSTRIAL APPLICATIONS 480
Organized by Frederic Alauzet, Thierry Coupez, Alain Dervieux, Adrien Loseille

MS 306 - LATTICE SPRING METHODS FOR LINEAR AND NONLINEAR CONTINUA 511
Organized by Ioannis Doltsinis

MS 403 - PARTICLE-BASED METHODS IN FLUID MECHANICS 539
Organized by Sergio Idelsohn, Eugenio Oñate

MS 404 - SIMULATION OF ENVIRONMENTAL FLOWS 650
Organized by Pablo Ortiz, Piotr K. Smolarkiewicz, Joanna Szmelter

MS 405 - COMPUTATIONAL MODELING OF MULTIPHASE FLOWS: ADVANCED METHODS, INTERFACE PHENOMENA AND ENVIRONMENTAL APPLICATIONS 750
Organized by Adeline Montlaur, Santiago Arias Calderón, Martin Kronbichler

MS 406 - ADVANCES IN COMPUTATIONAL METHODS FOR GAS-LIQUID TWO-PHASE FLOW 788
Organized by Byeong Rog Shin, Takeo Kajishima

MS 408 - MANIPULATION AND CONTROL OF TURBULENT FLOW 831
Organized by Markus Rütten, Christina Voß

MS 409 - CURRENT TRENDS IN MODELLING AND SIMULATION OF TURBULENT FLOWS 865
Organized by Suad Jakirlić

MS 410 - COMPLEX FLUID FLOWS IN ENGINEERING: MODELLING, SIMULATION AND OPTIMIZATION 949
Organized by Stefanie Elgeti, Philipp Knechtges

MS 411 - NON-NEWTONIAN HEAT AND FLUID FLOW SUBJECTED TO MAGNETIC FORCES 1061
Organized by Laszlo Konozsy, Dimitris Drikakis

MS 412 - NUMERICAL METHODS FOR WAVES AND FLOWS IN COASTAL AND DEEP WATER HYDRODYNAMICS 1093
Organized by Nina Shokina, Yuri Shokin, Leonid Chubarov, Gayaz Khakimzyanov, Vadym Aizinger, Denys Dutykh

MS 414 - NEW TRENDS IN NUMERICAL METHODS FOR MULTI-MATERIAL COMPRESSIBLE FLUID FLOWS 1139
Organized by Andy Barlow, Michael Dumbser, Raphaël Loubère, Pierre-Henri Maire, Rob Rieben, Mikhail Shashkov, François Vilar

MS 415 - COMPUTATIONAL NON-NEWTONIAN FLUID MECHANICS 1205
Organized by Georgios Georgiou, John Tsamopoulos

MS 501 - ALGORITHMIC ASPECTS OF HIGH-PERFORMANCE COMPUTING FOR MECHANICS AND PHYSICS 1237
Organized by Santiago Badia, Victor Calo, Javier Principe

MS 503 - HPC-BASED SIMULATIONS FOR THE ENGINEERING REALM AND INDUSTRIAL APPLICATIONS 1277
Organized by Makoto Tsubokura, Mariano Vázquez, Takayuki Aoki

MS 504 - NUMERICAL METHODS AND TOOLS FOR KEY EXASCALE COMPUTING CHALLENGES IN ENGINEERING AND APPLIED SCIENCES 1352
Organized by Eugenio Oñate, Manolis Papadrakakis, Peter Wriggers

MS 505 - INTERACTIVE SIMULATIONS IN COMPUTATIONAL ENGINEERING 1392
Organized by Adrian Harwood, Petra Wenisch

MS 506 - ACCURACY AND EFFICIENCY OF APPROXIMATE COMPUTATIONS IN SCIENCE AND ENGINEERING 1409
Organized by Aram Soroushian

MS 601 - SHOCK WAVE-BOUNDARY LAYER INTERACTION AND ITS CONTROL 1455
Organized by Piotr Doerffer, George Barakos

MS 602 - INNOVATIVE METHODS FOR FLUID-STRUCTURE-INTERACTION 1498
Organized by E. Harald van Brummelen, Roger Ohayon, Trond Kvamsdal

MS 603 - COMPUTATIONAL METHODS IN FLUID-STRUCTURE INTERACTION WITH IMPACT ON INDUSTRIAL APPLICATIONS 1582
Organized by Elisabeth Longatte, Yannick Hoarau, Marianna Braza

MS 605 - FRICTIONAL CONTACTS WITH LUBRICATION – BASICS AND APPLICATIONS 1623
Organized by Michael Müller, Thomas Hagemann

MS 606 - COMPUTATIONAL MODELING OF HYDRAULIC FRACTURING 1637
Organized by Gianluca Cusatis, Gilles Pijaudier-Cabot, Günther Meschke

MS 607 - ADVANCES IN COMPUTATIONAL METHODS FOR LIQUID-VAPOR FLOWS WITH PHASE TRANSFER PROCESSES 1672
Organized by Rémi Abgrall, Pietro M. Congedo, Tore Flåtten, Bernhard Müller, Marica Pelanti, Maria Giovanna Rodio

MS 608 - ADVANCES IN TIME INTEGRATION FOR SOLID, FLUID AND COUPLED SYSTEMS 1709
Organized by Ilinca Stanciulescu, Peter Betsch

MS 609 - ADVANCED COMPUTATIONAL MODELING OF BATTERIES AND FUEL CELLS 1838
Organized by Edwin Knobbe, Wolfgang A. Wall

MS 610 - NUMERICAL METHODS TO STUDY THE CONTACT MECHANICS OF DRY, ADHESIVE AND LUBRICATED ROUGH SURFACES 1850
Organized by Carmine Putignano, Daniele Dini

MS 612 - NUMERICAL SIMULATIONS FOR SMART-CITY APPLICATIONS 1876
Organized by Julien Waeytens, Rachida Chakir

MS 615 - COMPUTATIONAL MODELS IN MAGNETOHYDRODYNAMICS 1886
Organized by Oleg Zikanov

MS 701 - ADVANCED MATERIALS: COMPUTATIONAL ANALYSIS OF PROPERTIES AND PERFORMANCE 1893
Organized by Vadim V. Silberschmidt, Valery P. Matveenko

MS 702 - MODELING OF NANOFILLED COMPOSITES 1971
Organized by Konstantinos I. Tserpes

MS 705 - IDENTIFICATION OF MATERIAL MODELS 2003
Organized by Danuta Szeliga, Wacław Kuś, Tadeusz Burczyński, Jan Kusiak

MS 706 - MODELING OF FIBER-BASED STRUCTURES - TEXTILES AND TEXTILE REINFORCED COMPOSITES 2010
Organized by Yordan Kyosev, Philippe Boisse, Nahiene Hamila, Damien Durville

MS 707 - MICROMECHANICAL MODELLING: COMPETITION BETWEEN ANALYTICAL AND NUMERICAL METHODS 2039
Organized by Siegfried Schmauder, Vera Petrova

MS 711 - FOURIER-BASED METHODS FOR COMPUTING THE BEHAVIOR OF HETEROGENEOUS MATERIALS DEVELOPMENTS, EXTENSIONS AND APPLICATIONS 2099
Organized by Lionel Gélébart, Hervé Moulinec, Franz Roters, François Willot

MS 712 - SMART MATERIAL SYSTEMS AND STRUCTURES 2110
Organized by Mieczysław Kuczma, Pavel Krejči, Jörg Schröder, Georgios E. Stavroulakis, Gwidon Szefer

MS 713 - MICROSTRUCTURE-BASED MODELLING OF HETEROGENEOUS MATERIALS 2134
Organized by Jan Zeman, Jan Novak, Guillermo Díaz

MS 714 - STRENGTH, FATIGUE AND STABILITY OF COMPOSITE STRUCTURES 2182
Organized by Raimund Rolfes, Martin Ruess, Kai-Uwe Schröder

VOLUME II

MINISYMPOSIA

MS 715 - COMPUTATIONAL ANALYSIS OF COMPOSITE STRUCTURES 2210
Organized by Efstathios E. Theotokoglou

MS 801 - MULTISCALE COMPUTATIONAL HOMOGENIZATION FOR BRIDGING SCALES IN THE MECHANICS AND PHYSICS OF COMPLEX MATERIALS 2266
Organized by Julien Yvonnet, Kenjiro Terada, Peter Wriggers, Marc Geers

MS 802 - ADVANCES IN THE MODELLING OF MULTI-SCALE, MULTI-PHYSICS AND MULTI-UNCERTAINTY PROBLEMS 2323
Organized by Francisco M. Andrade Pires, Chengfeng Li

MS 805 - ADVANCED MULTI-PHYSICS AND MULTI-SCALE TECHNIQUES FOR MODELING INELASTIC PROCESSES IN SOLIDS: DAMAGE, FRACTURE AND CONTACT MECHANICS 2353
Organized by Mauro Corrado, Marco Paggi, José Reinoso

MS 806 - MULTISCALE MODELLING OF MATERIALS AND STRUCTURES 2363
Organized by Tadeusz Burczyński, Xavier Oliver, Maciej Pietrzyk, Alfredo Huespe

MS 807 - ADVANCED COMPUTATIONAL STRATEGIES FOR MODELLING, SIMULATION AND CHARACTERISATION OF MULTI-SCALE HETEROGENEOUS MATERIALS 2449
Organized by Stéphane Bordas, Daniel Dias-da-Costa, Fabrice Pierron, Timon Rabczuk, Pierre Kerfriden, Pascal Lava

MS 808 - MULTISCALE AND MULTIPHYSICS MODELING OF CEMENTITIOUS MATERIALS 2460
Organized by Jörg F. Unger, Thomas Titscher

MS 809 - MULTISCALE STOCHASTIC FINITE ELEMENT METHODS 2471
Organized by George Stefanou, Xi Frank Xu, Yu Ching Wu

MS 811 - MULTISCALE MODELING OF CONCRETE AND CONCRETE STRUCTURES 2501
Organized by Herbert Mang, Yong Yuan

MS 901 - ISOGEOMETRIC METHODS 2514
Organized by Yuri Bazilevs, David J. Benson, Rene De Borst, Thomas J.R. Hughes, Trond Kvamsdal, Alessandro Reali, Giancarlo Sangalli, Clemens V. Verhoosel

MS 902 - INNOVATIVE NUMERICAL APPROACHES FOR MULTI-PHYSICS PROBLEMS 2558
Organized by Anna Pandolfi, Laurent Stainier, Kerstin Weinberg

MS 903 - ADVANCES IN FICTITIOUS DOMAIN METHODS FOR SOLID MECHANICS 2566
Organized by Alexander Düster, Ernst Rank, Stefan Kollmannsberger, Andreas Schröder

MS 905 - DISCONTINUOUS GALERKIN METHODS: NEW TRENDS AND APPLICATIONS 2614
Organized by Sonia Fernández-Méndez, Nicoletta Franchina

MS 906 - MATHEMATICAL ADVANCES IN ISOGEOMETRIC ANALYSIS 2715
Organized by Annalisa Buffa, John A. Evans, Thomas J.R. Hughes, Giancarlo Sangalli

MS 907 - REGULARIZED ENRICHED APPROXIMATIONS AND QUADRATURE FOR DISCONTINUITIES, SINGULARITIES AND CONTINUOUS-DISCONTINUOUS TRANSITION 2726
Organized by Elena Benvenuti, Giulio Ventura, José M.A. César de Sá

MS 908 - VERIFICATION AND VALIDATION OF STRUCTURAL MECHANICS SIMULATION MODELS 2734
Organized by George Lampeas

MS 910 - HIGH ORDER CFD METHODS: CONCLUSIONS AND OUTLOOK 2843
Organized by Koen Hillewaert, John Ekaterinaris, Peter Vincent, Norbert Kroll, Norbert Huynh, Z.J. Wang

MS 911 - NUMERICAL METHODS IN THE MECHANICS OF GENERALIZED CONTINUA 2869
Organized by Elena Atroshchenko, Jack S. Hale, George Bourantas, Stéphane P.A Bordas

MS 912 - HIGH-ORDER METHODS, SENSITIVITY ANALYSIS AND ADAPTATION FOR THE NAVIER STOKES EQUATIONS 2883
Organized by Vincent Couaillier, Rémi Abgrall, Eusebio Valero

MS 913 - HIGH-ORDER METHODS FOR POLYGONAL AND POLYHEDRAL MESHES 2930
Organized by Lourenço Beirao da Veiga, Franco Brezzi, Donatella Marini, Alessandro Russo

MS 916 - DIRECT METHODS FOR LIMIT STATES OF MATERIALS AND STRUCTURES 2971
Organized by Konstantinos V. Spiliopoulos, Dieter Weichert

MS 917 - MESOSCOPIC METHODS FOR COMPLEX FLUIDS AND SOFT MATTER 3026
Organized by Zhen Li, Wenxiao Pan, Igor V. Pivkin

MS 918 - COMPUTER ALGEBRA SYSTEMS IN MODELLING STATIC AND DYNAMIC PROBLEMS IN MECHANICS OF SOLIDS 3045
Organized by Alexander V. Matrosov, Dmitriy P. Goloskokov

MS 919 - RECENT ADVANCES IN NUMERICAL SIMULATION AND ANALYSIS OF KINETIC MODELS 3113
Organized by E. Harald van Brummelen, Manuel Torrilhon

MS 921 - RECENT ADVANCES IN BOUNDARY ELEMENT METHODS 3139
Organized by Gernot Beer, Luiz Wrobel, Martin Schan

MS 922 - HIGH-ORDER METHODS FOR ELASTIC WAVES AND THEIR APPLICATION 3211
Organized by Thomas Hagstrom, Daniel Appelo

MS 923 - NOVEL DISCRETIZATION METHODS – MATHEMATICAL AND MECHANICAL ASPECTS 3224
Organized by Jörg Schröder, Peter Wriggers, Ferdinando Auricchio, Carsten Carstensen

MS 1001 - STRUCTURAL AND MULTIDISCIPLINARY OPTIMIZATION 3239
Organized by J.F. Aguilar Madeira, Helder C. Rodrigues

MS 1002 - EVOLUTIONARY ALGORITHMS AND METAHEURISTICS IN CIVIL ENGINEERING AND CONSTRUCTION MANAGEMENT 3473
Organized by Jorge Magalhaes-Mendes, David Greiner

MS 1003 - ADVANCES IN DESIGN OPTIMIZATION OF STRUCTURES AND MATERIALS 3534
Organized by Zhen Luo, Zhan Kang

MS 1004 - AERODYNAMIC STRATEGIES FOR THE GLOBAL OPTIMIZATION OF FLYING CONFIGURATIONS IN SUPERSONIC FLOW 3596
Organized by Adriana Nastase, Catalin Nae

MS 1005 - MONITORING AND CONTROL OF STRUCTURES 3630
Organized by Resat Oyguc

MS 1006 - PARAMETER IDENTIFICATION IN SOLID MECHANICS 3652
Organized by A. Gil Andrade-Campos, Marco Rossi, Sandrine Thuillier, Franck Toussaint, Marta C. Oliveira

MS 1007 - ADDITIVE MANUFACTURING AND OPTIMIZATION 3689
Organized by Ekkehard Ramm, Ole Sigmund, Pierre Duysinx, Wing Kam Liu

MS 1008 - ULTRASONIC GUIDED WAVES TESTING AND MONITORING 3753
Organized by Yaacoubi Slah, Nico Declercq

MS 1009 - ADJOINT METHODS FOR STEADY & UNSTEADY OPTIMIZATION 3770
Organized by Kyriakos C. Giannakoglou, Jens Dominik Mueller

MS 1010 - INVERSE PROBLEMS, DESIGN AND OPTIMIZATION 4061
Organized by Marcelo J. Colaço, Helcio R. B. Orlande, George S. Dulikravich

MS 1011 - SURROGATE-ASSISTED EVOLUTIONARY ALGORITHMS IN AERODYNAMIC DESIGN/OPTIMIZATION 4082
Organized by Varvara Asouti, Esther Andrés, Emiliano Iuliano

MS 1013 - SOLUTION OF LARGE-SCALE INVERSE PROBLEMS 4164
Organized by Clint Dawson, Steve Mattis, Troy Butler, Lindley Graham

MS 1014 - DESIGN OPTIMIZATION AND INVERSE PROBLEMS FOR WAVE PROPAGATION PROBLEMS 4178
Organized by Martin Berggren

MS 1101 - REDUCED BASIS, POD AND PGD MODEL ORDER REDUCTION TECHNIQUES 4187
Organized by

MS 1103 - MATHEMATICAL SURROGATE MODELLING IN ELECTROMAGNETICS 4284
Organized by Francisco Chinesta, Elias Cueto, Antonio Huerta, Pierre Ladeveze, Gianluigi Rozza

MS 1201 - COMPUTATIONAL STRUCTURAL DYNAMICS 4351
Organized by
Evangelos J. Sapountzakis, Andreas E. Kampitsis

VOLUME III

MINISYMPOSIA

MS 1202 - ADVANCED BEAM MODELS 4445
Organized by Dinar Camotim, Zuzana Dimitrovová, Rodrigo Gonçalves

MS 1203 - THE MODELS AND INVESTIGATIONS METHODS OF DYNAMICS OF THE SOLIDS SYSTEMS WITH DRY FRICTION 4553
Organized by Alexey A. Kireenkov, Alexander V. Karapetyan

MS 1204 - NONLINEAR DYNAMICS OF ROTATING STRUCTURES 4611
Organized by Evangeline Capiez-Lernout, Marc P. Mignolet, Christian Soize

MS 1206 - ADVANCES IN NUMERICAL METHODS FOR LINEAR AND NON-LINEAR DYNAMICS AND WAVE PROPAGATION 4674
Organized by Alexander Idesman

MS 1207 - COMPUTATIONAL SIMULATION OF SMART STRUCTURES AND MATERIALS 4728
Organized by Ruediger Schmidt, Kai-Uwe Schröder

MS 1208 - BIFURCATIONS AND STABILITY 4741
Organized by Pekka Neittaanmäki, Nikolay Banichuk, Juha Jeronen, Tero Tuovinen

MS 1209 - DYNAMIC ANALYSIS OF BEAMS UNDER MOVING VEHICLES: APPLICATION TO RAILWAY TRACK MODELLING, DESIGN AND REHABILITATION 4780
Organized by Fernardo Simões, Antonio Pinto da Costa

MS 1210 - ADVANCES IN MODELING AND ANALYSIS OF FGM STRUCTURES 4810
Organized by Justin Murin, Stephan Kugler, Mehdi Aminbaghai

MS 1211 - COMPUTATIONAL STRATEGIES FOR STRUCTURAL ROBUSTNESS ASSESSMENT 4900
Organized by Domenic Asprone, Fulvio Parisi

MS 1212 - DYNAMICS AND SEISMIC RESPONSE OF ROCKING AND SELF-CENTERING STRUCTURES..... 4959
Organized by Matthew DeJong, Elias Dimitrakopoulos, Michalis Fragiadakis

MS 1213 - INNOVATIVE STRUCTURAL SYSTEMS FOR SEISMIC RESISTANT BUILDINGS 5142
Organized by Carlo Castiglioni

MS 1214 - HISTORIC MASONRY STRUCTURES: MODELLING, ASSESSMENT & RETROFIT 5172
Organized by Panagiotis Asteris, Charilaos Maniatakis, Constantine Spyrakos

MS 1215 - NONLINEAR VIBRATIONS OF CONSERVATIVE AND NONCONSERVATIVE SYSTEMS: PHENOMENA AND ADVANCED NUMERICAL METHODS 5396
Organized by Malte Krack, Ludovic Renson, Gaëtan Kerschen

MS 1216 - STRUCTURAL ANALYSIS AND VIBRATIONS 5444
Organized by Diana V. Bambill, Carlos A. Rossit

MS 1217 - COMPUTATIONAL METHODS IN EARTHQUAKE ENGINEERING AND STRUCTURAL DYNAMICS 5464
Organized by Vagelis Plevris, Georgia Kremmyda, Yasin Fahjan

MS 1218 - STABILITY AND CONTROL OF FLEXIBLE STRUCTURES 5646
Organized by Ilaria Venanzi, Marco Lepidi

MS 1220 - DYNAMICS OF STRUCTURES SUBJECT TO SEISMIC EXCITATIONS 5712
Organized by Michel Géradin, Evtim Zahariev, Evangelos J. Sapountzakis

MS 1221 - COMPUTATIONAL STRUCTURAL STABILITY 5768
Organized by Herbert A. Mang, Yeon-Bin Yang

MS 1222 - INFLUENCE OF LIQUEFIABLE SOIL ON SINGLE AND CLOSELY CLUSTERED STRUCTURES 5817
Organized by Nawawi Chouw, Rolly Orense, Tam Larkin

MS 1224 - INNOVATIVE SOLUTIONS FOR THE SEISMIC PROTECTION OF INDUSTRIAL BUILDINGS 5834
Organized by Walter Salvatore, Carlo Castiglioni, Francesco Morelli, Nikolaos Bakas

MS 1225 - SEISMIC PERFORMANCE ASSESSMENT OF STRUCTURES AND SEISMIC RISK MITIGATION STRATEGIES 5934
Organized by Marco Vona

MS 1301 - THE STOCHASTIC COMPUTER METHODS IN MECHANICS 6029
Organized by Marcin Kamiński, Takahiko Kurahashi

MS 1303 - ANALYSIS AND DESIGN OF SAFETY CRITICAL SYSTEMS UNDER UNCERTAINTY 6091
Organized by Edoardo Patelli, Michael Beer, Matteo Broggi, Francisco Alejandro Díaz De la O

MS 1304 - STOCHASTIC MODELING AND IDENTIFICATION OF UNCERTAINTIES IN COMPUTATIONAL MECHANICS 6136
Organized by Johann Guilleminot, Maarten Arnst, Christian Soize

MS 1306 - ERCOFTAC SIG-45: UNCERTAINTY QUANTIFICATION IN CFD AND FLUID STRUCTURE INTERACTION 6196
Organized by Didier Lucor, Sunetra Sarkar

MS 1307 - NON-INTRUSIVE SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION IN HIGH DIMENSIONS 6250
Organized by Bruno Sudret, Eleni Chatzi, Jean-Marc Bourinet, Nicolas Gayton

MS 1308 - MODELLING AND INVERSE METHODS IN NONLINEAR DYNAMICAL SYSTEMS 6316
Organized by Sotirios Natsiavas, Costas Papadimitriou, Eleni Chatzi, Dimitrios Giagopoulos

MS 1309 - SCALABLE MULTI-FIDELITY MODELING FOR DESIGN, UNCERTAINTY QUANTIFICATION, AND INVERSE PROBLEMS 6383
Organized by Paris Perdikaris, George Em. Karniadakis

MS 1310 - COMPUTATIONAL METHODS FOR THE SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS 6398
Organized by Jianbing Chen, Ioannis Kougioumtzoglou, Vissarion Papadopoulos

MS 1401 - TOICA: THERMAL OVERALL INTEGRATED CONCEPT AIRCRAFT 6420
Organized by Pierre Arbez, Jean-Claude Dunyach

STS 3 - INNOVATIVE DESIGN OPTIMIZATION TOOLS LINKED TO INDUSTRIAL AERONAUTICAL APPLICATIONS: TARGETING GREENER PERFORMANCES 6471
Organized by Jacques Periaux, Gabriel Bugeda

STS 5 - TRANSITION LOCATION EFFECT ON SHOCK WAVE BOUNDARY LAYER INTERACTION 6485
Organized by Piotr Doerffer, Pawel Flaszynski

STS 7 - MORPHING TECHNOLOGIES FOR AIRCRAFT WINGS 6496
Organized by Hans Peter Monner

STS 9 - ADVANCED WING HIGH-LIFT SYSTEMS 6505
Organized by Jochen Wild

YOUNG INVESTIGATORS MINISYMPOSIUM 6515
Organized by Jaan-Willem Simon, Alexander Popp, Joan Baiges

ECCOMAS OLYMPIAD MINISYMPOSIUM 6526

VOLUME IV

CONTRIBUTED SESSIONS

CS 110 - NUMERICAL MODELS IN BIOMECHANICS 6539

CS 210 - NUMERICAL SIMULATION OF COMPOSITE MATERIALS 6637

CS 211 - CRACK PROPAGATION 6688

CS 212 - NUMERICAL MODELING OF DAMAGE, FAILURE AND FRACTURE 6699

CS 230 - DYNAMIC FAILURE AND FRACTURE 6799

CS 310 - CAD, CAM AND CAE 6847

CS 320 - GRID GENERATION AND ADAPTIVE TECHNIQUES 6911

CS 410 - COMPUTATIONAL FLUID MECHANICS 6955

CS 420 - MULTI-PHASE AND CHEMICALLY REACTING FLOWS 7138

CS 450 - NUMERICAL METHODS AND CONVERGENCE ACCELERATION IN CFD 7274

CS 460 - UNSTEADY FLOW COMPUTATION 7390

CS 500 - HIGH PERFORMANCE COMPUTING 7572

CS 610 - AERO-ACOUSTICS 7589

CS 620 - COMPUTATIONAL CONTACT MECHANICS 7628

CS 630 - SIMULATION OF FLUID-STRUCTURE INTERACTION 7640

CS 720 - COMPUTATIONAL MATERIALS SCIENCE 7711

CS 750 - COMPUTATIONAL MODELING OF COMPOSITES 7765

CS 751 - SMART MATERIALS AND STRUCTURES 7853

CS 830 - COMPUTATIONAL NANOTECHNOLOGY 7875

CS 840 - MULTI-SCALE COMPUTATIONAL METHODS FOR SOLIDS AND FLUIDS 7886

CS 930 - HIGH-ORDER DISCRETIZATION METHODS 7956

CS 940 - EXTENDED DISCRETIZATION METHODS 8066

CS 960 - MESHLESS METHODS 8115

CS 980 - NUMERICAL AND SYMBOLIC COMPUTATION 8162

CS 990 - PARTICLE-BASED METHODS 8196

CS 1010 - COMPUTATIONAL INVERSE PROBLEMS AND OPTIMIZATION 8247

CS 1020 - EVOLUTIONARY AND DETERMINISTIC METHODS FOR DESIGN, OPTIMIZATION AND CONTROL 8391

CS 1100 - REDUCTION ORDER METHODS 8539

CS 1200 - STRUCTURAL DYNAMICS 8597

CS 1201 - COMPUTATIONAL SOIL MECHANICS 8708

CS 1202 - STRUCTURAL ANALYSIS AND MULTIBODY DYNAMICS 8739

CS 1300 - UNCERTAINTY QUANTIFICATION AND ERROR ESTIMATION 8769

CONTENTS

Semi-plenary Lectures

NONPARAMETRIC PROBABILISTIC APPROACH OF MODEL UNCERTAINTIES INTRODUCED BY A PROJECTION-BASED NONLINEAR REDUCED-ORDER MODEL 1
Christian Soize, Charbel Farhat

Minisymposia

100 BIOLOGICAL SYSTEMS

MS 102: COMPUTATIONAL MODELS IN BIOMECHANICS AND MECHANOBIOLOGY

THE INFLUENCE OF MECHANICAL STIMULUS ON NUTRIENT TRANSPORT AND CELL GROWTH IN ENGINEERED CARTILAGE: A FINITE ELEMENT APPROACH 27
S. Cortez, A. Completo, J.L. Alves

HYBRID CELL CENTRED/VERTEX MODEL FOR LARGE TISSUE DEFORMATIONS 37
José J. Muñoz, Payman Mosaffa, Yanlan Mao, Rob Tetley, Nina Asadipour, Antonio Rodríguez-Ferran

SENSITIVITY OF TEMPERATURE FIELD IN THE SYSTEM PROTECTIVE CLOTHING - FOREARM WITH RESPECT TO PERTURBATIONS OF EXTERNAL HEATING CONDITIONS 46
Bohdan Mochnacki, Mariusz Ciesielski

MS 103: MECHANICS OF BIOLOGICAL TISSUES

INFLUENCE OF MUSCLE COMPRESSION ON DYNAMIC MUSCLE PERFORMANCE 55
Tobias Siebert, Norman Stutzig, Olaf Till, Christian Rode

A 3D ELECTROMECHANICAL FEM-BASED MODEL FOR CARDIAC TISSUE 63
Minh Tuan Duong, Alexander Jung, Ralf Frotscher, Manfred Staat

BIOMECHANICAL PROPERTIES OF CORNEAL EXTRACELLULAR MATRIX USING EXPERIMENTAL AND COMPUTATIONAL METHODS 74
Hamed Hatami-Marbini

MS 105: SIMULATION OF CARDIOVASCULAR PROCEDURES AND DEVICES

MODELLING OF PATIENT-SPECIFIC CASES OF ATHEROSCLEROSIS IN CAROTID ARTERIES 81
Timur Gamilov, Roman Pryamonosov, Sergey Simakov

PERSONALIZED COMPUTATION OF FRACTIONAL FLOW RESERVE IN CASE OF TWO CONSECUTIVE STENOSES ... 90
Yuri Vassilevski, Timur Gamilov, Philip Kopylov

MS 106: DIRECT AND INVERSE METHODS FOR CARDIOVASCULAR AND PULMONARY BIOMECHANICS

BLOOD FLOW IN THE COMMON CAROTID ARTERY WITH STENOSIS 98
Helena Henriques, Luisa Sousa, Catarina Castro, Carlos António, Rosa Santos, Pedro Castro, Elsa Azevedo

MS 108: NUMERICAL METHODS FOR COUPLED PROBLEMS IN BIOMEDICAL APPLICATIONS
A FENICS-HPC FRAMEWORK FOR MULTI-COMPARTMENT BLOCH-TORREY MODELS 105
Dang Van Nguyen, Johan Jansson, Johan Hoffman

MS 111: POPULATION BALANCE MODELING: CURRENT STATUS, FUTURE PROSPECTS AND NOVEL APPLICATIONS FROM NANOPARTICLES' SYNTHESIS TO (LUNG) CANCER
THERMODYNAMICS OF IRREVERSIBLE AGGREGATION 120
Themis Matsoukas

MS 112: ANEURYSMS: SOLID MECHANICS, FLUID MECHANICS, AND MECHANOBIOLOGY
RELATIVE ROLES OF MECHANICS AND BIOCHEMISTRY IN THE INITIATION AND PROGRESSION OF CEREBRAL ANEURYSM THROMBOSIS 137
Malebogo Ngoepe, Yiannis Ventikos

TECHNIQUES TO INTEGRATE PATIENT-SPECIFIC SIMULATION OF ANEURYSMAL BLOOD FLOW INTO THE CLINICAL WORKFLOW 147
Gábor Závodszky, Roland Joó-Kovács, György Paál, István Szikora

MS 113: MATHEMATICAL AND NUMERICAL MODELING OF THE HEART
PROXIMAL ISOVELOCITY SURFACE FOR DIFFERENT MITRAL VALVE HOLE GEOMETRIES 155
Alexandre This, Hernán G. Morales, Odile Bonnefous

COMPARISON OF TETRAHEDRAL AND HEXAHEDRAL MESHES FOR FINITE ELEMENT SIMULATION OF CARDIAC ELECTRO-MECHANICS 164
Bernardo Lino De Oliveira, Joakim Sundnes

MS 115: TUMOR GROWTH MODELING AND THE MECHANICAL ASPECTS OF CANCER
A PARAMETRIC STUDY OF A MULTIPHASE POROUS MEDIA MODEL FOR TUMOR SPHEROIDS AND ENVIRONMENT INTERACTIONS 178
P. Mascheroni, D.P. Boso, C. Stigliano, M. Carfagna, L. Preziosi, P. Decuzzi, B.A. Schrefler

200 DAMAGE, FRACTURE AND FAILURE

MS 201: MICROSTRUCTURE-DRIVEN DEFORMATION AND FAILURE IN CRYSTALLINE MATERIALS
STABILITY OF ASYMMETRIC GRAIN BOUNDARIES IN GRAPHENE 188
J.P. Mendez, F. Arca, M.P. Ariza

DEFORMATION-DIFFUSION COUPLED ANALYSIS OF LONG-TERM HYDROGEN DIFFUSION IN NANOFILMS 197
Xingsheng Sun, Pilar Ariza, Kevin Wang

MESH OBJECTIVE DAMAGE MODELING OF DUCTILE FRACTURE AT VISCO-PLASTIC CONTINUUM RESPONSE ... 209
Senad Razanica, Ragnar Larsson, B. Lennart Josefson

MS 202: CIVIL ENGINEERING MATERIALS AND STRUCTURES UNDER EXTREME LOADINGS

EQUIVALENT SDOF MODEL FOR ESTIMATING BLAST-INDUCED DYNAMIC REACTIONS OF EQUILATERAL TRIANGULAR HARDENED WALL ELEMENTS 216
Sebastian Mendes, Liling Cao, Douglas Heinze, Elisabeth Malsch

MODELLING STRATEGIES OF PRESTRESSING TENDONS AND REINFORCEMENT BARS IN CONCRETE STRUCTURES 234
Antoine Llau, Ludovic Jason, Frédéric Dufour, Julien Baroth

MODELING AND SIMULATION FOR AN OPTIMIZED DESIGN OF A DYNAMIC BEND TEST. 242
Rana Akiki, Fabrice Gatuingt, Cédric Giry, Nicolas Schmitt, Lavinia Stéfan

MS 203: COMPUTATIONAL METHODS FOR MODELLING INSTABILITIES IN SOLIDS & STRUCTURES

ON STABILITY LOSS OF A THIN-WALLED SPHERICAL SHELL SUBJECTED TO EXTERNAL PRESSURE AND INTERNAL HOMOGENEOUS CORROSION 254
Emmanuel Gutman, Rudolf Bergman, Semyon Levitski

EFFICIENCY OF DAMAGE-PLASTICITY MODELS IN CAPTURING COMPACTION-EXPANSION TRANSITION OF CONCRETE UNDER DIFFERENT COMPRESSION LOADING CONDITIONS 261
Reza Mousavi, Masoud D. Champiri, Kaspar J. Willam

MS 204: IMPACT AND CRASH MECHANICS

A SIMPLIFIED MODEL OF A STEEL COLUMN SUBJECTED TO IMPACT 273
Piseth Heng, Mohammed Hjiaj, Jean-Marc Battini

MATERIAL MODELLING OF CAST ALUMINIUM BY APPLICATION OF THE WILKINS DAMAGE MODEL 287
Christian Mühlstätter, Matthias Hartmann

FINITE ELEMENT SIMULATION OF CRACK PROPAGATION AND DELAMINATION IN LAYERED SHELLS DUE TO BLADE CUTTING 295
Federica Confalonieri, Aldo Ghisi, Umberto Perego

EVOLUTIONARY LEVEL SET METHOD FOR CRASHWORTHINESS TOPOLOGY OPTIMIZATION 309
Mariusz Bujny, Nikola Aulig, Markus Olhofer, Fabian Duddeck

MS 205: PROBABILISTIC APPROACH TO NUMERICAL SIMULATION OF FRACTURE

NUMERICAL SIMULATION OF HIGH-VELOCITY INTERACTION ELEMENTS WITH THIN-WALLED STRUCTURES ... 323
Alexander V. Gerasimov, Sergey V. Pashkov

NUMERICAL SIMULATION OF MULTILAYER COMPOSITES FAILURE UNDER DYNAMIC LOADING 334
Sergey A. Zelepugin, Aleksey S. Zelepugin, Vladimir F. Tolkachev

SIMULATION OF MECHANICAL PROPERTIES OF CERAMIC PARTS PRODUCED BY ADDITIVE TECHNOLOGIES IN WIDE RANGE OF LOADING RATES 341
Vladimir A. Skripnyak, Evgeniya Skripnyak, Vladimir V. Skripnyak, Irina K. Vaganova

FRACTURE OF THIN METAL SHEETS WITH DISTRIBUTION OF GRAIN SIZES IN THE LAYERS 355
Natalia V. Skripnyak, Vladimir V. Skripnyak, Vladimir A. Skripnyak

BRITTLE OR QUASI-BRITTLE FRACTURE OF CERAMIC NANOCOMPOSITES UNDER DYNAMIC LOADING 366
Evgeniya G. Skripnyak, Vladimir A. Skripnyak, Vladimir V. Skripnyak, Natalia V. Skripnyak, Irina K. Vaganova

300 DISCRETIZATION METHODS, GRID, MESH AND SOLID GENERATION

MS 301: METHODS FOR CUT AND COMPOSITE MESHES: THEORY, ALGORITHMS AND APPLICATIONS
REMOVING THE STABILIZATION PARAMETER IN FITTED AND UNFITTED SYMMETRIC NITSCHE FORMULATIONS 373
Christoph Lehrenfeld

MS 302: MESH GENERATION AND ADAPTION
A SEMI-UNSTRUCTURED TURBOMACHINERY MESHING LIBRARY WITH FOCUS ON MODELING OF SPECIFIC GEOMETRICAL FEATURES 384
Marco Stelldinger, Thomas Giersch, Felix Figaschewsky, Arnold Kühhorn

CARTESIAN MESHES WITH DYNAMIC LOCAL WAVELET-BASED REFINEMENT FOR FLOW AROUND BODY PROBLEMS. 402
Kirill Merkulov, Igor Menshov, Andrew Plenkin

A MESH GENERATION METHOD FOR HISTORICAL MONUMENTAL BUILDINGS: AN INNOVATIVE APPROACH ... 409
Giovanni Castellazzi, Antonio Maria D'Altri, Stefano De Miranda, Francesco Ubertini, Gabriele Bitelli, Alessandro Lambertini, Ilenia Selvaggi, Antonio Tralli

MS 303: CURVED MESH GENERATION FOR HIGH-ORDER METHODS
HIGH-ORDER CURVED MESH GENERATION BY USING A FINE LINEAR TARGET MESH 417
Verena Schmid, Hadrien Beriot, Onur Atak, Gwenael Gabard

AUTOMATIC GENERATION OF 3D UNSTRUCTURED HIGH-ORDER CURVILINEAR MESHE 428
Michael Turner, David Moxey, Spencerj.Sherwin, Joaquim Peiro

MESH GENERATION FOR THE 2D NURBS-ENHANCED FINITE ELEMENT METHOD 444
Ruben Sevilla, Luke Rees, Oubay Hassan

IMAGE SEGMENTATION TECHNIQUES FOR BIOMEDICAL MODELING: ELECTROPHYSIOLOGY AND HEMODYNAMICS 454
Alexander Danilov, Roman Pryamonosov, Alexandra Yurova

A CONSISTENTLY LINEARISED SOLID MECHANICS BASED MESH DEFORMATION TECHNIQUE FOR HIGH ORDER CURVED ELEMENTS 462
Roman Poya, Ruben Sevilla, Antonio Gil

MS 304: COMPUTATIONAL MODELLING OF ADDITIVE PRODUCTION PROCESSES
VALIDATION OF MODELLING ASSUMPTIONS FOR THE BUILDUP SIMULATION OF LASER BEAM MELTING ON THE BASIS OF THE RESIDUAL STRESS DISTRIBUTION 469
Fabian Bayerlein, Christian Zeller, Martin Wunderer, Johannes Weirather, Michael F. Zäh, Markus Schmid, Georg Schlick, Roland Hessert, Michael Hofmann, Thomas Uihlein

MS 305: ADVANCED MESHING METHODS FOR INDUSTRIAL APPLICATIONS
MESHING OCEAN DOMAINS FOR COASTAL ENGINEERING APPLICATIONS 480
Alexandros Avdis, Christian Jacobs, Simon Mouradian, Jon Hill, Matthew Piggott

HIGH-ORDER TRACKING METHOD FOR THE SIMULATION OF BURNING FRONTS IN 3D 493
Denis Gueyffier, Bastien Andrieu

SINGULAR MESH GENERATION FROM MULTIPLE OVERSET MESHES: A TOOL FOR INDUSTRIAL APPLICATIONS...501
Gennaro Abbruzzese, Marta Cordero Gracia, Mariola Gómez López, Nuno Vinha

MS 306: LATTICE SPRING METHODS FOR LINEAR AND NONLINEAR CONTINUA
GENERATING TRIANGULAR LATTICES FOR SURFACES WITH IRREGULAR BOUNDARY 511
Tatiana Sá Marques, Vitor Dias Da Silva

WAVE TRANSMISSION THROUGH NONLINEAR IMPACTING METAMATERIAL UNIT 526
Arnab Banerjee, Raj Das, Emilio Calius

400 FLOW PROBLEMS

MS 403: PARTICLE-BASED METHODS IN FLUID MECHANICS
A PARTICLE FINITE ELEMENT METHOD FOR MACHINING SIMULATIONS 539
Juan Manuel Rodríguez Prieto, Pär Jonsén, Ales Svoboda

DYNAMICS OF DUST PARTICLES NEAR THE ROUND BELL SUCTION INLET 554
Olga Averkova, Konstantin Logachev, Artur Logachev, Elena Tolmacheva

RECOVERY OF DIFFERENTIATION/INTEGRATION COMPATIBILITY OF MESHLESS OPERATORS VIA LOCAL ADAPTATION OF THE POINT CLOUD IN THE CONTEXT OF NODAL INTEGRATION 568
Gabriel Fougeron, Guillaume Pierrot, Denis Aubry

MODELING OF NON-SPHERICAL, ELONGATED PARTICLES FOR INDUSTRIAL SUSPENSION FLOW SIMULATION ... 586
Jakob Dominik Redlinger-Pohn, Lisa Maria König, Christoph Kloss, Christoph Goniva, Stefan Radl

A LAGRANGIAN FINITE ELEMENT METHOD FOR THE SIMULATION OF 3D COMPRESSIBLE FLOWS 600
Massimiliano Cremonesi, Attilio Frangi

A LAGRANGIAN PFEM APPROACH TO THE NUMERICAL SIMULATION OF 3D LARGE SCALE LANDSLIDES IMPINGING IN WATER RESERVOIRS 608
Massimiliano Cremonesi, Francesco Ferri, Umberto Perego

A NUMERICAL INVESTIGATION OF FLOW DYNAMICS OVER A TRAPEZOIDAL SMOOTH OPEN CHANNEL ……..... 619
Giacomo Viccione

A NUMERICAL INVESTIGATION OF LIQUID IMPACT ON PLANAR SURFACES. …….. 627
Giacomo Viccione, Vittorio Bovolin, Eugenio Pugliese Carratelli

AERODYNAMICS OF AN UNSTEADY PARTICLE FLOW IN THE CHUTE …….. 638
Olga Averkova, Ivan Logachev, Konstantin Logachev

MS 404: SIMULATION OF ENVIRONMENTAL FLOWS
PROJECTION-BASED MODEL REDUCTION FOR FINITE ELEMENT APPROXIMATION OF SHALLOW WATER FLOWS …….. 650
Matthew Farthing, Alexander Lozovskiy, Christopher Kees

EFFICIENT TWO-DIMENSIONAL SIMULATION MODELS FOR HYDRAULIC AND MORPHODYNAMIC TRANSIENTS..670
Pilar Garcia-Navarro, Javier Murillo, Mario Morales-Hernandez, Carmelo Juez, Asier Lacasta

ADDRESSING THE CHALLENGES OF IMPLEMENTATION OF HIGH-ORDER FINITE-VOLUME SCHEMES FOR ATMOSPHERIC DYNAMICS ON UNSTRUCTURED MESHES …….. 684
Panagiotis Tsoutsanis, Dimitris Drikakis

MODELLING OF LONG WAVES (SEICHES) IN CASCAIS BAY …….. 709
Vera Bras, Antonio Trigo-Teixeira

A MESHLESS METHOD FOR SIMULATION OF PARTICLE-DRIVEN GRAVITY CURRENTS WITH OBSTACLES ……. 720
Karel Kovářík, Jozef Mužík, Dana Sitányiová

APPLICATION OF THE SHALLOW WATER EQUATIONS TO REAL FLOODING CASE …….. 735
Hani Ali, Pierre-Yves Lagrée, Jose-Maria Fullana

MS 405: COMPUTATIONAL MODELING OF MULTIPHASE FLOWS: ADVANCED METHODS, INTERFACE PHENOMENA AND ENVIRONMENTAL APPLICATIONS
A MULTIPHASE MODEL FOR THE NUMERICAL SIMULATION OF ICE-FORMATION IN SEA-WATER …….. 750
Vanessa Covello, Antonella Abbà, Luca Bonaventura, Alessandro Della Rocca, Lorenzo Valdettaro

A 3D CFD NUMERICAL STUDY OF THE BUBBLE GENERATION PROCESS INTO A BUBBLE T-JUNCTION GENERATOR AND ITS COMPARISON WITH EXPERIMENTAL DATA: PART I …….. 772
Santiago Arias, Adeline Montlaur

MS 406: ADVANCES IN COMPUTATIONAL METHODS FOR GAS-LIQUID TWO-PHASE FLOW
COUPLING OF FLUID STUCTURE INTRACTION SOLVER WITH A VOF METHOD FOR MULTIPHASE STRUCTURE INTERACTION …….. 788
Daniele Cerroni, Roberto Da Via', Sandro Manservisi, Filippo Menghini

A METHOD OF CHAINED ANALYTICAL WAVE STRUCTURES FOR LARGE SCALE STRATIFIED TWO-PHASE PIPE FLOWS …….. 806
Andreas Holm Akselsen

MS 408: MANIPULATION AND CONTROL OF TURBULENT FLOW
MODELING OF EXCITATION OF CONTROLLING DISTURBANCES IN SWEPT WING BOUNDARY LAYER BY MEANS OF PLASMA ACTUATORS 831
Sergey Chernyshev, Andrey Kiselev, Aleksandr Kuryachii

NUMERICAL SIMULATION OF INTERACTION PROCESS BETWEEN DIELECTRIC BARRIER DISCHARGE AND DUCT FLOW 844
Pavel Semenev, Dmitriy Pudovikov, Pavel Toktaliev

IMPLEMENTATION OF POD AND DMD METHODS IN APACHE SPARK FRAMEWORK FOR SIMULATION OF UNSTEADY TURBULENT FLOW IN THE MODEL COMBUSTOR 857
Mikhail Kalugin, Sergei Strijhak

MS 409: CURRENT TRENDS IN MODELLING AND SIMULATION OF TURBULENT FLOWS
NUMERICAL VALIDATION OF A FOUR PARAMETER LOGARITHMIC TURBULENCE MODEL 865
Daniele Cerroni, Roberto Da Vià, Sandro Manservisi, Filippo Menghini

LARGE-EDDY SIMULATION AND FAR FIELD ACOUSTICS OF A SUBSONIC HOT JET 878
O. Labbé

ON THE APPLICATION OF MODERN TURBULENCE MODELS IN THE FLOW LAMINARIZATION PROBLEMS 893
Sergey Chernyshev, Aleksander Ivanov, Andrey Kiselev, Dmitriy Sboev, Leonid Teperin, Petr Vorotnikov, Valeriy Vozhdaev

ASSESSMENT OF RANS AND DES METHODS FOR THE AHMED BODY 912
Emmanuel Guilmineau, Gan Bo Deng, Alban Leroyer, Patrick Queutey, Michel Visonneau, Jeroen Wackers

A WAY TO IMPROVE JET MODELING WITHIN RANS EQUATION SYSTEM 922
Alexey Troshin

A MODIFIED SSG/LLR-ω REYNOLDS STRESS MODEL FOR PREDICTING BLUFF BODY AERODYNAMICS 936
Csaba Klajbár, László Könözsy, Karl W. Jenkins

MS 410: COMPLEX FLUID FLOWS IN ENGINEERING: MODELLING, SIMULATION AND OPTIMIZATION
NUMERICAL ANALYSIS OF RAREFIED GAS FLOWS USING THE ACADEMIC CFD CODE GALATEA 949
Angelos Klothakis, Georgios Lygidakis, Ioannis Nikolos

TOPOLOGY OPTIMIZATION FOR FLUID FLOW EMPLOYING LOCAL OPTIMALITY CRITERIA 967
Philip Sarstedt, Gerhard Kachel, Jörg Ettrich, Karl Bühler

NUMERICAL SIMULATION OF THE CONJUGATE HEAT TRANSFER IN THE COOLING SYSTEM OF THE COMBUSTION CHAMBERS OF THE AVIATION RAMJET ON THE ENDOTHERMIC FUELS 979
Vadim Volokhov, Pavel Toktaliev, Sergei Martynenko

SENSITIVITY STUDY OF THE NUMERICAL SETUP FOR AN AUTOMATIC OPTIMIZATION PROCEDURE FOR A HYDRAULIC MACHINE 992
Alexander Tismer, Markus Schlipf, Stefan Riedelbauch

ADVANCES IN GEOMETRICAL PARAMETRIZATION AND REDUCED ORDER MODELS AND METHODS FOR COMPUTATIONAL FLUID DYNAMICS PROBLEMS IN APPLIED SCIENCES AND ENGINEERING: OVERVIEW AND PERSPECTIVES 1013
F. Salmoiraghi, F. Ballarin, G. Corsi, A. Mola, M. Tezzele, G. Rozza

NEAR WALL BEHAVIOR OF IMPLICIT LARGE EDDY SIMULATIONS 1032
Ioannis Kokkinakis, Dimitris Drikakis

VALIDATION AND VERIFICATION OF A 2D LATTICE BOLTZMANN SOLVER FOR INCOMPRESSIBLE FLUID FLOW 1046
Tamás István Józsa, Máté Szőke, Tom-Robin Teschner, László Könözsy, Irene Moulitsas

MS 411: NON-NEWTONIAN HEAT AND FLUID FLOW SUBJECTED TO MAGNETIC FORCES
VALIDATION OF A MAGNETO- AND FERRO-HYDRODYNAMIC MODEL FOR NON-ISOTHERMAL FLOWS IN CONJUNCTION WITH NEWTONIAN AND NON-NEWTONIAN FLUIDS 1061
László Könözsy, Pietro Scienza, Dimitris Drikakis

NUMERICAL STUDY OF MAGNETIC PARTICLES CONCENTRATION IN BIOFLUID (BLOOD) UNDER THE INFLUENCE OF HIGH GRADIENT MAGNETIC FIELD IN MICROCHANNEL 1084
Vassilios Loukopoulos, George Bourantas, Dimitrios Labropoulos, Vassilios-Martin Nikiforidis, Stéphane Bordas, George Nikiforidis

MS 412: NUMERICAL METHODS FOR WAVES AND FLOWS IN COASTAL AND DEEP WATER HYDRODYNAMICS
NEW ALGORITHM FOR NUMERICAL SIMULATION OF SURFACE WAVES WITHIN THE FRAMEWORK OF THE FULL NONLINEAR DISPERSIVE MODEL 1093
Zinaida Fedotova, Oleg Gusev, Gayaz Khakimzyanov

NUMERICAL MODELLING OF SURFACE WAVES IN THE FRAMEWORK OF THE SHALLOW WATER MODEL 1104
Nina Shokina, Gayaz Khakimzyanov

ON NUMERICAL METHODS FOR SOLVING RUN-UP PROBLEMS. COMPARATIVE ANALYSIS OF NUMERICAL ALGORITHMS AND NUMERICAL RESULTS. 1127
Leonid Chubarov, Alexandr Rychkov, Gayaz Khakimzyanov, Yurii Shokin

MS 414: NEW TRENDS IN NUMERICAL METHODS FOR MULTI-MATERIAL COMPRESSIBLE FLUID FLOWS
EULERIAN CALCULATIONS OF MULTIMATERIAL FLOWS WITH SUB-CELL RECONSTRUCTION OF INTERFACES ... 1139
Igor Menshov, Pavel Zakharov

SOME ACOUSTIC-TRANSPORT SPLITTING SCHEMES FOR TWO-PHASE COMPRESSIBLE FLOWS 1151
Simon Peluchon, Gérard Gallice, Pierre-Henri Maire

LAGRANGE-FLUX EULERIAN SCHEMES FOR COMPRESSIBLE MULTIMATERIAL FLOWS 1165
Florian De Vuyst, Thibault Gasc, Renaud Motte, Mathieu Peybernes, Raphael Poncet

ISOTROPY–PRESERVING SLOPE LIMITERS FOR FINITE VOLUME METHODS ON SQUARE MESHES 1179
Jan Velechovsky, Marianne Francois

BUILDING A MORE EFFICIENT LAGRANGE-REMAP SCHEME THANKS TO PERFORMANCE MODELING 1191
Thibault Gasc, Florian De Vuyst, Mathieu Peybernes, Raphaël Poncet, Renaud Motte

MS 415: COMPUTATIONAL NON-NEWTONIAN FLUID MECHANICS
DEVELOPMENT OF CONFINED VISCOPLASTIC FLOWS WITH HETEROGENEOUS WALL SLIP 1205
Pandelitsa Panaseti, Maria Philippou, Zacharias Kountouriotis, Georgios Georgiou

SQUEEZE FLOW OF VISCOPLASTIC BINGHAM MATERIAL 1215
Larisa Muravleva

500 HIGH PERFORMANCE COMPUTING

MS 501: ALGORITHMIC ASPECTS OF HIGH-PERFORMANCE COMPUTING FOR MECHANICS AND PHYSICS
PSEUDOMULTIGRID GAUSS–SEIDEL METHOD FOR LARGE SCALE AND HIGH PERFORMANCE COMPUTING 1237
Sergey Martynenko, Vadim Volokhov, Pavel Toktaliev

ALGORITHM FOR FAST SIMULATIONS OF SPACE-TIME FINITE ELEMENT METHOD 1260
Marcin Skotniczny, Anna Paszynska, Maciej Paszynski

HYBRID PARALLELISATION OF AN ALGORITHMICALLY DIFFERENTIATED ADJOINT SOLVER 1268
Pavanakumar Mohanamuraly, Jan Christian Huckelheim, Jens-Dominik Mueller

MS 503: HPC-BASED SIMULATIONS FOR THE ENGINEERING REALM AND INDUSTRIAL APPLICATIONS
EFFICIENT PARALLEL GEOMETRY DISTRIBUTION FOR THE SIMULATION OF COMPLEX FLOWS 1277
Andreas Lintermann

LARGE-SCALE FREE-SURFACE FLOW SIMULATION USING LATTICE BOLTZMANN METHOD ON MULTI-GPU CLUSTERS 1294
Naoyuki Onodera, Kunihide Ohashi,

LARGE-SCALE LES ANALYSIS OF AUTOMOTIVE ENGINE COOLING FAN 1305
Yuji Kobayashi, Kenji Yoshida, Itsuhei Kohri, Masaharu Sakai, Hideo Asano

DIRECT AEROACOUSTIC SIMULATION RELATED WITH MODE CHANGE IN A RECORDER 1315
Hiroshi Yokoyama, Ryoma Hamasuna, Akira Miki, Hirofumi Onitsuka, Akiyoshi Iida

NUMERICAL ANALYSIS OF RADIO FREQUENCY BLACKOUT FOR ATMOSPHERIC REENTRY VEHICLE USING CFD-CEM COMBINED METHOD 1325
Yusuke Takahashi, Reo Nakasato, Nobuyuki Oshima

DIRECT FEM LARGE SCALE COMPUTATION OF TURBULENT MULTIPHASE FLOW IN URBAN WATER SYSTEMS AND MARINE ENERGY 1339
Ezhilmathi Krishnasamy, Johan Hoffman, Johan Jansson

MS 504: NUMERICAL METHODS AND TOOLS FOR KEY EXASCALE COMPUTING CHALLENGES IN ENGINEERING AND APPLIED SCIENCES
KINETIC MODELS AND ALGORITHMS FOR SOLUTION OF THE MAGNETOGASDYNAMIC PROBLEMS ON THE MODERN SUPERCOMPUTING SYSTEMS 1352
Boris Chetverushkin, Nicola D'Ascenzo, A. Saveliev, Valeri Saveliev

ERROR ANALYSIS AND QUANTIFICATION IN NEURON SIMULATIONS 1366
Francesco Casalegno, Francesco Cremonesi, Stuart Yates, Michael L. Hines, Felix Schürmann, Fabien Delalondre

EVALUATION OF SPARSE LINEAR ALGEBRA OPERATIONS IN TRILINOS 1381
Mohammad Siahatgar, Gabriele Von Voigt

MS 505: INTERACTIVE SIMULATIONS IN COMPUTATIONAL ENGINEERING
AUTOMATIC VISUALIZATION AND CONTROL OF ARBITRARY NUMERICAL SIMULATIONS 1392
Jan P. Springer, Helen Wright

MS 506: ACCURACY AND EFFICIENCY OF APPROXIMATE COMPUTATIONS IN SCIENCE AND ENGINEERING
DIRECT TIME-DOMAIN INTEGRATION APPROACH FOR VISCOELASTIC SYSTEMS INVOLVING VARIOUS DAMPING MODELS 1409
Zhe Ding, Li Li, Yujin Hu

ON THE PERFORMANCE OF A COMPUTATIONAL COST REDUCTION TECHNIQUE APPLIED TO COOLING TOWERS TRANSIENT ANALYSIS 1421
Aram Soroushian, Amin Jahani Mehrnoosh, Yalda Zarabimanesh, Mohammad Hadi Ghondaghsaz, Ali Baiani, Ali Zakizade

A DIFFERENT LOOK AT THE RICHARDSON EXTRAPOLATION LEADING TO A NEW PROPOSITION 1432
Aram Soroushian

ON THE FREQUENCY CONTENT OF ERRORS ORIGINATED IN A TIME INTEGRATION COMPUTATIONAL COST REDUCTION TECHNIQUE 1443
Aram Soroushian, Mahmood Hosseini, Seyed Mohammad Khalkhali

600 INTERDISCIPLINARY COUPLED AND CONTACT PROBLEMS

MS 601: SHOCK WAVE-BOUNDARY LAYER INTERACTION AND ITS CONTROL
NUMERICAL SIMULATION OF TRANSONIC BUFFET AND ITS CONTROL USING TANGENTIAL JET BLOWING 1455
Ksenia Abramova, Kamil Khairullin, Alexander Ryzhov, Vitaly Soudakov

SHOCK WAVE BOUNDARY LAYER INTERACTION IN PROFILE CASCADES REPRESENTING ROTOR BLADINGS OF THE LAST STAGE OF LARGE OUTPUT STEAM TURBINES 1464
Martin Luxa, David Simurda, Jaromir Prihoda, Jana Vachova

SHOCK WAVE BOUNDARY LAYER INTERACTION INVESTIGATION: CONTROL BY ROD VORTEX GENERATORS AND THEIR APPLICATION ON HELICOPTER ROTOR BLADES 1474
Fernando Tejero Embuena, Piotr Doerffer, Pawel Flaszynski

ANALYSIS OF EFFECTS OF SHAPE AND LOCATION OF MICRO - TURBULATORS ON UNSTEADY SHOCKWAVE - BOUNDARY LAYER INTERACTION IN TRANSONIC FLOW 1484
Janusz Sznajder, Tomasz Kwiatkowski

MS 602: INNOVATIVE METHODS FOR FLUID-STRUCTURE-INTERACTION
ASSESSMENT OF THE FLUID-STRUCTURE INTERACTION CAPABILITIES FOR AERONAUTICAL APPLICATIONS OF THE OPEN-SOURCE SOLVER SU2. 1498
Ruben Sanchez, Heather L. Kline, David Thomas, Anil Variyar, Marcello Righi, Thomas D. Economon, Juan J. Alonso, Rafael Palacios, Grigorios Dimitriadis, Vincent Terrapon

CONVERGENCE ANALYSIS OF COUPLING ITERATIONS FOR THE UNSTEADY TRANSMISSION PROBLEM WITH MIXED DISCRETIZATIONS 1530
Azahar Monge, Philipp Birken

USING IMPROVED RADIAL BASIS FUNCTIONS METHODS FOR FLUID-STRUCTURE COUPLING AND MESH DEFORMATION 1545
Giorgos Strofylas, Georgios Mazanakis, Sotirios Sarakinos, Georgios Lygidakis, Ioannis Nikolos

A MONOLITHIC FLUID-STRUCTURE INTERACTION METHOD, APPLICATION TO A PISTON PROBLEM 1564
Felix Ischinger, Martijn Anthonissen, Barry Koren

MS 603: COMPUTATIONAL METHODS IN FLUID-STRUCTURE INTERACTION WITH IMPACT ON INDUSTRIAL APPLICATIONS
SLAMMING IMPACT SIMULATION OF 2D WATER ENTRY FOR RIGID STRUCTURES 1582
Omar Hashim Hassoon, Mostapha Tarfaoui, A. El Malki

A MESH MORPHING BASED FSI METHOD USED IN AERONAUTICAL OPTIMIZATION APPLICATIONS 1594
Matej Andrejašič, David Eržen, Emiliano Costa, Stefano Porziani, Marco Evangelos Biancolini, Corrado Groth

A PATH-CONSERVATIVE OSHER-TYPE SCHEME FOR AXIALLY SYMMETRIC COMPRESSIBLE FLOWS IN FLEXIBLE TUBES 1613
Julia Leibinger, Michael Dumbser, Uwe Iben

MS 605: FRICTIONAL CONTACTS WITH LUBRICATION – BASICS AND APPLICATIONS
LIMITING SHEAR STRESS FORMULATION FOR TEHL SIMULATION 1623
Thomas Lohner, Klaus Michaelis, Karsten Stahl

MS 606: COMPUTATIONAL MODELING OF HYDRAULIC FRACTURING
THE XFEM FOR A SIMPLIFIED MODEL IN HYDRAULIC FRACTURING 1637
Markus Schätzer, Thomas-Peter Fries

SENSITIVITY ANALYSIS OF HYDRAULIC FRACTURING USING AN EXTENDED FINITE ELEMENT METHOD FOR THE PKN MODEL 1647
Hasini Garikapati, Clemens V. Verhoosel, Harald Van Brummelen, Pedro Diez

NUMERICAL MODELLING OF COUPLED HYDRO-THERMAL PROCESSES OF THE SOULTZ HETEROGENEOUS GEOTHERMAL SYSTEM 1659
Musa Dahiru Aliyu, Hua-Peng Chen

MS 607: ADVANCES IN COMPUTATIONAL METHODS FOR LIQUID-VAPOR FLOWS WITH PHASE TRANSFER PROCESSES
A LOW DIFFUSIVE APPROACH FOR TWO PHASE FLOWS 1672
Alberto Beccantini

A LARGE TIME STEP ROE SCHEME APPLIED TO TWO-PHASE FLOW 1693
Sofia Lindqvist, Halvor Lund

MS 608:ADVANCES IN TIME INTEGRATION FOR SOLID, FLUID AND COUPLED SYSTEMS
ENERGY CONSERVING TIME INTEGRATION BASED ON GALERKIN-VARIATIONAL INTEGRATORS WITH CONSTRAINTS 1709
Matthias Bartelt, Michael Groß

MIXED INTEGRATORS FOR STRUCTURE-PRESERVING SIMULATIONS IN NONLINEAR STRUCTURAL DYNAMICS 1736
Alexander Janz, Peter Betsch, Christian Hesch

ENERGY AND MOMENTUM CONSERVING VARIATIONAL BASED TIME INTEGRATION OF ANISOTROPIC HYPERELASTIC CONTINUA 1765
Michael Groß, Rajesh Ramesh, Julian Dietzsch

STRUCTURE-PRESERVING OPTIMAL CONTROL OF DISCRETE MECHANICAL SYSTEMS 1789
Peter Betsch, Christian Becker

VARIATIONAL INTEGRATORS OF MIXED ORDER FOR DYNAMICAL SYSTEMS WITH MULTIPLE TIME SCALES AND SPLIT POTENTIALS 1818
Theresa Wenger, Sina Ober-Blöbaum, Sigrid Leyendecker

CO-SIMULATION OF LARGE MBD, FEA AND DEM SYSTEMS 1832
Jose Ortiz, Neil Macdonald

MS 609: ADVANCED COMPUTATIONAL MODELING OF BATTERIES AND FUEL CELLS
STOCHASTIC 3D MODELING OF AMORPHOUS MICROSTRUCTURES - A POWERFUL TOOL FOR VIRTUAL MATERIALS TESTING 1838
Matthias Neumann, Volker Schmidt

MS 610: NUMERICAL METHODS TO STUDY THE CONTACT MECHANICS OF DRY, ADHESIVE AND LUBRICATED ROUGH SURFACES
PERCOLATION PROPERTIES OF THE FREE VOLUME GENERATED BY TWO ROUGH SURFACES IN CONTACT 1850
Paolo Cinat, Marco Paggi, Claudia Borri

A PARAMETRICALLY TIME-DEPENDENT METHODOLOGY FOR RECIPROCATING CONTACT MECHANICS BETWEEN VISCOELASTIC SOLIDS 1856
Carmine Putignano, Giuseppe Carbone, Daniele Dini

STATISTICAL AND GEOMETRICAL CHARACTERISTICS OF RANDOMLY ROUGH SURFACES USED FOR CONTACT SIMULATIONS. 1864
Rafael Schouwenaars, Carlos G. Figueroa, Víctor Hugo Jacobo, Armando Ortiz

MS 612: NUMERICAL SIMULATIONS FOR SMART-CITY APPLICATIONS
NUMERICAL OPTIMIZATION OF NEAR-ROAD VEGETATION BARRIERS 1876
Ludek Benes, Viktor Sip

MS 615: COMPUTATIONAL MODELS IN MAGNETOHYDRODYNAMICS
LARGE EDDY SIMULATIONS OF MAGNETIC FIELD EFFECT ON TURBULENT FLOW IN A SQUARE DUCT 1886
Jie Mao, Kunlei Zhang, Zhongquan Tan, Ke Liu

700 MATERIALS

MS 701: ADVANCED MATERIALS: COMPUTATIONAL ANALYSIS OF PROPERTIES AND PERFORMANCE
OPTIMAL DESIGN AND ADDITIVE MANUFACTURING OF NOVEL REINFORCING ELEMENTS FOR COMPOSITE MATERIALS 1893
Francesco Fabbrocino, Ilenia Farina, Ada Amendola, Luciano Feo, Fernando Fraternali

A TEMPERATURE SENSITIVE CRYSTAL PLASTICITY MODEL FOR THE PREDICTION OF HIGH TEMPERATURE MECHANICAL BEHAVIOUR OF MULTI-PHASE TIAL ALLOY 1909
M. Umer Ilyas, M. Rizviul Kabir

LAYOUT OPTIMIZATION OF PIEZOELECTRIC ELEMENTS WITH EXTERNAL ELECTRIC CIRCUITS IN SMART CONSTRUCTIONS BASED ON SOLUTION OF THE NATURAL VIBRATIONS PROBLEM 1920
Nataliia Iurlova, Valerii P. Matveenko, Dmitriy A. Oshmarin, Nataliya V. Sevodina, Maksim A. Yurlov

DETERMINATION OF OPTIMAL PARAMETERS FOR A PASSIVE RL-CIRCUIT BY SOLVING THE PROBLEM ON NATURAL VIBRATIONS OF ELECTROELASTIC BODIES 1930
Maksim Yurlov, Valerii P. Matveenko, Dmitriy A. Oshmarin, Nataliya V. Sevodina, Nataliya A. Iurlova

COMPUTATIONAL MODELLING OF MECHANICAL BEHAVIOR OF BIOLOGICAL TISSUES FOR BIOMEDICAL APPLICATIONS 1940
Simin Li, Yang Liu, Juan Du, Begum Zeybek, Lorenzo Zani, Mark P. Lewis, Vadim V. Silberschmidt

ANALYSIS OF FIBRE BRAGG GRATINGS SENSORS OPTIMAL PLACEMENT FOR MONITORING OF DAMAGE PROPAGATION IN LAMINATE COMPOSITES 1950
Mikhail Tashkinov

SEVEN DIFFERENT WAYS TO MODEL VISCOELASTICITY IN A GEOMETRICALLY EXACT SETTING 1959
Alexey Shutov

MS 702: MODELING OF NANOFILLED COMPOSITES
SIMULATING ATOMIC FORCE MICROSCOPY FOR THE DETERMINATION OF THE ELASTIC PROPERTIES OF NANOPARTICLE REINFORCED EPOXY RESIN 1971
Johannes Fankhänel, Andreas Kempe, Raimund Rolfes

MULTI-SCALE FINITE ELEMENT ANALYSIS OF GRAPHENE/POLYMER NANOCOMPOSITES ELECTRICAL PERFORMANCE 1984
Asimina Manta, Matthieu Gresil, Constantinos Soutis

MS 705: IDENTIFICATION OF MATERIAL MODELS
THE BOOTSTRAP APPROACH TO THE STATISTICAL SIGNIFICANCE OF PARAMETERS IN RSM MODEL 2003
Jacek Pietraszek, Leszek Wojnar

MS 706: MODELING OF FIBER-BASED STRUCTURES - TEXTILES AND TEXTILE REINFORCED COMPOSITES
INVERSE IDENTIFICATION OF THE BENDING STIFFNESS OF A BRAIDED POLYETHYLENE TWINE SUBJECT TO LARGE DEFORMATION: APPLICATION TO THE IDENTIFICATION OF THE MESH OPENING RIGIDITY OF FISHING NETS 2010
Barthélémy Morvan, Guilhem Bles, Nicolas Dumergue, Daniel Priour

MULTI-SCALE FRAMEWORK FOR MODELLING CARBON FIBRE WEAVES BASED ON STOCHASTIC REINFORCEMENT GEOMETRY 2029
Andy Vanaerschot, Karen Soete, Stepan Lomov, Dirk Vandepitte

MS 707: MICROMECHANICAL MODELLING: COMPETITION BETWEEN ANALYTICAL AND NUMERICAL METHODS
MULTISCALE MODELING OF THERMO-ELASTIC PROPERTIES OF MICROCRACKED MATERIAL 2039
Antonino Favata, Patrizia Trovalusci, Renato Masiani

EFFECTIVE ELASTICITY COEFFICIENTS IN DRY POROUS MATERIALS. NUMERICAL AND SEMI-ANALYTICAL APPROACHES 2050
Vladimir Sladek, Bruno Musil, Jan Sladek

INFLUENCE OF MICRO CRACKS ON EFFECTIVE MATERIAL PROPERTIES IN FIBER REINFORCED SMART COMPOSITE MATERIALS 2073
Michael Wünsche, Jan Sladek, Vladimir Sladek

DEFECT ACCUMULATION IN NANOPOROUS WEAR-RESISTANT COATINGS UNDER COLLECTIVE RECRYSTALLIZATION. SIMULATION BY HYBRID CELLULAR AUTOMATON METHOD 2080
Dmitry Moiseenko, Pavel Maksimov, Sergey Panin, Viktor Panin

MS 711: FOURIER-BASED METHODS FOR COMPUTING THE BEHAVIOR OF HETEROGENEOUS MATERIALS DEVELOPMENTS, EXTENSIONS AND APPLICATIONS
NONLINEAR COMPOSITE VOXELS AND FFT-BASED HOMOGENIZATION 2099
Matthias Kabel, Andreas Fink, Felix Ospald, Matti Schneider

MS 712: SMART MATERIAL SYSTEMS AND STRUCTURES
A THERMODYNAMICALLY CONSISTENT FINITE STRAIN MICRO-SPHERE FRAMEWORK FOR PHASE-TRANSFORMATION 2110
Richard Ostwald, Thorsten Bartel, Andreas Menzel

OPTIMAL CONTROL TUNNING IN SMART STRUCTURES WITH DELAMINATIONS 2123
Panos Koutsianitis, Amalia Moutsopoulou, Georgios Drosopoulos, Georgios Tairidis, Georgia Foutsitzi, Georgios Stavroulakis

MS 713: MICROSTRUCTURE-BASED MODELLING OF HETEROGENEOUS MATERIALS
DEPENDENCE OF THE MECHANICAL PROPERTIES OF PENTAMODE MATERIALS ON THE LATTICE MICROSTRUCTURE 2134
Ada Amendola, Francesco Fabbrocino, Luciano Feo, Ferdinando Auricchio, Fernando Fraternali

HOMOGENIZATION TECHNIQUES FOR ACCURATE AND APPROXIMATE ESTIMATES FOR OVERALL PROPERTIES OF MICROCRACKED VISCOELASTIC MASONRIES 2151
Amna Rekik, Alain Gasser

APPLICATION OF POLAR ELASTICITY TO THE PROBLEM OF PURE BENDING OF A THICK PLATE 2162
Svitlana Fedorova, Jiří Burša

ARTIFICIAL MICROSTRUCTURE GENERATOR FOR DUAL-PHASE STEELS 2172
Yuliang Hou, Alexandre Dumon, Pièrre Culière, Mohamed Rachik

MS 714: STRENGTH, FATIGUE AND STABILITY OF COMPOSITE STRUCTURES
ON THE STRUCTURAL DESIGN OF IMPERFECTION SENSITIVE LAMINATED COMPOSITE SHELL STRUCTURES SUBJECTED TO AXIAL COMPRESSION 2182
Linus Friedrich, Pawel Lyssakow, Garth Pearce, Martin Ruess, Chiara Bisagni, Kai-Uwe Schröder

STRENGTH AND WEIGHT EQUIVALENT SUBSTITUTION OF LARGE SANDWICH PANELS BY MONOLITHIC CFRP STRUCTURES 2190
Martin Meindlhumer, Martin Schagerl

ROBUST INTEGRAL COMPOSITE AIRCRAFT STRUCTURES 2202
Tamas Havar

MS 715: COMPUTATIONAL ANALYSIS OF COMPOSITE STRUCTURES
MINIMAL MASS DESIGN OF STRENGTHENING TECHNIQUES FOR PLANAR AND CURVED MASONRY STRUCTURES 2210
Gerardo Carpentieri, Francesco Fabbrocino, Mariella De Piano, Valentino Paolo Berardi, Luciano Feo, Fernando Fraternali

MODELING OF FIBER-REINFORCED PLASTICS TAKING INTO ACCOUNT THE MANUFACTURING PROCESS …….. 2220
Cherry Ann Reclusado, Sumito Nagasawa

CONVERGENCE ANALYSIS OF STRESS FIELDS TO HOMOGENIZATION PREDICTIONS IN OPTIMAL PERIODIC COMPOSITE DESIGN …….. 2236
Pedro Coelho, Rui Reis, José Guedes

A NUMERICAL STUDY FOR THE BUCKLING CAPACITY OF WIND TURBINE BLADES: GEOMETRY, LOADING AND MATERIAL INFLUENCE …….. 2247
Efstathios E. Theotokoglou, George A. Balokas, Evgenia K. Savvaki

800 MULTISCALE PROBLEMS

MS 801: MULTISCALE COMPUTATIONAL HOMOGENIZATION FOR BRIDGING SCALES IN THE MECHANICS AND PHYSICS OF COMPLEX MATERIALS

A SEMI-IMPLICIT MICROPLAR DISCRETE-TO-CONTINUUM METHOD FOR GRANULAR MATERIALS …….. 2266
Kun Wang, Waiching Sun

MODELING OF LOW-ALLOYED TRIP-STEELS BASED ON DIRECT MICRO-MACRO SIMULATIONS …….. 2280
Stefan Prüger, Ashutosh Gandhi, Daniel Balzani

HOMOGENIZATION METHOD FOR DESIGNING NOVEL ARCHITECTURED CELLULAR MATERIALS …….. 2292
Zheng-Dong Ma

A HOMOGENIZATION TECHNIQUE FOR ELASTO-PLASTIC COMPOSITES …….. 2316
Federica Covezzi, Stefano De Miranda, Sonia Marfia, Elio Sacco

MS 802: ADVANCES IN THE MODELLING OF MULTI-SCALE, MULTI-PHYSICS AND MULTI-UNCERTAINTY PROBLEMS

THEORETICAL PREDICTION OF PLASTIC INSTABILITY FOR ASYMMETRICAL MATERIALS …….. 2323
Shenghua Wu, Nannan Song, F. M. Andrade Pires

MULTISCALE SIMULATION OF THREE-DIMENSIONAL UNSTEADY GAS FLOWS IN MICROCHANNELS OF TECHNICAL SYSTEMS …….. 2331
Viktoriia Podryga, Yury Karamzin, Tatiana Kudryashova, Sergey Polyakov

HIGHER-ORDER ASYMPTOTIC HOMOGENIZATION OF PERIODIC MATERIALS WITH LOW SCALE SEPARATION …….. 2346
M. Mohammed Ameen, R.H.J. Peerlings, M.G.D. Geers

MS 805: ADVANCED MULTI-PHYSICS AND MULTI-SCALE TECHNIQUES FOR MODELING INELASTIC PROCESSES IN SOLIDS: DAMAGE, FRACTURE AND CONTACT MECHANICS

MESO-SCALE MODELING OF HYBRID INDUSTRIAL/RECYCLED STEEL FIBER-REINFORCED CONCRETE …….. 2353
Antonio Caggiano, Diego Said Schicchi, Guillermo Etse, Enzo Martinelli

MS 806: MULTISCALE MODELLING OF MATERIALS AND STRUCTURES
THE KIRKENDALL SHIFT AND FRENKEL EFFECT DURING MULTI-COMPONENT DIFFUSION PROCESS 2363
Bartek Wierzba

CRACK PATH FIELD AND STRAIN INJECTION TECHNIQUES IN DYNAMIC FRACTURE SIMULATIONS 2372
Oriol Lloberas-Valls, Alfredo E. Huespe, Javier Oliver, Ivo F. Dias

HYBRID FE/XFE FINITE ELEMENT MODEL FOR SIMULATION OF BRITTLE-DUCTILE FRACTURES IN DUAL-PHASE STEEL GRADES. 2390
Konrad Perzynski, Lukasz Madej

EFFECTIVE MESO AND MACRO PROPERTIES FOR FIBRE-REINFORCED-POLYMER CURING COUPLED TO VISCO-ELASTICITY 2398
Rolf Mahnken, Christian Dammann

ON APPLICATION OF THE MESHLESS FINITE DIFFERENCE METHOD TO NUMERICAL HOMOGENIZATION 2417
Irena Jaworska

A MESOSCALE STUDY OF DEFORMATION AND FAILURE OF ANGLE-PLY LAMINATES UNDER TENSILE LOADING BY MEANS OF NUMERICAL HOMOGENIZATION 2426
Marek Romanowicz

MODELING OF PHASE CHANGES IN MICRO-DOMAIN INDUCED BY AN ULTRASHORT LASER PULSE 2437
Ewa Majchrzak, Lukasz Turchan, Jolanta Dziatkiewicz

MS 807: ADVANCED COMPUTATIONAL STRATEGIES FOR MODELLING, SIMULATION AND CHARACTERISATION OF MULTI-SCALE HETEROGENEOUS MATERIALS
VALIDATION OF A DISCRETE CRACK MODEL FOR LIGHTWEIGHT AGGREGATE CONCRETE BEAMS 2449
D. Dias-Da-Costa, R. Graça-E-Costa, R.N.F. Carmo

MS 808: MULTISCALE AND MULTIPHYSICS MODELING OF CEMENTITIOUS MATERIALS
REINFORCEMENT CORROSION IN CONCRETE DUE TO CARBONATION AND CHLORIDE INGRESS UP AND BEYOND INDUCTION PERIOD 2460
Karolina Hájková, Libor Jendele, Vit Šmilauer, T. Sajdlová, Jan Červenka

MS 809: MULTISCALE STOCHASTIC FINITE ELEMENT METHODS
THE STOCHASTIC FINITE ELEMENT METHOD FOR NUCLEAR APPLICATIONS 2471
José David Arregui Mena, Lee Margetts, Llion Evans, D. V. Griffiths, Anton Shterenlikht, Luis Cebamanos, Paul M Mummery

DETERMINATION OF MESOSCALE RANDOM FIELDS FOR THE APPARENT PROPERTIES OF SPATIALLY RANDOM COMPOSITES 2484
Dimitrios Savvas, George Stefanou

MS 811: MULTISCALE MODELING OF CONCRETE AND CONCRETE STRUCTURES
NUMERICAL SIMULATION OF THE AUTOGENOUS SHRINKAGE OF HARDENING PORTLAND CEMENT PASTE ... 2501
Peng Gao, Guang Ye, Jiangxiong Wei, Qijun Yu

900 NUMERICAL SIMULATION METHODS

MS 901: ISOGEOMETRIC METHODS
ANALYSIS OF LAMINATED COMPOSITE PLATES USING ISOGEOMETRIC COLLOCATION METHOD 2514
G. S. Pavan, K. S. Nanjunda Rao

MODELLING STIFFENED LIGHTWEIGHT STRUCTURES WITH ISOGEOMETRIC ANALYSIS VIA MORTAR METHODS 2531
Malte Woidt, Kay Sommerwerk, Matthias C. Haupt, Peter Horst

DOMAIN DECOMPOSITION SOLUTION SCHEMES FOR LARGE-SCALE IGA PROBLEMS 2541
George Stavroulakis, Dimitris Tsapetis, Manolis Papadrakakis

MS 902: INNOVATIVE NUMERICAL APPROACHES FOR MULTI-PHYSICS PROBLEMS
NUMERICAL PROPERTIES OF A DISCONTINUOUS GALERKIN FOMULATION FOR ELECTRO-THERMAL COUPLED PROBLEMS 2558
Lina Homsi, Christophe Geuzaine, Ludovic Noels

MS 903: ADVANCES IN FICTITIOUS DOMAIN METHODS FOR SOLID MECHANICS
TOWARDS LATTICE-BOLTZMANN ON DYNAMICALLY ADAPTIVE GRIDS — MINIMALLY-INVASIVE GRID EXCHANGE IN ESPRESSO 2566
Michael Lahnert, Carsten Burstedde, Christian Holm, Miriam Mehl, Georg Rempfer, Florian Weik

AUTOMATIC CONFORMAL DECOMPOSITION OF ELEMENTS CUT BY NURBS 2591
Jakob W. Steidl, Thomas-Peter Fries

A DESIGN-THROUGH-ANALYSIS APPROACH USING THE FINITE CELL METHOD 2601
Benjamin Wassermann, Tino Bog, Stefan Kollmannsberger, Ernst Rank

MS 905: DISCONTINUOUS GALERKIN METHODS: NEW TRENDS AND APPLICATIONS
DISCONTINUOUS GALERKIN COMPUTATION OF GASEOUS MIXTURE COAXIAL JETS 2614
Nicoletta Franchina, Marco Savini, Francesco Bassi

DISCONTINUOUS GALERKIN SOLUTION OF THE REYNOLDS-AVERAGED NAVIER–STOKES AND KL-KT-LOG(W) TRANSITION MODEL EQUATIONS 2632
A. Bassi, Alessandro Colombo, Antonio Ghidoni, Marco Lorini, Gianmaria Noventa

A STUDY OF MULTIGRID SMOOTHERS USED IN COMPRESSIBLE CFD BASED ON THE CONVECTION DIFFUSION EQUATION 2648
Philipp Birken, Jonathan Bull, Antony Jameson

HIGH-ORDER LINEARLY IMPLICIT TWO-STEP PEER METHODS FOR THE DISCONTINUOUS GALERKIN SOLUTION OF THE INCOMPRESSIBLE RANS EQUATIONS 2664
Francesco Carlo Massa, Gianmaria Noventa, Francesco Bassi, Alessandro Colombo, Antonio Ghidoni, Marco Lorini
OUTPUT-BASED SPACE-TIME ADAPTATION WITH NON-VARIATIONAL TIME INTEGRATION 2684
Krzysztof Fidkowski

MATRIX-FREE MODIFIED EXTENDED BACKWARD DIFFERENTIATION FORMULAE APPLIED TO THE DISCONTINUOUS GALERKIN SOLUTION OF COMPRESSIBLE UNSTEADY VISCOUS FLOWS 2697
Alessandra Nigro, Carmine De Bartolo, Andrea Crivellini, Francesco Bassi

MS 906: MATHEMATICAL ADVANCES IN ISOGEOMETRIC ANALYSIS
PRECONDITIONERS FOR ISOGEOMETRIC ANALYSIS BASED ON SOLVERS FOR SYLVESTER EQUATION 2715
Giancarlo Sangalli, Mattia Tani

MS 907: REGULARIZED ENRICHED APPROXIMATIONS AND QUADRATURE FOR DISCONTINUITIES, SINGULARITIES AND CONTINUOUS-DISCONTINUOUS TRANSITION
EFFECTIVE 3D REGULARIZED XFEM FOR PULL-OUT OF STEEL BARS IN CONCRETE, BENDING AND SHEAR TESTS ON FRP-REINFORCED CONCRETE BEAMS 2726
Elena Benvenuti, Nicola Orlando

MS 908: VERIFICATION AND VALIDATION OF STRUCTURAL MECHANICS SIMULATION MODELS
EXPERIMENTAL VALIDATION OF COMPOSITE STRUCTURES IN EXPLICIT IMPACT ANALYSIS 2734
Konstantinos Fotopoulos, George Lampeas

CODE VERIFICATION IN SHELL ANALYSIS BY THE METHOD OF MANUFACTURED SOLUTIONS 2745
Michael Gfrerer, Martin Schanz

EXPERIMENTAL VALIDATION OF A MODEL UPDATE PROCEDURE FOCUSING ON SMALL GEOMETRIC DEVIATIONS 2757
Thomas Maywald, Arnold Kühhorn, Sven Schrape

OPTIMUM DESIGN, FINITE ELEMENT MODEL UPDATING AND DYNAMIC ANALYSIS OF A FULL LAMINATED GLASS PANORAMIC CAR ELEVATOR 2774
Dimitrios Giagopoulos, Iraklis Chatziparasidis

INVESTIGATION IN THE MECHANICAL BEHAVIOUR OF LOW-VELOCITY IMPACTED CFRP PLATES UNDER DIFFERENT BOUNDARY CONDITIONS 2786
Süleyman Arslan, Zamaan Sadeghi, Athanasios Dafnis, Kai-Uwe Schröder

A NEW SHELL FINITE ELEMENT WITH DRILLING DEGREES OF FREEDOM AND ITS RELATION TO EXISTING FORMULATIONS 2803
Robert Winkler, Dimitrios Plakomytis

MS 910: HIGH ORDER CFD METHODS: CONCLUSIONS AND OUTLOOK
CURVED GRID GENERATION AND DG COMPUTATION FOR THE DLR-F11 HIGH LIFT CONFIGURATION 2843
Ralf Hartmann, Harlan Mcmorris, Tobias Leicht

BS1 - DNS OF THE TAYLOR-GREEN VORTEX AT RE = 1600 2859
Andrea Mastellone, Luigi Cutrone, Francesco Capuano

MS 911: NUMERICAL METHODS IN THE MECHANICS OF GENERALIZED CONTINUA
THE STRUCTURAL SYMMETRY WITHIN THE CONTEXT OF NONLOCAL ELASTICITY 2869
Aurora Angela Pisano, Paolo Fuschi

ISOGEOMETRIC GALERKIN METHODS FOR GRADIENT-ELASTIC BARS, BEAMS, MEMBRANES AND PLATES 2876
Jarkko Niiranen, Sergei Khakalo, Viacheslav Balobanov, Josef Kiendl, Antti H. Niemi, Bahram Hosseini, Alessandro Reali

MS 912: HIGH-ORDER METHODS, SENSITIVITY ANALYSIS AND ADAPTATION FOR THE NAVIER STOKES EQUATIONS
ADAPTIVE MESH REFINEMENT TECHNIQUES FOR HIGH-ORDER FINITE-VOLUME WENO SCHEME 2883
Harshavardhana Srinivasan, Panagiotis Tsoutsanis

IMPLEMENTATION OF A LOW-MACH NUMBER MODIFICATION FOR HIGH-ORDER FINITE-VOLUME SCHEMES FOR ARBITRARY HYBRID UNSTRUCTURED MESHES 2900
Nicholas Simmonds, Panagiotis Tsoutsanis, Adrian Gaylard

APPROPRIATE COMBINATIONS OF CONTROLLER PARAMETERS FOR UNSTEADY FLOW SIMULATIONS WITH ADAPTIVE TIME STEP CONTROL 2916
Kathrin Kozulovic, Graham Ashcroft

MS 913: HIGH-ORDER METHODS FOR POLYGONAL AND POLYHEDRAL MESHES
ADAPTIVE BEM-BASED FEM ON POLYGONAL MESHES FROM VIRTUAL ELEMENT METHODS 2930
Steffen Weißer

VEM AND TOPOLOGY OPTIMIZATION ON POLYGONAL MESHES 2941
Paola F. Antonietti, Matteo Bruggi, Simone Scacchi, Marco Verani

THE VIRTUAL ELEMENT METHOD FOR DISCRETE FRACTURE NETWORK FLOW AND TRANSPORT SIMULATIONS 2953
Matias Fernando Benedetto, Stefano Berrone, Andrea Borio, Sandra Pieraccini, Stefano Scialo

MS 916: DIRECT METHODS FOR LIMIT STATES OF MATERIALS AND STRUCTURES
COMPOSITE FEM MODELS FOR LIMIT AND SHAKEDOWN ANALYSIS 2971
Leonardo Leonetti, Giovanni Garcea, Hung Nguyen-Xuan

ON SHAKEDOWN OF ELASTIC-PLASTIC BODIES WITH TEMPERATURE DEPENDENT PROPERTIES 2983
Michael Peigney

A NUMERICAL STUDY ON THE ENDURANCE LIMIT OF PARTICULATE REINFORCED METAL MATRIX COMPOSITES (PRMMCS) USING THE DIRECT METHOD AND THE STATISTICAL LEARNING 2994
Geng Chen, Bezold Alexander, Christoph Broeckmann, Dieter Weichert

SHAKEDOWN ANALYSIS OF PLATE BENDING UNDER STOCHASTIC UNCERTAINTY BY CHANCE CONSTRAINED PROGRAMMING 3007
Ngoc Trinh Tran, Thanh Ngoc Tran, H.G. Matthies, G.E. Stavroulakis, M. Staat

LIMIT ANALYSIS LOCUS OF HIGH STRENGTH STEEL PLATES UNDER NON-QUADRATIC YIELD CRITERIA 3020
Konstantinos Nikolaou, Christos Bisbos

MS 917: MESOSCOPIC METHODS FOR COMPLEX FLUIDS AND SOFT MATTER
PARTICLE BASED SIMULATION OF FLUID FLOW IN PERIODICALLY GROOVED CHANNELS 3026
Dorothea Kasiteropoulou, Theodoros Karakasidis, Antonios Liakopoulos

MS 918: COMPUTER ALGEBRA SYSTEMS IN MODELLING STATIC AND DYNAMIC PROBLEMS IN MECHANICS OF SOLIDS
HIGH COMPUTATIONAL EFFICIENCY THROUGH GENERIC ANALYTICAL FORMULATION FOR LINEAR SOIL PRESSURE DISTRIBUTION OF RIGID SPREAD RECTANGULAR FOOTINGS 3045
John Bellos, Nikolaos P. Bakas

A REAL-SPACE MODAL ANALYSIS METHOD FOR NON-PROPORTIONAL DAMPED STRUCTURES 3065
Evgueni Stanoev

A FORCE-BASED FORMULATION FOR THE ANALYSIS OF 3-DIMENSIONAL INELASTIC STRUCTURAL FRAMES ... 3091
Theodoros Patsios, Konstantinos Spiliopoulos

THE FEATURES OF ANALYTICAL SOLUTIONS OF BOUNDARY VALUE PROBLEMS OF THE ELASTICITY THEORY FOR FINITE DOMAINS WITH ANGULAR POINTS OF A BOUNDARY 3107
Alexander Kerzhaev, Mikhail Kovalenko, Irina Menshova

MS 919: RECENT ADVANCES IN NUMERICAL SIMULATION AND ANALYSIS OF KINETIC MODELS
DISCRETE VELOCITY MODELS: A STUDY OF THE HYDRODYNAMIC LIMIT 3113
Hans Babovsky

MESO - MACRO MODELS FOR A HARD SPHERE GAS 3121
Sergey Bogomolov, Natalya Esikova, Artem Kuvshinnikov

MS 921: RECENT ADVANCES IN BOUNDARY ELEMENT METHODS
ADVANCES IN THE BOUNDARY ELEMENT METHOD IN GEOMECHANICS 3139
Gernot Beer

ENERGETIC BEM FOR THE NUMERICAL SOLUTION OF DAMPED WAVE PROPAGATION EXTERIOR PROBLEMS 3159
Alessandra Aimi, Mauro Diligenti, Chiara Guardasoni

MODELLING THE PROCESS OF SEQUENTIAL EXCAVATION WITH THE BOUNDARY ELEMENT METHOD 3168
Christian Duenser, Gernot Beer

GREEN'S FUNCTION FOR THE EVALUATION OF ANCHOR LOSSES IN MEMS 3181
Attilio Frangi, Massimiliano Cremonesi

EFFICIENT BOUNDARY ELEMENT FORMULATION OF THERMOELASTICITY 3190
Relindis Rott, Martin Schanz

A CHEBYSHEV INTERPOALTION-BASED FAST MULTIPOLE METHOD FOR POROELASTICITY 3201
Barbara Knöbl, Thomas Traub, Martin Schanz

MS 922: HIGH-ORDER METHODS FOR ELASTIC WAVES AND THEIR APPLICATION
ENERGY STABLE HIGH ORDER FINITE DIFFERENCE METHODS ON STAGGERED GRIDS: AN INITIAL INVESTIGATION 3211
Ossian O'Reilly, Tomas Lundquist, Jan Nordström

MS 923: NOVEL DISCRETIZATION METHODS – MATHEMATICAL AND MECHANICAL ASPECTS
ALTERNATIVE TO RETURN -MAPPING ALGORITH FOR COMPUTING PLASTIC STRAIN. APPLICATION TO DILATANT MATERIALS 3224
Siegfried Maiolino

DISCONTINUOUS GALERKIN METHOD WITH REDUCED INTEGRATION SCHEME FOR THE BOUNDARY TERMS IN ALMOST INCOMPRESSIBLE LINEAR ELASTICITY 3231
Hamid Reza Bayat, Stephan Wulfinghoff, Stefanie Reese

1000 OPTIMIZATION, INVERSE PROBLEMS AND CONTROL

MS 1001: STRUCTURAL AND MULTIDISCIPLINARY OPTIMIZATION
A METHOD FOR BI-DIRECTIONAL COUPLING OF STRUCTURE AND SYSTEM IN THE OPTIMISATION OF MULTI-FUNCTIONAL COMPONENTS 3239
Ajit Panesar, Ian Ashcroft, Ricky Wildman, Richard Hague

SHAPE-TOPOLOGY OPTIMIZATION FOR DESIGNING SHELL STRUCTURES 3249
Hirotaka Nakayama, Masatoshi Shimoda

ON A NASH GAME FOR TOPOLOGY OPTIMIZATION UNDER LOAD-UNCERTAINTY – FINDING THE WORST LOAD 3259
Carl-Johan Thore

ANALYSIS AND DESIGN OF REINFORCED CONCRETE STRUCTURES AS A TOPOLOGY OPTIMIZATION PROBLEM 3271
Matteo Bruggi

ADJOINT OPTIMAL CONTROL PROBLEMS FOR FLUID-STRUCTURE INTERACTION SYSTEMS 3280
Daniele Cerroni, Roberto Da Vià, Sandro Manservisi, Filippo Menghini, Luca Zaniboni

APPLICATION OF WHIRL FLUTTER OPTIMIZATION-BASED SOLUTION TO FULL-SPAN MODEL OF TWIN TURBOPROP AIRCRAFT 3293
Jiri Cecrdle

ON THE CONSIDERATION OF UNCERTAINTY IN DESIGN: OPTIMIZATION – RELIABILITY – ROBUSTNESS 3310
Nicolas Lelièvre, Pierre Beaurepaire, Cécile Mattrand, Nicolas Gayton, Abdelkader Ostmane

A NEW OPTIMIZED ANISOTROPIC MOVING LEAST SQUARES SURROGATE MODEL WITH MAXIMIZED PROGNOSIS 3321
Kevin Cremanns, Dirk Roos

LATIN HYPERCUBE SAMPLING BASED ON ADAPTIVE ORTHOGONAL DECOMPOSITION 3333
Dirk Roos

A BENCHMARK OF CONTEMPORARY METAMODELING ALGORITHMS 3344
Can Bogoclu, Dirk Roos

TOPOLOGY OPTIMIZATION USING A KRIGING-ASSISTED GENETIC ALGORITHM WITH A NOVEL LEVEL SET REPRESENTATION APPROACH 3361
Mitsuo Yoshimura, Koji Shimoyama, Takashi Misaka, Shigeru Obayashi

TOPOLOGY OPTIMIZATION OF CONTACT PROBLEMS BASED ON ALLEN CAHN APPROACH 3377
Andrzej Myslinski

ON THE OPTIMAL DESIGN OF CABLE-STAYED BRIDGES 3386
Gerardo Carpentieri, Mariano Modano, Francesco Fabbrocino, Luciano Feo, Fernando Fraternali

A MIXED-INTEGER LINEAR PROGRAMMING APPROACH FOR GLOBAL DISCRETE SIZE OPTIMIZATION OF FRAME STRUCTURES 3395
Roxane Van Mellaert, Kristo Mela, Teemu Tiainen, Markku Heinisuo, Geert Lombaert, Mattias Schevenels

A SUPER-STRUCTURE BASED OPTIMISATION APPROACH FOR BUILDING SPATIAL DESIGNS 3409
Koen Van Der Blom, Sjonnie Boonstra, Hèrm Hofmeyer, Michael Emmerich

BENDING STIFFNESS OF A MULTILAYERED PLATE 3423
Petr Tovstik, Tatiana M. Tovstik

EFFICIENT SIZING OF STRUCTURES UNDER STRESS CONSTRAINTS 3436
Zhi Hong, Mostafa Abdalla

TOPOLOGY OPTIMIZATION OF WAVE BARRIERS FOR RAILWAY INDUCED VIBRATIONS IN BUILDINGS 3461
Cédric Van Hoorickx, Mattias Schevenels, Geert Lombaert

MS 1002: EVOLUTIONARY ALGORITHMS AND METAHEURISTICS IN CIVIL ENGINEERING AND CONSTRUCTION MANAGEMENT

MULTIOBJECTIVE OPTIMIZATION USING GENETIC ALGORITHMS IN TIME-COST CONSTRUCTION PROJECT SCHEDULING PROBLEM 3473
Jorge Magalhães-Mendes

APPLICATION OF GENETIC ALGORITHMS TO STRUTTED SHEET PILE WALL DESIGN OPTIMIZATION 3485
Mohamed Eid, Remon Isaac

NASH EVOLUTIONARY ALGORITHMS: TESTING PROBLEM SIZE IN RECONSTRUCTION PROBLEMS IN FRAME STRUCTURES 3493
David Greiner, Jacques Periaux, Jose M. Emperador, Blas Galván, Gabriel Winter

OPTIMUM DESIGN OF REINFORCED CONCRETE FRAMES ACCORDING TO EC8 AND MC2010 WITH GENETIC ALGORITHMS 3505
Panagiotis Mergos

CALIBRATION OF THE NUMERICAL MODEL OF A STAND IN DRAGÃO STADIUM BASED ON GENETIC ALGORITHMS 3519
Jorge Leite, Diogo Ribeiro, Hugo Marques, Rui Calçada

MS 1003: ADVANCES IN DESIGN OPTIMIZATION OF STRUCTURES AND MATERIALS
DESIGN OF ISOTROPIC MICROSTRUCTURES VIA A TWO-SCALE APPROACH 3534
Alexis Faure, Georgios Michailidis, Rafael Estevez, Guillaume Parry, Grégoire Allaire

SHAPE OPTIMIZATION OF GRAPHENE SHEETS FOR MAXIMUM FUNDAMENTAL FREQUENCY 3546
Jin-Xing Shi, Masatoshi Shimoda

TOPOLOGY OPTIMIZATION WITH AN IMPLICIT FUNCTION AND PARAMETERIZED CUTTING SURFACE 3553
Peter Dunning

TOPOLOGY OPTIMIZATION OF PERIODIC STRUCTURES FOR COUPLED ACOUSTIC-STRUCTURE SYSTEMS 3565
William Martins Vicente, Renato Picelli, Renato Pavanello, Yi Min Xie

BESO APPROACH TO TOPOLOGY OPTIMIZATION OF GAN PHONONIC CRYSTALS 3583
Luca D'Alessandro, Bichoy Bahr, Luca Daniel, Dana Weinstein, Raffaele Ardito

MS 1004: AERODYNAMIC STRATEGIES FOR THE GLOBAL OPTIMIZATION OF FLYING CONFIGURATIONS IN SUPERSONIC FLOW
LOW BOOM / LOW DRAG SMALL SIZE SUPERSONIC AIRCRAFT DESIGN 3596
Atsushi Ueno, Yasushi Watanabe, Itham Salah El Din, Richard Grenon, Gerald Carrier

DEVELOPMENT OF PHYSICAL METHODS OF THE SUPERSONIC AIRPLANE NEAR-FIELD INVESTIGATION AIMED AT THE SONIC BOOM MINIMIZATION 3612
Sergey Chernyshev, Alexander Ivanov, Andrey Kiselev, Vladimir Mosharov, Leonid Teperin

GLOBAL OPTIMIZATION OF THE SHAPE OF AN AEROSPACE VEHICLE, VIA ITERATIVE OPTIMUM-OPTIMORUM STRATEGY 3621
A. Nastase

MS 1005: MONITORING AND CONTROL OF STRUCTURES
SHAPE IDENTIFICATION ANALYSIS OF CAVITY IN RESIN STRUCTURE BASED ON THERMAL NONDESTRUCTIVE TESTING METHOD 3630
Kotaro Maruoka, Takahiko Kurahashi, Tetsuro Iyama

DESIGN AND INSTALLATION OF A PERMANENT MONITORING SYSTEM FOR PALAZZO LOMBARDIA IN MILANO, ITALY 3640
Marta Berardengo, Alfredo Cigada, Stefano Manzoni, Marcello Vanali

MS 1006: PARAMETER IDENTIFICATION IN SOLID MECHANICS
AN INVERSE PROBLEM STRATEGY BASED ON FORWARD MODEL EVALUATIONS: GRADIENT-BASED OPTIMIZATION WITHOUT ADJOINT SOLVES 3652
Miguel Aguilo

COHESIVE ZONE MODEL IDENTIFICATION WITH DCB TEST PARAMETER ESTIMATION SENSITIVITY 3676
Racine Ly, Julien Jumel, Martin Shanahan, Florian Lavelle

MS 1007: ADDITIVE MANUFACTURING AND OPTIMIZATION
TOPOLOGY OPTIMIZATION FOR ADDITIVE MANUFACTURING WITH CONTROLLABLE SUPPORT STRUCTURE COSTS 3689
Matthijs Langelaar

STUDY OF TOPOLOGY OPTIMIZATION PARAMETERS AND SCAFFOLD STRUCTURES IN ADDITIVE MANUFACTURING 3700
Alain Garaigordobil, Ruben Ansola, Estrella Veguería

FINITE ELEMENT APPROXIMATION OF A TIME-DEPENDENT TOPOLOGY OPTIMIZATION PROBLEM 3711
Matteo Bruggi, Nicola Parolini, Francesco Regazzoni, Marco Verani

DESIGNING MANUFACTURABLE VISCOELASTIC DEVICES USING A TOPOLOGY OPTIMIZATION APPROACH WITHIN A TRULY-MIXED FEM FRAMEWORK 3724
Paolo Venini, Marco Pingaro, Carlo Cinquini

ISO-XFEM FOR TOPOLOGY OPTIMIZATION OF STRUCTURES UNDER MULTIPLE LOAD CASES AND ACCELERATION LOADING 3739
Meisam Abdi, Ian Ashcroft, Ricky Wildman

MS 1008: ULTRASONIC GUIDED WAVES TESTING AND MONITORING
LAMINATE ELEMENT METHOD AND ITS APPLICATION TO THE STUDY OF GUIDED WAVE RESONANCE PHENOMENA IN LAYERED ELASTIC STRUCTURES WITH DEFECTS 3753
Evgeny Glushkov, Natalia Glushkova, Artem Eremin, Rolf Lammering

INVESTIGATIONS OF MECHANICAL GUIDED WAVES PROPAGATION IN PIPES REPAIRED LOCALLY BY COMPOSITE PATCHES 3761
Slah Yaacoubi, Weina Ke, F. Dahmene, M. El Mountassir

MS 1009: ADJOINT METHODS FOR STEADY & UNSTEADY OPTIMIZATION

TWO-STEPS SHAPE OPTIMIZATION ALGORITHM IMPROVING HYDRODYNAMICS STABILITY 3770
Takashi Nakazawa

ON THE CORRECT APPLICATION OF AD CHECKPOINTING TO ADJOINT MPI-PARALLEL PROGRAMS 3778
Ala Taftaf, Laurent Hascoët

SPHERICITY: MESH OPTIMIZATION FOR ARBITRARY ELEMENT TOPOLOGY 3797
Pavlos Alexias, Eugene De Villiers

LINKING PARAMETRIC CAD WITH ADJOINT SURFACE SENSITIVITIES 3812
Ilias Vasilopoulos, Dheeraj Agarwal, Marcus Meyer, Trevor T. Robinson, Cecil G. Armstrong

A CODE-COUPLING APPROACH TO THE IMPLEMENTATION OF DISCRETE ADJOINT SOLVERS BASED ON AUTOMATIC DIFFERENTIATION 3828
Jan Backhaus, Anna Engels-Putzka, Christian Frey

CAD KERNEL AND GRID GENERATION ALGORITHMIC DIFFERENTIATION FOR TURBOMACHINERY ADJOINT OPTIMIZATION 3843
Ismael Sanchez Torreguitart, Tom Verstraete, Lasse Mueller

ALTERNATIVE SOLUTION ALGORITHMS FOR PRIMAL AND ADJOINT INCOMPRESSIBLE NAVIER-STOKES 3858
Mattia Oriani, Guillaume Pierrot

DEFROSTER NOZZLE SHAPE OPTIMIZATION USING THE CONTINUOUS ADJOINT METHOD 3883
Lefki Germanou, Evangelos Papoutsis-Kiachagias, Antoine Delacroix, Kyriakos Giannakoglou

SENSITIVITY ANALYSIS FOR FORCED RESPONSE IN TURBOMACHINERY USING AN ADJOINT HARMONIC BALANCE METHOD 3893
Anna Engels-Putzka, Christian Frey

A TWO-LEVEL HYBRID APPROACH FOR OPTIMAL ACTIVE FLOW CONTROL ON A THREE-ELEMENT AIRFOIL 3912
Anil Nemili, Emre Özkaya, Nicolas R. Gauger, Felix Kramer, Frank Thiele

TRANSITION FROM 2D CONTINUOUS ADJOINT LEVEL SET TOPOLOGY TO SHAPE OPTIMIZATION 3928
J.R.L. Koch, E.M. Papoutsis-Kiachagias, K.C. Giannakoglou

AERODYNAMIC SHAPE OPTIMIZATION USING THE TRUNCATED NEWTON METHOD AND CONTINUOUS ADJOINT 3942
Mehdi Ghavami Nejad, Evangelos M. Papoutsis-Kiachagias, Kyriakos C. Giannakoglou

UNSTEADY ADJOINT TO THE CUT-CELL METHOD USING MESH ADAPTATION ON GPU'S 3957
Konstantinos Samouchos, Stergios Katsanoulis, Kyriakos Giannakoglou

SHAPE OPTIMIZATION OF TURBOMACHINERY ROWS USING A PARAMETRIC BLADE MODELLER AND THE CONTINUOUS ADJOINT METHOD RUNNING ON GPUS 3972
K.T. Tsiakas, F. Gagliardi, X.S. Trompoukis, K.C. Giannakoglou

GEOMETRIC CONTINUITY CONSTRAINTS FOR ADJACENT NURBS PATCHES IN SHAPE OPTIMISATION 3985
Xingchen Zhang, Yang Wang, Mateusz Gugala, Jens-Dominik Mueller

TOWARDS MULTIDISCIPLINARY ADJOINT OPTIMIZATION OF TURBOMACHINERY COMPONENTS 3999
Marc Schwalbach, Tom Verstraete

TOWARDS AN OUTPUT-BASED RE-MESHING FOR TURBOMACHINERY APPLICATIONS 4011
Mateusz Gugala, Marcus Meyer, Jens-Dominik Mueller

OPTIMISATION OF A U-BEND USING A CAD-BASED ADJOINT METHOD WITH DIFFERENTIATED CAD KERNEL 4023
Salvatore Auriemma, Mladen Banovic, Orest Mykhaskiv, Herve Legrand, Jens-Dominik Mueller, Tom Verstraete, Andrea Walther

A CONTINUOUS ADJOINT APPROACH FOR VEHICLE INTERIOR NOISE REDUCTION 4035
Christos Kapellos, Michael Hartmann

A MIXED OPERATOR OVERLOADING AND SOURCE TRANSFORMATION APPROACH FOR ADJOINT CFD COMPUTATION 4047
Zahrasadat Dastouri, Sinan Gezgin, Uwe Naumann

MS 1010: INVERSE PROBLEMS, DESIGN AND OPTIMIZATION

EFFICIENT CALIBRATION OF DISCRETE ELEMENT MATERIAL MODEL PARAMETERS USING LATIN HYPERCUBE SAMPLING AND KRIGING 4061
Michael Rackl, Carolin D. Görnig, Kevin J. Hanley, Willibald A. Günthner

UPDATING PRIOR PARAMETERS BASED ON LIKELIHOOD FUNCTION-BAYESIAN METHOD FOR PARAMETER ESTIMATION AT HIGH MEASUREMENT UNCERTAINTY 4073
Sankaran Somasundharam, K Srinivas Reddy

MS 1011: SURROGATE-ASSISTED EVOLUTIONARY ALGORITHMS IN AERODYNAMIC DESIGN/OPTIMIZATION

CONSTRAINED MULTI-POINT AERODYNAMIC SHAPE OPTIMIZATION OF THE VISCOUS DPW WING THROUGH EVOLUTIONARY PROGRAMMING AND SUPPORT VECTOR MACHINES 4082
Daniel González Juárez, Esther Andrés Pérez, Mario Jaime Martin Burgos

SURROGATE-BASED GLOBAL OPTIMIZATION OF A CYLINDER BY THE USE OF EVOLUTIONARY ALGORITHMS, SUPPORT VECTOR MACHINES AND NON UNIFORM B-SPLINES 4092
Eesther Andrés-Pérez, Daniel González-Juárez, Mario J. Martin-Burgos, Leopoldo Carro-Calvo, Sancho Salcedo-Sanz

PERFORMANCE COMPARISON OF KRIGING AND SVR SURROGATE MODELS APPLIED TO THE OBJECTIVE FUNCTION PREDICTION WITHIN AERODYNAMIC SHAPE OPTIMIZATION 4104
Daniel Viúdez-Moreiras, Esther Andrés-Pérez, Daniel González-Juárez, Mario J. Martin Burgos

HEAT TRANSFER OPTIMIZATION OF A RIBBED SURFACE USING SURROGATE-ASSISTED GENETIC ALGORITHMS 4117
Panagiotis Tsirikoglou, Ghader Ghorbaniasl, Simon Abraham, Chris Lacor

MULTI-FIDELITY EXTENSION TO NON-INTRUSIVE PROPER ORTHOGONAL DECOMPOSITION BASED SURROGATES 4129
Tariq Benamara, Piotr Breitkopf, I. Lepot, Caroline Sainvitu

EVOLUTIONARY AERODYNAMIC SHAPE OPTIMIZATION THROUGH THE RBF4AERO PLATFORM 4146
Dimitrios Kapsoulis, Varvara Asouti, Kyriakos Giannakoglou, Stefano Porziani, Emiliano Costa, Corrado Groth, Ubaldo Cella, Marco Evangelos Biancolini

THE RBF4AERO BENCHMARK TECHNOLOGY PLATFORM 4156
Massimo Bernaschi, Alessandro Sabellico, Giorgio Urso, Emiliano Costa, Stefano Porziani, Fabrizio Lagasco, Corrado Groth, Ubaldo Cella, Marco Evangelos Biancolini, Dimitrios H.Kapsoulis, Varvara G. Asouti, Kyriakos C. Giannakoglou

MS 1013: SOLUTION OF LARGE-SCALE INVERSE PROBLEMS
DATA ASSIMILATION USING MRI DATA 4164
Geir Nævdal, Ove Sævareid, Rolf J. Lorentzen

MS 1014: DESIGN OPTIMIZATION AND INVERSE PROBLEMS FOR WAVE PROPAGATION PROBLEMS
FULL-WAVEFORM INVERSION FOR THE RECONSTRUCTION OF ELECTROMAGNETIC PROPERTY PROFILES USING PLANE MICROWAVES 4178
Namho Joh, Jun Won Kang

1100 REDUCTION METHODS

MS 1101: REDUCED BASIS, POD AND PGD MODEL ORDER REDUCTION TECHNIQUES
REDUCED BASIS METHOD FOR POISSON-BOLTZMANN EQUATION 4187
Cleophas Kweyu, Martin Hess, Lihong Feng, Matthias Stein, Peter Benner

AN ITERATIVE MODEL ORDER REDUCTION SCHEME FOR A SPECIAL CLASS OF BILINEAR DESCRIPTOR SYSTEMS APPEARING IN CONSTRAINT CIRCUIT SIMULATION 4196
Pawan Goyal, Peter Benner

ELECTRICAL FIELDS SIMULATION IN HETEROGENOUS DOMAINS USING THE PROPER GENERALIZED DECOMPOSITION 4213
Chady Ghnatios, Francisco Chinesta, Anais Barasinski

MODEL ORDER REDUCTION IN STRUCTURAL DYNAMICS 4221
Raul Rodriguez Sanchez, Martin Buchschmid, Gerhard Müller

TOWARDS A NON-INTRUSIVE PROPER GENERALIZED DECOMPOSITION SCHEME FOR MODEL ORDER REDUCTION 4248
Xi Zou, Pedro Díez, Michele Conti, Ferdinando Auricchio

A PGD-BASED TIME SPACE DECOMPOSITION FOR THE UNSTEADY NAVIER-STOKES EQUATIONS APPLIED TO INCOMPRESSIBLE FLOWS 4263
Michel Visonneau

MS 1103: MATHEMATICAL SURROGATE MODELLING IN ELECTROMAGNETICS

RATIONAL MODELING OF MULTIVARIATE MULTI-FIDELITY DATA 4284
Elizabeth Rita Samuel, Dirk Deschrijver, Luc Knockaert, Tom Dhaene, Annie Cuyt

ON THE USE OF S-PARAMETER TRANSFORMATIONS TO IMPROVE SURROGATE MODEL BEHAVIOUR OF MULTIPORT NETWORKS 4295
Petrie Meyer

RECENT ADVANCES IN SURROGATE MODELLING OF REFLECTOR ANTENNA SYSTEMS 4302
Dirk De Villiers

A SHORT-TIME PRONY METHOD FOR THE DETECTION OF TRANSIENTS 4313
Annie Cuyt, Wen-Shin Lee, Min-Nan Tsai

SURROGATE MODELING OF ANTENNA RADIATION CHARACTERISTICS BY GAUSSIAN PROCESSES 4322
Jan Pieter Jacobs, Dirk De Villiers

ACCELERATION OF MESH-BASED PHYSICAL OPTICS FOR ELECTROMAGNETIC SCATTERING ANALYSIS 4330
Dao P. Xiang, Matthys M. Botha

DESIGN OF AXIALLY SYMMETRIC POWER COMBINERS USING SURROGATE BASED OPTIMIZATION 4339
Ryno Beyers, Dirk De Villiers

1200 STRUCTURAL STABILITY AND DYNAMICS

MS 1201: COMPUTATIONAL STRUCTURAL DYNAMICS

FLEXURAL VIBRATION ANALYSIS OF GRAPHENE NANOPLATELETS REINFORCED NANOCOMPOSITE BEAMS 4351
Jie Yang, Chuang Feng

DYNAMIC RESPONSE OF REAL OFFSHORE WIND TURBINES ON MONOPILES IN STRATIFIED SEABED 4361
Guillermo M. Álamo, Juan J. Aznárez, Luis A. Padrón, Alejandro E. Martínez-Castro, Rafael Gallego, Orlando Maeso

DYNAMIC ANALYSIS OF SPECIAL CARS ON UNEVEN ROADS 4377
Szymon Tengler

STRUCTURED SPATIAL DISCRETIZATION OF DYNAMICAL SYSTEMS 4398
Marko Jokic, Andrej Jokic, Bruno Dogancic

WAVE PROPAGATION IN THIN PRETWISTED ANISOTROPIC STRIPS 4406
Maloth Thirupathi, Mira Mitra, P J Guruprasad

NUMERICAL ROUTINE FOR DYNAMIC ANALYSIS OF TRANSMISSION LINES GUYED TOWERS SUBMITTED TO BROKEN CABLE 4416
Thiago Brazeiro Carlos, João Kaminski Jr.

INVARIANCE OF EIGENFREQUENCIES AND EIGENMODES UNDER GEOMETRIC TRANSFORMATION IN ELONGATED ELASTIC STRUCTURES 4436
Maryam Morvaridi, Michele Brun

MS 1202: ADVANCED BEAM MODELS

A COMPUTATIONALLY EFFECTIVE FORMULATION TO FINITE ROTATIONS - SMALL STRAINS DESCRIPTION OF BEAM ELEMENTS 4445
Salvatore Lopez

A COROTATIONAL FINITE ELEMENT TO MODEL BENDING VIBRATIONS OF METALLIC STRANDS 4455
Francesco Foti

3D BEAM-COLUMN FINITE ELEMENT UNDER NON-UNIFORM SHEAR STRESS DISTRIBUTION DUE TO SHEAR AND TORSION 4467
Paolo Di Re, Daniela Addessi, Filip C. Filippou

SINGULARLY PERTURBED PROBLEMS IN MECHANICS (SOME FUNDAMENTAL ASPECTS OF DESIGNING AND COMPUTING) 4481
Lyudmila K. Kuzmina

A DEGENERATE-CONTINUUM BASED TIMOSHENKO BEAM APPROACH FOR THE AEROELASTIC ANALYSIS OF THE WIND TURBINE BLADES 4488
Anthoula Panteli, Dimitris Manolas, Konstantinos Spiliopoulos

DYNAMICS OF WIND TURBINE BLADES USING A GEOMETRICALLY-EXACT BEAM FORMULATION 4503
Celso Faccio Junior, Alfredo Gay Neto

EXTENDED FORMULA FOR THE CRITICAL VELOCITY OF A LOAD MOVING ON A BEAM SUPPORTED BY A FINITE DEPTH FOUNDATION 4520
Zuzana Dimitrovová

FIRST-ORDER DISPLACEMENT-BASED ZIGZAG THEORIES FOR COMPOSITE LAMINATES AND SANDWICH STRUCTURES: A REVIEW 4528
Marco Di Sciuva

MS 1203: THE MODELS AND INVESTIGATIONS METHODS OF DYNAMICS OF THE SOLIDS SYSTEMS WITH DRY FRICTION

VIRTUAL TESTBENCH FOR THE OMNI WHEEL DYNAMICS SIMULATION: NEW CONTACT TRACKING ALGORITHM 4553
Ivan Kosenko, Sergey Stepanov, Kirill Gerasimov, Mikhail Stavrovskiy

MODEL AND INVESTIGATION OF DYNAMICS OF SOLID SYSTEM WITH TWO MASSIVE ECCENTRICS ON A ROUGH PLANE 4572
Sergey Semendyaev, Alexey Tsyganov

NON-REGULAR VEHICLE DYNAMICS. APPLICATION TO COLLISION 4584
Ahmed Bouzar Essaidi, Bachir Menkouz, Moussa Haddad, Taha Chettibi

TESTING, SIMULATING AND UNDERSTANDING UNDER-PLATFORM DAMPER DYNAMICS 4599
Chiara Gastaldi, Muzio M. Gola

MS 1204: NONLINEAR DYNAMICS OF ROTATING STRUCTURES
A NEW DYNAMIC SUBSTRUCTURING METHOD FOR NONLINEAR AND DISSIPATIVE SYSTEMS 4611
Colas Joannin, Benjamin Chouvion, Fabrice Thouverez

SOME REMARKS ON TIME INTEGRATION OF 3D ROTOR-STATOR ASSEMBLY 4624
Benoit Prabel

DELAYED FEEDBACK CONTROL METHOD FOR CALCULATING SPACE-TIME PERIODIC SOLUTIONS OF VISCOELASTIC PROBLEMS 4638
Ustim Khristenko, Patrick Le Tallec

COMBINING FINITE ELEMENT ANALYSIS AND ANALYTICAL MODELLING FOR EFFICIENT SIMULATIONS OF NON-LINEAR GEAR DYNAMICS 4655
Shadi Sweiki, Jakub Korta, Antonio Palermo, Rocco Adduci, Domenico Mundo

DAMPING INDUCED BY DRY FRICTION: ANALYSES AND EXPERIMENTS FOR MODELING IMPROVEMENT 4664
Marc-André Douville, Béatrice Faverjon, Eric Chatelet, Georges Jacquet-Richardet

MS 1206: ADVANCES IN NUMERICAL METHODS FOR LINEAR AND NON-LINEAR DYNAMICS AND WAVE PROPAGATION
RAYLEIGH METHOD APPLIED TO A 46-M-HIGH CONCRETE MAST 4674
Alexandre De M. Wahrhaftig, Reyolando M. L. R. F. Brasil

A DISPERSION MINIMIZED MIMETIC METHOD FOR A COLD PLASMA MODEL 4682
Vrushali A. Bokil, Vitaliy Gyrya, Duncan A. Mcgregor

A VOLUME-AGGLOMERATION MULTIRATE TIME ADVANCING APPROACH 4703
Emmanuelle Itam, Stephen Wornom, Bruno Koobus, Alain Dervieux

THERMOMECHANICAL NUMERICAL SIMULATION OF IMPACTS ON ELASTIC-PLASTIC SOLIDS WITH THE FINITE VOLUME METHOD 4717
Thomas Heuzé

MS 1207: COMPUTATIONAL SIMULATION OF SMART STRUCTURES AND MATERIALS
TRANSVERSE WAVE PROPAGATION IN A ONE-DIMENSIONAL STRUCTURE COUPLED TO ITS ELECTRICAL ANALOGUE: COMPARISON OF TRANSFER MATRIX MODELS 4728
Boris Lossouarn, Mathieu Aucejo, Jean-François Deü

MS 1208: BIFURCATIONS AND STABILITY
BUCKLING OF ANNULAR PLATE JOINT WITH CIRCULAR BEAM 4741
Sergei Filippov, Maria Boyarskaya

THE CURL AND FLUTING OF PAPER: THE EFFECT OF ELASTO-PLASTICITY 4752
Anna-Leena Erkkilä, Teemu Leppänen, Tero Tuovinen

STABILITY OF STEEL STRUCTURES WITH CLEARANCES AND IMPERFECTION 4770
Katarzyna Rzeszut, Andrzej Garstecki

MS 1209: DYNAMIC ANALYSIS OF BEAMS UNDER MOVING VEHICLES: APPLICATION TO RAILWAY TRACK MODELLING, DESIGN AND REHABILITATION
NUMERICAL ANALYSIS FOR THE DYNAMIC RESPONSE CHARACTERISTICS OF THE PRESTRESSED CONCRETE SLEEPER 4780
Tsutomu Watanabe, Kodai Matsuoka, Shintaro Minoura

NUMERICAL DYNAMIC ANALYSIS OF BEAMS ON NONLINEAR ELASTIC FOUNDATIONS UNDER HARMONIC MOVING LOAD 4794
Diego Froio, Roberto Moioli, Egidio Rizzi

MS 1210: ADVANCES IN MODELING AND ANALYSIS OF FGM STRUCTURES
ELASTOSTATIC AND MODAL AND BUCKLING ANALYSIS OF SPATIAL FGM BEAM STRUCTURES 4810
Justin Murin, Mehdi Aminbaghai, Juraj Hrabovsky, Vladimir Kutis, Juraj Paulech, Stephan Kugler

THERMO-ELASTICITY IN SHELL STRUCTURES MADE OF FUNCTIONALLY GRADED MATERIALS 4838
Stephan Kugler, Peter Fotiu, Justin Murin

FINITE BEAM ELEMENT WITH PIEZOELECTRIC LAYERS AND FUNCTIONALLY GRADED MATERIAL OF CORE 4863
Vladimír Kutiš, Justín Murín, Juraj Paulech, Juraj Hrabovský, Roman Gogola, Jakub Jakubec

NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS WITH NON-CONSTANT COEFFICIENTS 4878
Juraj Hrabovský, Justín Murín, Mehdi Aminbaghai, Vladimír Kutiš, Juraj Paulech

TWO-WAY COUPLED ELECTRO-THERMAL ANALYSIS OF FGM SYSTEM CALCULATED BY THE NEW LINK FINITE ELEMENT 4888
Juraj Paulech, Vladimír Kutiš, Juraj Hrabovský, Justín Murín, Roman Gogola, Tibor Sedlár, Gabriel Gálik

MS 1211: COMPUTATIONAL STRATEGIES FOR STRUCTURAL ROBUSTNESS ASSESSMENT
PROGRESSIVE COLLAPSE FRAGILITY MODELS OF RC FRAMED BUILDINGS BASED ON PUSHDOWN ANALYSIS 4900
Emanuele Brunesi, Roberto Nascimbene, Fulvio Parisi

MECHANICAL PROPERTIES OF S355 UNDER EXTREME COUPLED EFFECT OF HIGH TEMPERATURES AND HIGH STRAIN RATES 4920
Daniele Forni, Bernardino Chiaia, Ezio Cadoni

IMPLEMENTATION OF PROGRESSIVE DAMAGE IN FINITE ELEMENT CODES FOR THE ASSESSMENT OF ROBUSNTESS 4928
Domenico Asprone, Bernardino Chiaia, Valerio De Biagi, Gaetano Manfredi, Fulvio Parisi

SEISMIC RESIDUAL CAPACITY ASSESSMENT OF FRAMED STRUCTURES DAMAGED BY EXCEPTIONAL ACTIONS 4942
Antonio Formisano, Giuseppe Iazzetta, Giuseppe Marino, Francesco Fabbrocino, Raffaele Landolfo

MS 1212: DYNAMICS AND SEISMIC RESPONSE OF ROCKING AND SELF-CENTERING STRUCTURES
ROCKING RESPONSE OF MASONRY BLOCK STRUCTURES USING MATHEMATICAL PROGRAMMING 4959
Francesco Portioli, Lucrezia Cascini, Raffaele Landolfo

CONTROLLED ROCKING, DISSIPATIVE CONTROLLED ROCKING AND MULTI-HIERARCHICAL ACTIVATION: NUMERICAL ANALYSIS AND EXPERIMENTAL TESTING 4969
Royce Liu, Alessandro Palermo

THE ROLE OF THE PRESTRESSED TENDONS ON THE SEISMIC PERFORMANCE OF HYBRID ROCKING BRIDGE BENTS 4983
Anastasios I. Giouvanidis, Elias G. Dimitrakopoulos

MODELLING CONTACT IN ROCKING STRUCTURES WITH A NONSMOOTH DYNAMICS APPROACH 4997
Anastasios I. Giouvanidis, Elias G. Dimitrakopoulos

A SIMPLE ANALYTICAL MODEL FOR THE ROCKING PREWEC SYSTEM 5010
Dimitrios Kalliontzis, Sri Sritharan

AN ANALYTICAL MODEL FOR DYNAMIC RESPONSE OF AN ELASTIC SDOF SYSTEM FIXED ON TOP OF A ROCKING SINGLE-STORY FRAME STRUCTURE: EXPERIMENTAL VALIDATION 5023
Jonas A. Bachmann, Christoph Jost, Quentin Studemann, Michalis F. Vassiliou, Bozidar Stojadinovic

SMOOTH-ROCKING OSCILLATOR UNDER NATURAL ACCELEROGRAMS 5063
Blerta Lipo, Gianmarco De Felice

EXPERIMENTAL INVESTIGATION OF THE SEISMIC RESPONSE OF A COLUMN ROCKING AND ROLLING ON A CONCAVE BASE 5074
Jonas A. Bachmann, Patrick Blöchlinger, Matthias Wellauer, Michalis F. Vassiliou, Bozidar Stojadinovic

SEISMIC ASSESSMENT OF MASONRY ROCKING COLUMNS AND FRAMES UNDER GROUND MOTION EXCITATIONS 5098
Ioannis Kavvadias, Lazaros Vasiliadis

PARAMETRIC INVESTIGATION OF THE DYNAMIC RESPONSE OF RIGID BLOCKS SUBJECTED TO SYNTHETIC NEAR-SOURCE GROUND MOTION RECORDS 5113
Michalis Fragiadakis, Ioannis Psycharis, Yenan Cao, George P. Mavroeidis

A MACRO-ELEMENT FORMULATION FOR ROCKING FLEXIBLE BODIES WITH A DEFORMABLE BASE 5129
Evangelos Avgenakis, Ioannis N. Psycharis

MS 1213: INNOVATIVE STRUCTURAL SYSTEMS FOR SEISMIC RESISTANT BUILDINGS

NONLINEAR ANALYSIS AND EXPERIMENTAL BEHAVIOUR OF AN INNOVATIVE STEEL FRAME WITH REINFORCED CONCRETE INFILL WALLS 5142
Andrea Dall'Asta, Graziano Leoni, Francesco Morelli, Walter Salvatore, Alessandro Zona

MODAL PROPERTIES AND SEISMIC RESPONSE OF EXISTING BUILDING RETROFITTED BY EXTERNAL BRACINGS WITH VISCOUS DAMPERS 5156
Laura Gioiella, Enrico Tubaldi, Fabrizio Gara, Andrea Dall'Asta

MS 1214: HISTORIC MASONRY STRUCTURES: MODELLING, ASSESSMENT & RETROFIT

SIMPLE CLOSED FORM HOMOGENIZATION MODEL FOR THE NON LINEAR STATIC AND DYNAMIC ANALYSIS OF RUNNING BOND MASONRY WALLS IN AND OUT OF PLANE LOADED 5172
Elisa Bertolesi, Gabriele Milani

CONSTRUCTION PHASES ANALYSIS OF UNREINFORCED MASONRY BUILDINGS THROUGH EQUIVALENT FRAME MODEL 5190
Francesco Pugi, Alessio Francioso, Giacomo Sevieri

A COMBINED EXPERIMENTAL AND NUMERICAL STUDY OF THE PULL-OUT MECHANISM OF THREADED TITANIUM BARS EMBEDDED IN MARBLE BLOCKS 5207
Stavros K. Kourkoulis, Aikaterini Marinelli, Ioanna Dakanali

FINITE-DISCRETE ELEMENT MODELLING OF MASONRY INFILL WALLS SUBJECTED TO OUT-OF-PLANE LOADS 5219
Laura Liberatore, Marta Bruno, Omar Al Shawa, Monica Pasca, Luigi Sorrentino

PALAZZO LA SAPIENZA IN PISA: STRUCTURAL ASSESSMENT AND RETROFIT OF AN HISTORICAL MASONRY BUILDING IN ITALY 5230
Silvia Caprili, Federico Mangini, Nicola Mussini, Walter Salvatore

THREE-DIMENSIONAL NONLINEAR BEHAVIOUR OF MASONRY WALLS MODELLED WITH DISCRETE ELEMENTS 5248
Daniele Baraldi, Antonella Cecchi

NUMERICAL VALIDATION OF EQUIVALENT-FRAME MODELS FOR URM WALLS 5262
Rossella Siano, Guido Camata, Vincenzo Sepe, Enrico Spacone, Pere Roca, Luca Pelà

SEISMIC ASSESSMENT OF HISTORICAL MASONRY TOWERS IN THE NORTH-EAST REGION OF ITALY 5274
Marco Valente, Gabriele Milani

NATIVITY CHURCH IN BETHLEHEM: FULL 3D NON-LINEAR FE APPROACH FOR STRUCTURAL DAMAGE PREDICTION 5287
Gabriele Milani, Marco Valente, Claudio Alessandri

DYNAMIC BEHAVIOUR AND EARTHQUAKE PERFORMANCE OF GREEK BASILICA CHURCHES WITH FOUNDATION DEFORMABILITY 5301
George Manos, Evaggelos Kozikopoulos, Lambros Kotoulas

PROBLEMS RELATED TO THE USE OF FIBER REINFORCED CEMENTITIOUS MATERIALS AS STRENGTHENING OF MASONRY MEMBERS 5314
Daniela Sinicropi, Antonio Borri, Marco Corradi, Michele Paradiso

FAST KINEMATIC LIMIT ANALYSIS OF FRP REINFORCED MASONRY VAULTS THROUGH A NEW GENETIC ALGORITHM NURBS-BASED APPROACH 5326
Andrea Chiozzi, Gabriele Milani, Antonio Tralli

THE ROLE OF RESTORATION MORTARS IN THE EARTHQUAKE PROTECTION OF KAISARIANI MONASTERY 5340
Antonia Moropoulou, Maria Apostolopoulou, Petros Moundoulas, Eleni Aggelakopoulou, Louiza Siouta, Asterios Bakolas, Maria Douvika, Panagiotis Karakitsios, Panagiotis G. Asteris

FIRST RESULTS OF THE VIBRATION-BASED STRUCTURAL HEALTH MONITORING OF A MASONRY DOME 5359
Nicola Cavalagli, Gabriele Comanducci, Massimiliano Gioffrè, Vittorio Gusella, Filippo Ubertini

BEHAVIOR OF SINGLE STORY BEARING WALL MASONRY STRUCTURES IN VARIOUS CONDITIONS 5373
Abdelraouf Kassem

SEISMIC PROTECTION OF MONUMENTS AND HISTORIC STRUCTURES – THE SEISMO RESEARCH PROJECT 5382
Constantine Spyrakos, Charilaos Maniatakis

MS 1215: NONLINEAR VIBRATIONS OF CONSERVATIVE AND NONCONSERVATIVE SYSTEMS: PHENOMENA AND ADVANCED NUMERICAL METHODS

FORCED RESPONSE OF SHROUDED BLADES WITH VARIABLE OPERATING POINTS 5396
Ferhat Kaptan, Lars Panning-Von Scheidt, Jörg Wallaschek, Victor Salit

GEOMETRY EFFECTS ON THE NONLINEAR OSCILLATIONS OF VISCOELASTIC CYLINDRICAL SHELLS 5406
Zenon Del Prado, Marco Amabili, Paulo Gonçalves, Frederico Da Silva

REDUCED ORDER MODELS FOR DYNAMIC BEHAVIOR OF PRESTRESSED ELASTOMER DAMPING DEVICES 5417
Antoine Legay, Jean-François Deü, Benjamin Morin

LOCALIZATION OF VIBRATORY ENERGY OF A LINEAR SYSTEM IN A CHAIN OF FOUR NONLINEAR OSCILLATORS 5427
Simon Charlemagne, Claude-Henri Lamarque, Alireza Ture Savadkoohi

CONTINUATION OF QUASI-PERIODIC SOLUTIONS WITH TWO-FREQUENCY HARMONIC BALANCE METHOD .. 5438
Louis Guillot, Pierre Vigué, Christophe Vergez, Bruno Cochelin

MS 1216: STRUCTURAL ANALYSIS AND VIBRATIONS

ASYMPTOTIC ANALYSIS OF DEFORMATIONS OF THE SLIGHTLY ORTHOTROPIC SPHERICAL LAYER UNDER NORMAL PRESSURE 5444
Andrei L. Smirnov, Svetlana M. Bauer, Liudmila A. Venatovskaia, Eva B. Voronkova

EXPERIMENTAL VALIDATION OF FINITE ELEMENT MODELS FOR REINFORCED CONCRETE BEAMS WITH AND WITHOUT DISCONTINUITIES 5454
Marios Filippoupolitis, Carl Hopkins, Siu-Kui Au

MS 1217: COMPUTATIONAL METHODS IN EARTHQUAKE ENGINEERING AND STRUCTURAL DYNAMICS

SEISMIC BEHAVIOR INDICES OF OLD TYPE REINFORCED CONCRETE MEMBERS 5464
Anthos Ioannou, Stavroula J. Pantazopoulou

EFFECT OF SOIL SPATIAL VARIABILITY ON THE DYNAMIC BEHAVIOR OF A SLOPE 5483
Michael Michael, Tamara Al-Bittar, Abdul-Hamid Soubra

ENERGY-MOMENTUM METHOD FOR NONLINEAR DYNAMIC OF 2D COROTATIONAL BEAMS 5496
Sophy Chhang, Mohammed Hjiaj, Jean-Marc Battini, Carlo Sansour

THE INFLUENCE OF UNCERTAINTY OF DESIGN PARAMETERS ON DYNAMIC CHARACTERISTICS OF STRUCTURE WITH DAMPERS 5507
Magdalena Lasecka-Plura, Roman Lewandowski

FINITE-DISCRETE NUMERICAL MODELLING OF REINFORCED CONCRETE STRUCTURES 5523
Nikolina Zivaljic, Zeljana Nikolic, Hrvoje Smoljanovic, Ivan Balic

BEAM DYNAMIC STRESSES INCREMENTS AFTER PARTIAL DECONSTRUCTION OF FOUNDATION 5533
Vladimir Gordon, Olga Pilipenko, Timur Gasimov

RELIABILITY OF SYSTEMS EQUIPPED WITH VISCOUS DAMPERS WITH VARIABLE PROPERTIES 5550
Andrea Dall'Asta, Laura Ragni, Fabrizio Scozzese, Enrico Tubaldi

AN INVESTIGATION ON EFFECT OF SUBGRADE REACTION MODULUS OF SOIL ON NATURAL PERIOD OF LOW RISE BUILDINGS 5567
Rafet Sisman, Abdurrahman Sahin

A PROBABILISTIC APPROACH TOWARDS AN EVALUATION OF EXISTING CODE PROVISIONS FOR SEISMICALLY ISOLATED STRUCTURES 5574
Anastasios Tsiavos, Bozidar Stojadinovic

EFFECT OF THE SEISMIC EXCITATION'S INCIDENCE ANGLE ON THE NONLINEAR BEHAVIOR OF BASE ISOLATED BUILDINGS CONSIDERING POUNDING TO ADJACENT MOAT WALLS 5585
Eftychia A. Mavronicola, Panayiotis C. Polycarpou, Petros Komodromos

NUMERICAL NON-LINEAR SIMULATION OF THE IN-PLANE BEHAVIOUR OF R/C FRAMES WITH MASONRY INFILLS UNDER SEISMIC TYPE LOADING 5598
George Manos, Vasilios Soulis

MODELING THE MECHANICAL RESPONSE OF A LEAD-CORE BEARING DEVICE: DAMAGE MECHANICS APPROACH 5613
Todor Zhelyazov, Rajesh Rupakhety, Simon Olafsson

INVESTIGATION OF THE PERFORMANCE OF VARIOUS MODAL CORRELATION CRITERIA IN STRUCTURAL DAMAGE IDENTIFICATION 5626
Manolis Georgioudakis, Vagelis Plevris

MS 1218: STABILITY AND CONTROL OF FLEXIBLE STRUCTURES

EXPLOITING MULTIPLE REFERENCE MODELS FOR ADAPTIVE CONTROL OF FLEXIBLE STRUCTURES 5646
Ilaria Venanzi, Laura Ierimonti

PASSIVE CONTROL OF WAVE PROPAGATION IN PERIODIC ANTI-TETRACHIRAL META-MATERIALS 5658
Marco Lepidi, Andrea Bacigalupo

FLEXIBLE COMPLEX SYSTEM OF A DOUBLE-STRING UNDER EXTREME MOVING LOADS 5670
Jarosław Rusin

ON THE MODELING OF SELF-DAMPING IN STRANDED CABLES 5680
Francesco Foti, Luca Martinelli, Federico Perotti

COMPUTATIONAL EFFECTIVENESS OF LMI DESIGN STRATEGIES FOR VIBRATION CONTROL OF LARGE STRUCTURES 5689
Francisco Palacios-Quiñonero, Josep Rubió-Massegú, Josep M. Rossell, Hamid Reza Karimi

A SMART BASE RESTRAINT FOR WIND TURBINES TO MITIGATE UNDESIRED EFFECTS DUE TO STRUCTURAL VIBRATIONS 5702
Nicola Caterino, Christos T. Georgakis, Mariacristina Spizzuoco, Antonio Occhiuzzi

MS 1220: DYNAMICS OF STRUCTURES SUBJECT TO SEISMIC EXCITATIONS

NONLINEAR DYNAMIC BEHAVIOUR OF BASE-ISOLATED BUILDINGS WITH THE FRICTION PENDULUM SYSTEM SUBJECTED TO NEAR-FAULT EARTHQUAKES 5712
Fabio Mazza, Sandro Sisinno

ABSOLUTE FINITE ELEMENT COORDINATES IN THE DYNAMICS OF LARGE FLEXIBLE STRUCTURES 5726
Evtim Zahariev

MODAL ANALYSIS OF A FRAME MODEL UNDER UNMEASURED SEISMIC INPUT 5742
Claudio Valente, Vincenzo Sepe, Marco Di Pilla, Fabrizio Iezzi, Rossella Siano, Luigia Zuccarino

EQUIVALENT UNIAXIAL ACCELEROGRAM FOR CSS-BASED ISOLATION SYSTEMS ASSESSMENT UNDER TWO-COMPONENTS SEISMIC EVENTS 5753
Marco Furinghetti, Alberto Pavese

MS 1221: COMPUTATIONAL STRUCTURAL STABILITY

MIXED SOLID MODELS IN NUMERICAL ANALYSIS OF SLENDER STRUCTURES 5768
Domenico Magisano, Leonardo Leonetti, Giovanni Garcea

COMPUTATIONAL ANALYSIS OF THE COLLAPSE BEHAVIOUR OF THIN-WALLED POLYGONAL STEEL BEAMS 5789
Raffaele Ardito

INFLUENCE OF STRAIN DEFINITIONS ON TRUSSES CRITICAL LOADS 5799
Reyolando Brasil, Jose Balthazar

ULTIMATE STRENGTH ANALYSIS OF STIFFENED PANELS OF SHIP STRUCTURES UNDER COMBINED LOAD 5810
Li Hong, Meng Linghua, Qin Zhongwen, Zhang Enguo, Li Li

MS 1222: INFLUENCE OF LIQUEFIABLE SOIL ON SINGLE AND CLOSELY CLUSTERED STRUCTURES
NUMERICAL ANALYSES OF THE INFLUENCE OF STRUCTURAL SLENDERNESS ON THE SEISMIC RESPONSE OF SINGLE AND CLUSTERED STRUCTURES 5817
Gonzalo Barrios, Tam Larkin, Nawawi Chouw

EFFECT OF HIGHER MODES ON STRUCTURAL RESPONSE WITH NONLINEAR SOIL-FOUNDATION-STRUCTURE INTERACTION 5826
Xiaoyang Qin, Nawawi Chouw, Tam Larkin

MS 1224: INNOVATIVE SOLUTIONS FOR THE SEISMIC PROTECTION OF INDUSTRIAL BUILDINGS
SEISMIC RETROFIT OF AN INDUSTRIAL STRUCTURE THROUGH AN INNOVATIVE SELF-CENTERING HYSTERETIC DAMPER: MODELLING, ANALYSIS AND OPTIMIZATION. 5834
Francesco Morelli, Andrea Piscini, Walter Salvatore

PERFORMANCE ASSESSMENT OF SEISMIC RETROFITTING MEASURES ON SILO STRUCTURES USING INNOVATIVE SEISMIC PROTECTION SYSTEMS 5851
Marius Pinkawa, Benno Hoffmeister, Markus Feldmann

SEISMIC RETROFIT OF INDUSTRIAL SILOS BY MEANS OF BASE ISOLATION DEVICES 5868
Alper Kanyilmaz, Carlo Andrea Castiglioni, Julia Georgi

EFFICIENCY OF SEISMIC ISOLATION ON INDUSTRIAL PLANTS - CASE STUDY OF A GAS TANK 5896
José Henriques, Francesco Morelli, Bram Vandoren, Walter Salvatore, Hervé Degée

PERFORMANCE-BASED NONLINEAR RESPONSE HISTORY ANALYSIS FRAMEWORK FOR THE "PROINDUSTRY" PROJECT CASE STUDIES 5912
Marco Faggella, Raffaele Laguardia, Rosario Gigliotti, Francesco Morelli, Franco Braga, Walter Salvatore

PERFORMANCE BASED EARTHQUAKE ASSESSMENT OF AN INDUSTRIAL SILOS STRUCTURE AND RETROFIT WITH SLIDING ISOLATORS 5926
Edoardo Rossi, Michelangelo Ventrella, Marco Faggella, Rosario Gigliotti, Franco Braga

MS 1225: SEISMIC PERFORMANCE ASSESSMENT OF STRUCTURES AND SEISMIC RISK MITIGATION STRATEGIES
A CRITICAL REVIEW OF FRAGILITY CURVES FOR EXISTING RC BUILDINGS 5934
Monica Mastroberti, Marco Vona

DYNAMIC IDENTIFICATION AND MODELLING OF A RC HOSPITAL BUILDING 5942
Giorgio Lacanna, Pauline Deguy, Maurizio Ripepe, Massimo Baglione, Marco Tanganelli, Stefania Viti, Mario De Stefano

SEISMIC HAZARD OF URBAN AREAS: A CASE-STUDY 5952
Giorgio Lacanna, Pauline Deguy, Maurizio Ripepe, Massimo Coli, Barbara Paoletti, Sara Barducci, Marco Tanganelli, Stefania Viti, Mario De Stefano

SPO2FRAG V1.0: SOFTWARE FOR PUSHOVER-BASED DERIVATION OF SEISMIC FRAGILITY CURVES 5962
Iunio Iervolino, Georgios Baltzopoulos, Dimitrios Vamvatsikos, Roberto Baraschino

COMPUTER-AIDED SEISMIC RISK ASSESSMENT AT URBAN SCALE. MODEL DEFINITION AND VALIDATION ON A CASE STUDY 5977
Alberto Basaglia, Alessandra Aprile, Francesco Pilla, Enrico Spacone

TOWARDS INTEGRATED SEISMIC RISK ASSESSMENT IN PALESTINE - APPLICATION TO THE CITY OF NABLUS ... 5987
Ricardo Monteiro, Paola Ceresa, Vania Cerchiello, Jamal Dabeek, Antonella Di Meo, Barbara Borzi

REASSESS V1.0: A COMPUTATIONALLY-EFFICIENT SOFTWARE FOR PROBABILISTIC SEISMIC HAZARD ANALYSIS 5999
Iunio Iervolino, Eugenio Chioccarelli, Pasquale Cito

SIMPLIFIED PERIOD ESTIMATION OF ITALIAN RC BRIDGES FOR LARGE-SCALE SEISMIC ASSESSMENT 6013
C. Zelaschi, R. Monteiro, R. Pinho

1300 UNCERTAINTY QUANTIFICATION AND ERROR ESTIMATION

MS 1301: THE STOCHASTIC COMPUTER METHODS IN MECHANICS
FLOW FIELD ESTIMATION IN OPEN CHANNEL BASED ON KALMAN FILTER FINITE ELEMENT METHOD 6029
Taichi Yoshiara, Takahiko Kurahashi, Yasuhide Kobayashi, Toshihiko Eto

STOCHASTIC TECHNIQUES FOR THE NUMERICAL SOLUTION OF ENGINEERING BOUNDARY VALUE PROBLEMS 6041
Victor Maceiras, Manuel Casteleiro

THE BOOTSTRAP APPROACH TO THE STATISTICAL SIGNIFICANCE OF PARAMETERS IN THE FIXED EFFECTS MODEL 6061
Jacek Pietraszek, Renata Dwornicka, Agnieszka Szczotok

ON FULLY COUPLED THERMO-ELASTO-PLASTIC STOCHASTIC FINITE ELEMENT ANALYSIS OF STEEL STRUCTURES 6069
Marcin Kaminski, Michał Strakowski

AN INTRODUCTION TO STOCHASTIC FINITE ELEMENT METHOD ANALYSIS OF HYPERELASTIC STRUCTURES ... 6078
Marcin Kamiński, Damian Sokołowski

MS 1303: ANALYSIS AND DESIGN OF SAFETY CRITICAL SYSTEMS UNDER UNCERTAINTY
IDENTIFICATION OF INTERVAL FIELDS FOR SPATIAL UNCERTAINTY REPRESENTATION IN FINITE ELEMENT MODELS 6091
Matthias Faes, Jasper Cerneels, Dirk Vandepitte, David Moens

INTRODUCTION OF BEAD INGS INTO A CRASH TUBE USING A ROBUST OPTIMISATION APPROACH 6099
Stefan Scheiblhofer

DYNAMICAL ANALYSIS OF PLATE MODELS WITH UNCERTAIN STRUCTURAL PROPERTIES USING THE INTERVAL FIELD METHOD 6107
Maurice Imholz, Dirk Vandepitte, David Moens

POWER GRID ROBUSTENSS TO SEVERE FAILURES: TOPOLGICAL AND FLOW BASED METRICS COMPARISION 6121
Roberto Rocchetta, Edoardo Patelli

MS 1304: STOCHASTIC MODELING AND IDENTIFICATION OF UNCERTAINTIES IN COMPUTATIONAL MECHANICS
MULTILEVEL STOCHASTIC REDUCED-ORDER MODEL IN LINEAR STRUCTURAL DYNAMICS FOR COMPLEX STRUCTURES 6136
Olivier Ezvan, Anas Batou, Christian Soize

TO ESTIMATE NON-STATIONARY STOCHASTIC DISTRIBUTED LOAD ON A BEAM STRUCTURE FROM RESPONSE SAMPLES 6148
S.Q. Wu, J. Zhu

A ROBUST POLYNOMIAL CHAOS KALMAN FILTER FRAMEWORK FOR CORROSION DETECTION IN REINFORCED CONCRETE STRUCTURES 6160
Wael Slika, George Saad

ANALYSIS OF STOCHASTIC DYNAMIC SOIL-STRUCTURE INTERACTION PROBLEMS BY MEANS OF COUPLED FINITE ELEMENTS-PERFECTLY MATCHED LAYERS 6172
Manthos Papadopoulos, Stijn François, Geert Degrande, Geert Lombaert

A MULTISCALE FRAMEWORK FOR THE STOCHASTIC ASSIMILATION AND MODELING OF UNCERTAINTY ASSOCIATED NCF COMPOSITE MATERIALS 6186
Loujaine Mehrez, Roger Ghanem, Colin Mcauliffe, William R. Rodgers, Venkat Aitharaju

MS 1306: ERCOFTAC SIG-45: UNCERTAINTY QUANTIFICATION IN CFD AND FLUID STRUCTURE INTERACTION
A MULTI-FIDELITY ADAPTIVE SAMPLING METHOD FOR METAMODEL-BASED UNCERTAINTY QUANTIFICATION OF COMPUTER SIMULATIONS 6196
Riccardo Pellegrini, Cecilia Leotardi, Umberto Iemma, Emilio Campana, Matteo Diez

A LEAST-SQUARES, ADAPTIVE UNCERTAINTY PROPAGATION APPROACH FOR A PLASMA-COUPLED COMBUSTION SYSTEM 6213
Kunkun Tang, Luca Massa, Jonathan Wang, Jonathan Freund

UNCERTAINTY QUANTIFICATION IN NUMERICAL SIMULATIONS OF THE FLOW IN THORACIC AORTIC ANEURYSMS 6226
Alessandro Boccadifuoco, Alessandro Mariotti, Simona Celi, Nicola Martini, Maria Vittoria Salvetti

MS 1307: NON-INTRUSIVE SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION IN HIGH DIMENSIONS
OPTIMIZATION OF A PHOTOACOUSTIC GAS SENSOR USING MULTIFIDELITY RBF METAMODELING 6250
Cedric Durantin, Justin Rouxel, Jean-Antoine Desideri, Alain Gliere

BEARING CAPACITY OF STRIP FOOTINGS ON SPATIALLY RANDOM SOILS USING KRIGING AND MONTE CARLO SIMULATION 6260
Jawad Thajeel, Tamara Al-Bittar, Nour Issa, Abdul Hamid Soubra

AN OPTIMAL SENSOR PLACEMENT METHOD FOR SHM BASED ON BAYESIAN EXPERIMENTAL DESIGN AND POLYNOMIAL CHAOS EXPANSION 6272
Giovanni Capellari, Eleni Chatzi, Stefano Mariani

CALIBRATION OF NESTED COMPUTER MODELS 6283
Sophie Marque-Pucheu, Guillaume Perrin, Josselin Garnier

A STATISTICAL APPROACH FOR BUILDING SPARSE POLYNOMIAL CHAOS EXPANSIONS 6307
Simon Abraham, Ghader Ghorbaniasl, Chris Lacor

MS 1308: MODELLING AND INVERSE METHODS IN NONLINEAR DYNAMICAL SYSTEMS
FINITE ELEMENT METAMODELING OF UNCERTAIN STRUCTURES 6316
Vasilis Dertimanis, Dimitris Giagopoulos, Eleni Chatzi

DYNAMICS OF MULTIBODY SYSTEMS SUBJECT TO UNILATERAL MOTION CONSTRAINTS USING A NEW NUMERICAL ALGORITHM FOR CONTACT DETECTION 6328
Antonios Pournaras, Fotios Karaoulanis, Sotirios Natsiavas

EQUATIONS OF MOTION FOR MECHANICAL SYSTEMS SUBJECT TO ACATASTATIC CONSTRAINTS 6342
Elias Paraskevopoulos, Sotirios Natsiavas

PARAMETER ESTIMATION OF NONLINEAR LARGE SCALE SYSTEMS THROUGH STOCHASTIC METHODS AND MEASUREMENT OF ITS DYNAMIC RESPONSE 6369
Dimitrios Giagopoulos, Alexandros Arailopoulos

MS 1309: SCALABLE MULTI-FIDELITY MODELING FOR DESIGN, UNCERTAINTY QUANTIFICATION, AND INVERSE PROBLEMS
VARIABLE FIDELITY MODELLING IN MODERN AIRCRAFT DESIGN 6383
Marian Zastawny

MS 1310: COMPUTATIONAL METHODS FOR THE SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS
FOUR-DIMENSIONAL PATH INTEGRATION METHOD FOR ESTIMATING THE STOCHASTIC ROLL RESPONSE 6398
Wei Chai, Arvid Naess, Bernt Leira

SEISMIC RELIABILITY ASSESSMENT OF FRAME STRUCTURES WITH FLEXIBILITY-BASED STOCHASTIC BEAM-COLUMN ELEMENTS 6407
Georgios Balokas, Michalis Fragiadakis

1400 SPECIAL TECHNOLOGICAL SESSIONS

MS 1401: TOICA: THERMAL OVERALL INTEGRATED CONCEPT AIRCRAFT
THERMAL TRADE OFF SUSTAINED BY MULTI DISCIPLINARY AND MULTI LEVEL OPTIMIZATION 6420
Olivier Tabaste, Cedric Campguilhem

COLLABORATIVE ENGINEERING TECHNOLOGIES ENABLING MULTI-PARTNER THERMAL ANALYSIS IN EARLY DESIGN STAGES OF AIRCRAFT 6436
Erik Baalbergen, Wim Lammen, Bert De Wit, Robert Maas, Marie Moghadasi, Johan Kos, Fabio Chiacchio

VISUAL ANALYTICS FOR EVALUATION OF VALUE IMPACT IN ENGINEERING DESIGN 6461
Timoleon Kipouros, Ola Isaksson

STS 3: INNOVATIVE DESIGN OPTIMIZATION TOOLS LINKED TO INDUSTRIAL AERONAUTICAL APPLICATIONS: TARGETING GREENER PERFORMANCES
COMBINING AN RBF-BASED MORPHER WITH CONTINUOUS ADJOINT FOR LOW-SPEED AERONAUTICAL OPTIMIZATION APPLICATIONS 6471
Evangelos Papoutsis-Kiachagias, Matej Andrejasic, Stefano Porziani, Corrado Groth, David Erzen, Marco Evangelos Biancolini, Emiliano Costa, Kyriakos Giannakoglou

STS 5: TRANSITION LOCATION EFFECT ON SHOCK WAVE BOUNDARY LAYER INTERACTION
LOAD CONTROL OF NATURAL-LAMINAR-FLOW WING VIA BOUNDARY LAYER CONTROL 6485
Wieńczysław Stalewski, Janusz Sznajder

STS 7: MORPHING TECHNOLOGIES FOR AIRCRAFT WINGS
AN ADAPTIVE TRAILING EDGE FOR LARGE COMMERCIAL AIRCRAFT 6496
Antonio Concilio, Ignazio Dimino, Rosario Pecora

STS 9: ADVANCED WING HIGH-LIFT SYSTEMS
APPLICATION OF SYNTHETIC JETS ACTUATORS IN WING-PYLON JUNCTION AREA TO IMPROVE THE HIGH LIFT PERFORMANCES 6505
Petr Vrchota

YOUNG INVESTIGATORS MINISYMPOSIUM
A NANO-MACRO BOTTOM-UP APPROACH TOWARDS BRITTLE FRACTURE 6515
Sandeep Patil, Yousef Heider, Carlos Hernandez Padilla, Eduardo Cruz Chu, Bernd Markert

ECCOMAS OLYMPIAD MINISYMPOSIUM
INTEGRATION OF DESIGN AND ANALYSIS THROUGH BOUNDARY INTEGRAL EQUATIONS 6526
Benjamin Marussig, Jürgen Zechner, Gernot Beer, Thomas-Peter Fries

Contributed Sessions

CS 110: NUMERICAL MODELS IN BIOMECHANICS

EVALUATION OF ANGULAR KINEMATICS OF LOWER LIMB AMPUTEES USING QUANTITATIVE FLUOROSCOPIC IMAGING 6539
Alexander Breen, Mihai Dupac

PRELIMINARY STUDY TO INVESTIGATE THE EFFECT OF PISTON-LIKE AND ROCKING MOTIONS OF THE STAPES FOOTPLATE ON THE BASILAR MEMBRANE VIBRATION 6547
Philipp Wahl, Sebastian Ihrle, Pascal Ziegler, Peter Eberhard

IN-HOUSE MULTIBODY HUMAN MODEL BASED ON EULER PARAMETERS FOR THE FAST IMPACT SCENARIO CALCULATION 6556
Jan Spicka, Michal Hajzman, Tomasz Bonkowski

EXPERIMENTAL STUDY ON THE FUZZY-PID HYBRID CONTROL ALGORITHM FOR UNLOADING SYSTEM IN MECHATRONIC DEVICE FOR GAIT RE-EDUCATION 6567
Slawomir Duda, Grzegorz Gembalczyk

FE BONE STRUCTURAL ANALYSIS WITH CT MAPPING OF INHOMOGENEOUS MATERIAL PROPERTIES 6574
Miguel Tobias Bahia, Emílio Graciliano Ferreira Mercuri, Mildred Ballin Hecke

SOLVING THE INVERSE PROBLEM OF ESTIMATING FUZZY VISCOELASTIC CONSTITUTIVE PARAMET ERS 6588
Ruifei Peng, Haitian Yang

DEVELOPMENT OF A CENTRIFUGAL BLOOD PUMP FOR ECMO AND VAD OPERATIONS 6606
Shinhwa Choi, Nahmkeon Hur, Mohammad Moshfeghi, Seongwon Kang, Wonjung Kim

EFFECT OF PERIVASCULAR TISSUE ON INFLATION-EXTENSION BEHAVIOR OF ABDOMINAL AORTA 6616
Tereza Voňavková, Lukáš Horný, Jan Veselý, Tomáš Adámek, Rudolf Žitný

EFFECT OF STERILIZATION ON MECHANICAL PROPERTIES OF BIOLOGICAL COMPOSITE 6625
Jan Vesely, Hynek Chlup, Rudolf Zitny, Tomas Grus

FINITE-ELEMENT MODEL OF INTRAOCULAR PRESSURE MEASUREMENT BY MAKLAKOV APPLANATION TONOMETER 6631
Dmitry V. Franus

CS 210: NUMERICAL SIMULATION OF COMPOSITE MATERIALS

MOMENT-CURVATURE DIAGRAMS FOR EVALUATION OF SECOND ORDER EFFECTS IN RC ELEMENTS 6637
Helena Barros, Carla Ferreira, Tatiana Marques

INVESTIGATIONS ON AN INNOVATIVE METAL-THERMOPLASTIC COMPOSITE ASSEMBLY 6652
Simon Paroissien, Patrick Rozycki, Thierry Renault

A FAILURE CRITERIA FOR UNIDIRECTIONAL FIBER REINFORCED COMPOSITES BASED ON MICROMECHANICS BY ASYMPTOTIC HOMOGENIZATION 6664
Rafael Quelho De Macedo, José Miranda Guedes, Rafael Thiago Luiz Ferreira, Maurício Vicente Donadon

EVOLUTION EQUATIONS BASED APPROACH FOR MODELING FATIGUE IN AMORPHOUS GLASSY POLYMERS. ON THE INVESTIGATION OF FATIGUE DAMAGE DEVELOPMENT IN POLYCARBONATE 6675
Sami Holopainen

CS 211: CRACK PROPAGATION
ULTRASONIC GUIDED WAVES INSPECTION OF PIPES FROM ONE END TO THE FIRST BENT 6688
Mihai Valentin Predoi, Marian Soare, Ovidiu Vasile, Mihail Boiangiu

CS 212: NUMERICAL MODELING OF DAMAGE, FAILURE AND FRACTURE
DUCTILE FRACTURE CRITERIA IN PREDICTION OF SLANT FRACTURE 6699
Petr Kubik, Frantisek Sebek, Jindrich Petruska

DAMAGE MODELING USING STRAIN GRADIENT BASED FINITE ELEMENT FORMULATION 6711
Filip Putar, Jurica Soric, Tomislav Lesicar, Zdenko Tonkovic

RELIABILITY ANALYSIS IN FRACTURE MECHANICS ACCORDING TO COMBINED FAILURE CRITERIA 6721
Rudy Chocat, Paul Beaucaire, Loïc Debeugny, Jean-Pierre Lefebvre, Caroline Sainvitu, Piotr Breitkopf, Eric Wyart

NEW BENCHMARK FOR THE LIFE ASSESSMENT OF A THIN-WALLED PIPE SUBJECTED TO STRESS ASSISTED CORROSION 6734
Yulia Pronina, Elena Sedova

A HIGHER ORDER PHASE-FIELD APPROACH TO FRACTURE FOR FINITE-DEFORMATION CONTACT PROBLEMS 6741
Marlon Franke, Christian Hesch, Maik Dittmann

MODELLING THERMAL BARRIER COATINGS AND THEIR INFLUENCE ON THE LIFETIME OF ROCKET ENGINE NOZZLE STRUCTURES 6764
Marek Fassin, Stephan Wulfinghoff, Stefanie Reese

ANALYTICAL SOLUTION FOR DYNAMIC FRACTURE OF TWO COPLANAR LIMITED-PERMEABLE CRACKS IN MAGNETO-ELECTRO-ELASTIC MATERIAL 6774
Peiwei Zhang

DAMAGE DETECTION OF A BRIDGE BY PARAMETRIC STATISTICAL MOMENT METHOD 6788
Isabella Failla, Nicola Impollonia, Giuseppe Ricciardi

CS 230: MODELLING OF CONCRETE AND MEASURING
A COHESIVE ZONE MODEL FOR THE CHARACTERISATION OF THE INTERFACIAL TRANSITION ZONE (ITZ) BETWEEN CEMENT PASTE AND AGGREGATE 6799
Etienne Malachanne, Marie Salgues, Mouad Jebli, Frederic Jamin

AN EXACT SHEAR STRAIN APPROACH FOR RC FRAME ELEMENTS WITH AXIAL-SHEAR INTERACTION 6810
Alexander Kagermanov, Paola Ceresa

CROSS-SECTIONAL FAILURE CRITERION COMBINED WITH STRAIN-HARDENING DAMAGE MODEL FOR SIMULATION OF THIN-WALLED TEXTILE-REINFORCED CONCRETE SHELLS 6823
Ehsan Sharei, Rostislav Chudoba, Alexander Scholzen

SEISMIC ASSESSMENT OF A 5-STOREY RETROFITTED RC BUILDING 6832
Hussein Bark, George Markou, Christos Mourlas, Manolis Papadrakakis

CS 310: CAD, CAM AND CAE
VECTOR FIELD GUIDED TOOL PATHS FOR FIVE-AXIS MACHINING 6847
Stanislav Makhanov

AN 'A PRIORI' MODEL REDUCTION FOR ISOGEOMETRIC BOUNDARY ELEMENT METHOD 6882
Shengze Li, Jon Trevelyan, Weihua Zhang, Zhuxuan Meng

RESEARCH OF DEFORMATION QUALITIES OF POROUS MATERIALS ENHANCING COMPUTER DESIGN OF WETSUIT 6892
Irina Cherunova, Elena Sirota, Nikolai Kornev, Mathias Paschen, Sebastian Schreier, Tatyana Lesnikova, Pavel Cherunov

MATHEMATICAL MODELLING OF CITY AERODYNAMICS 6900
Svetlana Valger, Natalya Fedorova, Alexander Fedorov

CS 320: GRID GENERATION AND ADAPTIVE TECHNIQUES
A STABLE AND CONSERVATIVE TIME-DEPENDENT INTERFACE FORMULATION ON SUMMATION-BY-PARTS FORM: AN INITIAL INVESTIGATION 6911
Samira Nikkar, Jan Nordström

A LAPLACIAN MESH DEFORMATION TECHNIQUE FOR SIMULATION-DRIVEN DESIGN OPTIMIZATION 6928
Mario J. Martin-Burgos, Daniel González-Juárez, Esther Andrés-Pérez

AN INTEGRATED FRAMEWORK FOR WRAPPING AND MESH GENERATION OF COMPLEX GEOMETRIES 6938
David Martineau, Jeremy Gould, Jacques Papper

CS 410: COMPUTATIONAL FLUID MECHANICS
THE VORTICITY CREATION PROCESS AT PHYSICAL SURFACES 6955
Horia Dumitrescu, Vladimir Cardos

HEAT RECOVERY BY CROSS FLOW 6966
Karel Adamek, Jan Kolar, Pavel Peukert

THE INLET AND OUTLET BOUNDARY PROBLEM WITH THE PREFERENCE OF MASS FLOW 6979
Martin Kyncl, Jaroslav Pelant

BUDGET ANALYSIS OF TURBULENT KINETIC ENERGY IN CORNER SEPARATION : RANS VS LES 6999
Jean-François Monier, Feng Gao, Jérôme Boudet, Liang Shao, Lipeng Lu

A SHALLOW WATER EQUATION BASED ON DISPLACEMENT AND PRESSURE AND THE ZU-CLASS SYMPLECTIC METHOD 7014
Feng Wu, Wan-Xie Zhong

IMMERSED BOUNDARY METHODS FOR COMPRESSIBLE LAMINAR FLOWS 7029
Rakesh Ramakrishnan, Anant Girdhar, Santanu Ghosh

IMPACT OF EXTERNAL SURROUNDINGS ON NATURAL CONVECTION IN A VERTICAL CHANNEL ASYMMETRICALLY HEATED 7045
Delphine Ramalingom, Alain Bastide

INFLUENCE OF ADJACENT ROOMS ON THE DEVELOPMENT OF GAS EXPLOSION 7056
Iurii Polandov, Sergei Dobrikov

DOMAIN DECOMPOSITION APPROACH FOR NEAR-WALL TURBULENCE MODELING 7066
Sergey Utyuzhnikov

LARGE EDDY SIMULATION OF SALTATION OVER GAUSSIAN HILLS 7074
Gang Huang, Catherine Le Ribault, Serge Simoëns, Ivana Vinkovic, J.M.Vignon

AN EXPERIMENTAL AND NUMERICAL STUDY ON THE PERFORMANCE OF AN INNOVATIVE VERTICAL-AXIS WIND TURBINE 7087
Katarzyna Kludzinska, Krzysztof Tesch, Piotr Doerffer

NUMERICAL SIMULATION OF NONLINEAR FREE SURFACE WATER WAVES USING A COUPLED POTENTIAL FLOW-URANS/VOF APPROACH 7095
Bulent Duz, Tim Bunnik, Geert Kapsenberg

THE OPTIMAL CONTROL OF A MULTI-MASS VIBRATION PROPULSION SYSTEM IN A VISCOUS INCOMPRESSIBLE FLUID 7121
Artem Nuriev, Olga Zakharova

HIGH-RESOLUTION SIMULATION OF INTERNAL WAVES ATTRACTORS AND IMPACT OF INTERACTION OF HIGH AMPLITUDE INTERNAL WAVES WITH WALLS ON DYNAMICS OF WAVES ATTRACTORS 7130
Ilias Sibgatullin, Michael Kalugin

NUMERICAL SIMULATION OF HEAT AND MASS TRANSFER PROCESSES IN LARGE-SCALE FLUIDIZED BED COMPLEX STRUCTURE APPARATUS AS AN EXAMPLE OF THE REACTOR OF ISOPARAFFINS DEHYDROGENATION 7138
Sergei Solovev, Svetlana Egorova, Alexander Lamberov, Olga Soloveva

CS 420: MULTI-PHASE AND CHEMICALLY REACTING FLOWS

SHOCK CAPTURING COMPUTATIONS WITH STABILIZED POWELL-SABIN ELEMENTS 7149
Giorgio Giorgiani, Herve Guillard, Boniface Nkonga

NUMERICAL INVESTIGATION OF METASTABLE CONDENSING FLOWS WITH AN IMPLICIT UPWIND METHOD ... 7165
Lucia Azzini, Teus Van Der Stelt, Matteo Pini

LEVEL SET METHOD FOR SIMULATING THE INTERFACE KINEMATICS: APPLICATION OF A DISCONTINUOUS GALERKIN METHOD 7182
Roozbeh Mousavi, Florian Kummer, Martin Oberlack, Peter F. Pelz

LARGE EDDY SIMULATION OF TURBULENT COMPRESSIBLE FLOWS USING THE CHARACTERISTIC BASED SPLIT SCHEME AND MESH ADAPTATION 7222
Renato Linn, Armando Awruch

NUMERICAL SIMULATION OF HYDROGEN JET INJECTION AND IGNITION IN SUPERSONIC FLOW 7237
Yulia Zakharova, Natalya Fedorova, Svetlana Valger, Marat Goldfeld, Olga Vankova

A 3D UNSTEADY NUMERICAL SIMULATION OF THE REFRIGERANT FLOW EVAPORATION IN A PLATE HEAT EXCHANGER 7247
Mirza Popovac, Gerwin Schmid, Michael Lauermann

RESEARCH OF UNSTEADY FLOW REGIMES IN CHANNEL OF HYPERSONIC INLET 7255
Natalya Fedorova, Marat Goldfeld

EXPLICIT DARCY'S LAW BOUNDARY CONDITION WITH COMBINED CONTINUUM AND DISCRETE MODEL FOR PRESSURE DRIVEN MEMBRANE APPLICATIONS 7266
Tomi Naukkarinen, Teemu Turunen-Saaresti

CS 450: NUMERICAL METHODS AND CONVERGENCE ACCELERATION IN CFD

A MULTIGRID FORMULATION FOR FINITE DIFFERENCE METHODS ON SUMMATION-BY-PARTS FORM: AN INITIAL INVESTIGATION 7274
Andrea Alessandro Ruggiu, Per Weinerfelt, Tomas Lundquist, Jan Nordström

IMPROVED DUAL TIME–STEPPING USING SECOND DERIVATIVES 7285
Jan Nordström, Andrea Ruggiu

A DIRECT SOLVER FOR THE ADVECTION-DIFFUSION EQUATION USING GREEN'S FUNCTIONS AND LOW-RANK APPROXIMATION 7302
Jonathan Bull

AN IMPROVED DISCRETIZATION METHOD FOR BOUNDED CONVECTIVE SCHEMES ON UNSTRUCTURED CO-LOCATED GRIDS 7317
Vlado Przulj

A LOW-MACH, LOW-REYNOLDS PRECONDITIONING SCHEME WITH PARTICULAR ATTENTION TO VISCOUS TIME-STEPPING 7334
Jens Fiedler, Graham Ashcroft

PERFORMANCE IMPROVEMENT OF FLOW COMPUTATIONS WITH AN OVERSET-GRID METHOD INCLUDING BODY MOTIONS USING A FULL MULTIGRID METHOD 7347
Kunihide Ohashi, Hiroshi Kobayashi

SOLVING LINEAR SYSTEMS WITH MULTIPLE RIGHT-HAND SIDES WITH GMRES : AN APPLICATION TO AIRCRAFT DESIGN 7358
Aloïs Bissuel, Grégoire Allaire, Laurent Daumas, Frédéric Chalot, Michel Mallet

COMPARISON OF DIFFERENT SPATIAL/ANGULAR AGGLOMERATION MULTIGRID SCHEMES FOR RADIATIVE HEAT TRANSFER COMPUTATIONS 7372
Georgios Lygidakis, Ioannis Nikolos

CS 460: UNSTEADY FLOW COMPUTATION
LINEARISED FREQUENCY DOMAIN GUST ANALYSIS OF LARGE CIVIL AIRCRAFT 7390
Philipp Bekemeyer, Reik Thormann, Sebastian Timme

CONSISTENT NON-REFLECTING BOUNDARY CONDITIONS FOR BOTH STEADY AND UNSTEADY FLOW SIMULATIONS IN TURBOMACHINERY APPLICATIONS 7403
Daniel Schlüß, Christian Frey, Graham Ashcroft

REDUCED ORDER MODELLING OF GUST ANALYSIS USING COMPUTATIONAL FLUID DYNAMICS 7423
Reik Thormann, Philipp Bekemeyer, Sebastian Timme

ON THE REGULARISATION OF NON-REFLECTING BOUNDARY CONDITIONS NEAR ACOUSTIC RESONANCE 7435
Christian Frey, Hans-Peter Kersken

ON GRID RESOLUTION REQUIREMENTS FOR LES OF WALL-BOUNDED FLOWS 7454
Saleh Rezaeiravesh, Mattias Liefvendahl, Christer Fureby

A HYBRID MESH HARMONIC BALANCE SOLVER FOR THE AEROELASTIC ANALYSIS OF TURBOMACHINERY 7466
Hans-Peter Kersken, Graham Ashcroft, Christian Frey

NUMERICAL AND EXPERIMENTAL INVESTIGATIONS OF DYNAMIC STALL PROBLEM ON A 2D HELICOPTER ROTOR BLADE SECTION 7477
Antonello Marino, Serena Russo, Giovanni Paolo Reina, Gennaro Esposito, Francesco Capizzano, Carlo De Nicola

INVESTIGATION OF A MODERN & HYBRID TURBULENCE MODELLING APPROACH FOR TRANSIENT AUTOMOTIVE AERODYNAMICS SIMULATIONS 7488
Rene Devaradja, Petr Simanek, Jacques Papper, Pavla Policka

FREE-SURFACE FLOW SIMULATIONS FOR MOORED AND FLOATING OFFSHORE PLATFORMS 7515
Arthur Veldman, Roel Luppes, Peter Van Der Plas, Henri Van Der Heiden, Bulent Duz, Henk Seubers, Joop Helder, Tim Bunnik, Rene Huijsmans

MODELLING OF UNSTEADY SECONDARY VORTICES GENERATED BEHIND THE RADIAL GAP OF THE AXIAL TURBINE BLADE WHEEL 7532
Petr Straka

PATHWAYS TO IMPROVED AERODYNAMIC DESIGN 7544
James Page, Paul Hield, Callum Mantell, Paul Tucker

SIMULATIONS OF BLAST WAVE PROPAGATION IN OPEN AND CLOSED SPACE 7557
Natalya Fedorova, Svetlana Valger, Yulia Zakharova

CS 500: HIGH PERFORMANCE COMPUTING
ON A DYNAMIC SCHEDULING ALGORITHM FOR MASSIVELY PARALLEL COMPUTATIONS OF ATOMIC ISOTOPE 7572
Elizaveta Dorofeeva, Jatin Arora, Stefan Typel, Gevorg Poghosyan, Peter Sanders, Achim Streit

APPLICATION THE SPARSE MATRIX METHOD TO CALCULATE THE METAL ELASTIC STRESS-STRAIN STATE, USING THE FINITE ELEMENT METHOD 7581
Inna Barankova, Uliana Mikhailova, Victor Demidovich

CS 610: AERO-ACOUSTICS
FLUID ANALYSIS USING FICTITIOUS DOMAIN FINITE ELEMENT METHOD 7589
Yukihiro Terakado, Tahahiko Kurahashi

ON THE SIMULATION OF AERODYNAMIC NOISE WITH DIFFERENT TURBULENCE MODELS 7599
Xin Huang, Michael Schäfer

NUMERICAL INVESTIGATION OF THERMOACOUSTIC ENGINE USING IMPLICIT LARGE EDDY SIMULATION 7609
Mustapha Mahdaoui, Rachid Bennacer, Smaïne Kouidri, N. Martaj

SEMI-DISCRETE COINCIDENCE IN THE MID-FREQUENCY SOUND TRANSMISSION THROUGH RIB-STIFFENED PANELS 7620
Edwin Reynders, Cédric Van Hoorickx, Arne Dijckmans

CS 620: COMPUTATIONAL CONTACT MECHANICS
SOLVING UNILATERAL CONTACT PROBLEMS IN MULTIBODY DYNAMICS USING A PRIMAL-DUAL INTERIOR POINT METHOD 7628
Dario Mangoni, Alessandro Tasora

CS 630: SIMULATION OF FLUID-STRUCTURE INTERACTION
3D DRBEM MODELING FOR ROTATING INITIALLY STRESSED ANISOTROPIC FUNCTIONALLY GRADED PIEZOELECTRIC PLATES 7640
Mohamed Abdelsabour Fahmy

DIRECT EDDY CURRENT METHOD FOR VOLUMETRIC FLAWS OF CYLINDRICAL SHAPE 7659
Valentina Koliskina, Andrei Kolyshkin, Rauno Gordon, Olev Märtens

NUMERICAL SIMULATION OF BUOYANT PLUMES USING A FIXED POINT ITERATIVE METHOD 7666
Blanca Bermúdez, Alejandro Rangel-Huerta, W. Fermín Guerrero S., José David Alanis Urquieta

INVESTIGATION OF THE FLOW OVER AN OSCILLATING CYLINDER WITH A VERY LARGE EDDY SIMULATION MODEL 7674
Anastasia Kondratyuk, Michael Schäfer, Awais Ali

NUMERICAL STUDY OF TWO OPTIMIZED COUPLING INTERFACE TREATMENTS FOR STEADY CONJUGATE HEAT TRANSFER PROBLEMS 7682
Marc Paul Errera, Roch Roukoz El Khoury

EFFECT OF DISC GEOMETRY ON THE DYNAMIC STABILITY OF DIRECT SPRING OPERATED PRESSURE RELIEF VALVES 7695
István Erdődi, Csaba Hős

CFD SIMULATION OF THERMO- AERODYNAMIC INTERACTION IN A SYSTEM HUMAN-CLOTH-ENVIRONMENT UNDER VERY LOW TEMPERATURE AND WIND CONDITIONS 7703
Irina Cherunova, Sina Samarbaksch, Nikolai Kornev

CS 720: COMPUTATIONAL MATERIALS SCIENCE
ON THE INFLUENCE OF THE INCLUSIONS' MORPHOLOGY ON THE ACCURACY OF THE PREDICTION OF REINFORCED COMPOSITES MECHANICAL BEHAVIOUR 7711
Viwanou Hounkpati, Vladimir Salnikov, Philippe Karamian-Surville, Alexandre Vivet

NUMERICAL SIMULATION OF ADVANCED PULTRUSION PROCESSES WITH MICROWAVE HEATING 7720
Evgeny Barkanov, Pavel Akishin, Rudolf Emmerich, Matthias Graf

EFFICIENT NUMERICAL SIMULATION OF INDUSTRIAL SHEET METAL BENDING PROCESSES 7739
Christian Zehetner, Paula Reimer, Franz Hammelmüller, Hans Irschik, Wolfgang Kunze

NUMERICAL MODELLING AND SIMULATION OF SHEET METAL CUTTING PROCESSES 7749
Paula Reimer, Christian Zehetner, Franz Hammelmüller, Wolfgang Kunze

SIMULATION OF PRECIPITATION IN V-CONTAINING HSLA STEEL FOR THE STRENGTHENING ENHANCEMENT 7757
Piyada Suwanpinij, Paolo Massaro, Annalisa Pola, Prasonk Srichareonchai

CS 750: COMPUTATIONAL MODELING OF COMPOSITES
NUMERICAL ANALYSIS OF SANDWICH PANELS SUBJECTED TO POINT LOADS 7765
Zbigniew Pozorski, Jolanta Pozorska

BOND OF REINFORCEMENT IN REACTIVE POWDER CONCRETE: EXPERIMENTAL STUDY 7771
Mingde Sun, Ri Gao, Aili Li, Yongjing Wang

MODELLING SOIL-FIBRE COMPOSITE BEHAVIOUR USING A MICROMECHANICAL APPROACH 7779
Thomas Bower, Anthony Jefferson, Peter Cleall, Paul Lyons

NUMERICAL SIMULATION OF MECHANICAL BEHAVIOR OF WOVEN COMPOSITE AT DIFFERENT STRAIN RATE BY A COLLABORATIVE ELASTO-PLASTO-DAMAGE MODEL WITH FRACTIONAL DERIVATIVES 7786
Alina Krasnobrizha, Patrick Rozycki, Laurent Gornet, Pascal Cosson

NUMERICAL SIMULATION OF MECHANICAL BEHAVIOUR OF COMPOSITE SANDWICH PANELS WITH DEFECTS 7800
Aleksandr Anoshkin, Valerii Zuiko, Mikhail Alikin, Anna Tchugaynova

TIMOSHENKO BEAM ELEMENT WITH ANISOTROPIC CROSS-SECTIONAL PROPERTIES ……. 7810
Alexander R. Stäblein, Morten H. Hansen

STABILIZATION OF THE NUMERICAL PRIMING TO STABILIZATION PHENOMENON IN THE CRUSHING OF A COMPOSITE PANEL USING LOCAL AND DIFFUSE DAMAGE LAWS …….. 7820
Christine Espinosa, Floran Tostain, Samuel Rivallant

STRUCTURAL ANALYSIS OF ADHESIVE BONDING FOR THICK-WALLED TUBULAR COMPOSITE PROFILES …….. 7837
Geminiano Mancusi, Agostina Orefice, Luciano Feo, Fernando Fraternali

CS 751: SMART MATERIALS AND STRUCTURES
TOWARDS REAL-TIME STRUCTURAL HEALTH MONITORING DAMAGE DETECTION WITHOUT USER INPUT ….. 7853
Mohmmad Salmanpour, Zahra Sharif Khodaei, Ferri Aliabadi

FIBROUS TECTONICS:A RETHINKING OF COMPOSITE PRODUCTION THROUGH INNOVATION AND EXPLORATION OF MOLDING TECHNIQUES AND METHODOLOGIES …….. 7863
David Costanza

CS 830: COMPUTATIONAL NANOTECHNOLOGY
EFFECT OF NANOSIZED ASPERITIES AT THE SURFACE OF A NANOHOLE …….. 7875
Mikhail Grekov, Aleksandra Vakaeva

CS 840: MULTI-SCALE COMPUTATIONAL METHODS FOR SOLIDS AND FLUIDS
THE GEMA FRAMEWORK – AN INNOVATIVE FRAMEWORK FOR THE DEVELOPMENT OF MULTIPHYSICS AND MULTISCALE SIMULATIONS …….. 7886
Carlos Augusto Teixeira Mendes, Marcelo Gattass, Deane Roehl

NONLINEAR PATCH METHOD AND APPLICATION …….. 7895
Antoine Brunet, Pierre Sarrailh, François Rogier, Jean-François Roussel, Denis Payan

ASSESSMENT OF A CONTINUUM MICROMECHANICS-BASED MULTISCALE MODEL FOR CONCRETE BY MEANS OF SENSITIVITY ANALYSIS AND UNCERTAINTY PROPAGATION …….. 7905
Luise Göbel, Andrea Osburg, Tom Lahmer

SPATIAL CLUSTERING STRATEGIES FOR HIERARCHICAL MULTI-SCALE MODELLING OF METAL PLASTICITY …… 7922
Md Khairullah, Jerzy Gawad, Dirk Roose, Albert Van Bael

BLACK-BOX SOLVER FOR ONE-DIMENSIONAL MULTISCALE MODELLING USING THE QTT FORMAT …….. 7938
Ivan Oseledets, Maxim Rakhuba, Andrei Chertkov

SCALE EFFECT IN MICROFLOWS MODELLING WITH THE MICROPOLAR FLUID THEORY …….. 7948
Anna Kucaba-Pietal

CS 930: HIGH-ORDER DISCRETIZATION METHODS

A COMPARISON OF VARIOUS NODAL DISCONTINUOUS GALERKIN METHODS FOR THE 3D EULER EQUATIONS 7956
Michael Bergmann, Svetlana Drapkina, Graham Ashcroft, Christian Frey

A 3D ISOGEOMETRICAL BOUNDARY ELEMENT ANALYSIS FOR NON-LINEAR GRAVITY WAVE PROPAGATION ... 7967
Jorge Maestre, Jordi Pallarés, Ildefonso Cuesta

HYBRID OPENMP/MPI PARALLELIZATION OF A HIGH–ORDER DISCONTINUOUS GALERKIN CFD SOLVER 7992
Francesco Bassi, Alessandro Colombo, Andrea Crivellini, Matteo Franciolini

ON THE CHOICE OF SHOCK CAPTURING SCHEMES FOR THE SOLUTION OF THE LWR TRAFFIC FLOW EQUATION USING A HIGH ORDER MODAL DISCONTINUOUS GALERKIN DISCRETIZATION 8013
Alberto Costa Nogueira Junior, João Lucas De Sousa Almeida, Cláudio Alessandro De Carvalho Silva

COMPARISON OF FINITE VOLUME HIGH-ORDER SCHEMES FOR THE WO-DIMENSIONAL EULER EQUATIONS 8033
Jens Wellner

NURBS-BASED ISOGEOMETRIC ANALYSIS FOR BALLISTIC EVALUATION OF TITANIUM PLATES 8047
Mattia Montanari, Petros Siegkas, Antonio Pellegrino, Nik Petrinic

HYBRID RIEMANN SOLVERS FOR LARGE SYSTEMS OF CONSERVATION LAWS 8057
Birte Schmidtmann, Mariia Astrakhantceva, Manuel Torrilhon

CS 940: EXTENDED DISCRETIZATION METHODS

THE USE OF ENRICHED BASE FUNCTIONS IN THE THREE-DIMENSIONAL SCALED BOUNDARY FINITE ELEMENT METHOD 8066
Sascha Hell, Wilfried Becker

METHODOLOGY OF COMBINED APPLICATION OF DIRECTIONAL DERIVATIVES AND THE EXTENDED FINITE ELEMENT METHOD (X-FEM) FOR SOLVING VIBRATION EIGENVALUE PROBLEMS 8085
Daria Serbichenko, Pascal Cosson, Patrick Rozycki

XFEM ANALYSIS – INCLUDING BUCKLING – OF COMPOSITE SHELLS CONTAINING DELAMINATION 8103
Wilhelm J.H. Rust, Saleh Yazdani, Peter Wriggers

CS 960: MESHLESS METHODS

THREE-DIMENSIONAL CRACK PROPAGATION ANALYSIS USING MESHLESS POINT COLLOCATION METHOD 8115
Eiji Tanaka

A CELL-BASED SMOOTHED POINT INTERPOLATION METHOD FOR AXISYMMETRIC PROBLEMS 8133
Arash Tootoonchi, Arman Khoshghalb

EFFICIENT SHAPE FUNCTION COSTRUCTION OF EFG MESHLESS METHOD 8142
P. Metsis, M. Papadrakakis

CS 980: NUMERICAL AND SYMBOLIC COMPUTATION
MODERNIZING SOFTWARE IN SCIENCE AND ENGINEERING: FROM C/C++ APPLICATIONS TO MOBILE PLATFORMS 8162
Liliana Favre, Claudia Pereira, Liliana Martinez

NEW EDUCATIONAL MATLAB TOOL TO EXPLAIN TWO DIMENSIONAL FINITE ELEMENT METHOD TO GRADUATE STUDENTS 8177
Rafet Sisman, Abdurrahman Sahin

CONTRIBUTION OF NUMERICAL THERMAL SIMULATION OF MASONRY TO THE GLOBAL OBJECTIVE OF BUILDINGS ENERGY EFFICIENCY 8187
Hipólito Sousa, Rui Sousa, Luís Silva, Gonçalo Sousa

CS 990: PARTICLE-BASED METHODS
TOWARDS MULTIPHYSICS SIMULATION OF DEEP PENETRATION LASER WELDING USING SMOOTHED PARTICLE HYDRODYNAMICS 8196
Haoyue Hu, Peter Eberhard, Florian Fetzer, Peter Berger

MULTIPLE TIMESCALE MODELLING OF PARTICLE SUSPENSIONS IN METAL MELTS SUBJECTED TO EXTERNAL FORCES 8207
Anton Manoylov, Georgi Djambazov, Valdis Bojarevics, Koulis Pericleous

STATISTICAL DISTRIBUTION FUNCTION OF CHARGED PARTICLES IN MAGNETIC FIELD 8224
B.Chetverushkin, Nicola D'Ascenzo, Valeri Saveliev

A STABILISED TOTAL LAGRANGIAN CORRECTED SMOOTH PARTICLE HYDRODYNAMICS TECHNIQUE IN LARGE STRAIN EXPLICIT FAST SOLID DYNAMICS 8231
Giorgio Greto, Sivakumar Kulasegaram, Chun H. Lee, Antonio J. Gil, Javier Bonet

BLOOD FLOW SIMULATION USING SMOOTHED PARTICLE HYDRODYNAMICS 8241
Mohammed Al-Saad, Sivakumar Kulasegaram, Stephane P.A. Bordas

CS 1010: COMPUTATIONAL INVERSE PROBLEMS AND OPTIMIZATION
ON THE VALIDITY OF TIDAL TURBINE ARRAY CONFIGURATIONS OBTAINED FROM STEADY-STATE ADJOINT OPTIMISATION 8247
Christian T. Jacobs, Matthew D. Piggott, Stephan C. Kramer, Simon W. Funke

DEVELOPMENT OF ACCURATE PNEUMATIC TYRE FINITE ELEMENT MODELS BASED ON AN OPTIMISATION PROCEDURE 8262
Chrysostomos-Alexandros Bekakos, George Papazafeiropoulos, Dan J. O'Boy, Jan Prins

METEOROLOGICAL DATA ASSIMILATION USING AN ADJOINT PROGRAM GENERATED BY AUTOMATIC DIFFERENTIATION 8275
Yasuyoshi Horibata

EVOLUTIONARY TOPOLOGY OPIMIZATION USING PARAMETERIZED B-SPLINE SURFACE 8283
Igor Pehnec, Damir Vucina, Frane Vlak

GEARBOX DESIGN VIA MIXED-INTEGER PROGRAMMING 8294
Bastian Dörig, Thorsten Ederer, Peter F. Pelz, Marc E. Pfetsch, Jan Wolf

A FRAMEWORK FOR THE DESIGN BY OPTIMIZATION OF HYDROFOILS UNDER CAVITATING CONDITIONS 8305
Paolo Olivucci, Stefano Gaggero

A COMPUTATIONAL STRATEGY FOR TRAJECTORY OPTIMIZATION OF UNDERACTUATED MULTIBODY SYSTEMS WITH CONTACTS 8324
Silvia Manara, Alessio Artoni, Marco Gabiccini

ANALYSIS OF NON-LINEAR ENVIRONMENTAL LOAD COMBINATIONS BY EXTENDED CONTOUR-LINE ALGORITHMS 8337
Bernt Leira

ADJOINT-BASED ROBUST OPTIMIZATION USING POLYNOMIAL CHAOS EXPANSIONS 8351
Joao Miranda, Dinesh Kumar, Chris Lacor

APPLICATION OF THE DECONVOLUTION METHODS FOR PROCESSING OF MEASUREMENT SIGNALS IN THE FAST PROCESSES 8365
Marat Goldfeld, Valery Pickalov

INVERSION METHODS BASED ON KALMAN FILTERING FOR IDENTIFICATION IN TUNNELING PROBLEMS 8378
Tamara Nestorović, Luan T. Nguyen

CS 1020: EVOLUTIONARY AND DETERMINISTIC METHODS FOR DESIGN, OPTIMIZATION AND CONTROL
EVALUATION OF DIFFERENT APPROACHES FOR THE OPTIMIZATION OF LAYOUT AND CONTROL OF BOOSTER STATIONS 8391
Philipp Pöttgen, Peter F. Pelz

MANIFESTATION OF ASCENDANCY OF EXTINCTION-REIGNITION ON SOUNDING HYBRID ROCKET USING DESIGN INFORMATICS 8403
Kazuhisa Chiba, Masahiro Kanazaki, Toru Shimada

COMPUTING AN OPERATING STRATEGY FOR AN ACTIVE BODY CONTROL WITH DYNAMIC PROGRAMMING IN THE AUTOMOTIVE AREA 8419
Marlene Utz, Phillipp Hedrich, Peter F. Pelz

SEISMIC RELIABILITY-BASED DESIGN OF STRUCTUTRES ISOLATED BY FPS 8431
Paolo Castaldo, Guglielmo Amendola, Bruno Palazzo

SEISMIC RETROFIT OF EXISTING BUILDINGS THROUGH THE DISSIPATIVE COLUMNS 8445
Paolo Castaldo, Bruno Palazzo, Francesco Perri, Ivana Marino, Marco Maria Faraco

AN EXAMPLE OF ENERGY DISSIPATION OPTIMIZATION FOR STEEL MRFS WITH PIN-JOINTED COLUMN BASES 8460
Elide Nastri, Annabella Paciello

ECCENTRICALLY BRACED FRAMES DESIGNED FOR THE ENERGY DISSIPATION OPTIMIZATION 8476
Elide Nastri

AN EXAMPLE OF COLUMN DESIGN OPTIMIZATION FOR FAILURE MODE CONTROL OF REINFORCED CONCRETE FRAMES 8492
Roberta Muscati

A COMPARISON OF MILP AND MINLP SOLVER PERFOMANCE ON THE EXAMPLE OF A DRINKING WATER SUPPLY SYSTEM DESIGN PROBLEM 8509
Lea Rausch, Philipp Leise, Thorsten Ederer, Lena C. Altherr, Peter F. Pelz

RELIABILITY TEST DEMONSTRATION METHOD FOR EXPONENTIAL LIFE SYSTEM WITH RELIABILITY GROWTH UNDER THE CONDITION OF IN-TIME CORRECTIVE STRATEGY 8528
Yunyan Xing, Ping Jiang, Feng Yao, Zhiwei Yang, Michael Emmerich, Thomas Bäck

CS 1100: REDUCTION ORDER METHODS
EVALUATION OF LINE-FITTING METHOD OF MODEL ORDER REDUCTION 8539
Valery Makhavikou, Roland Kasper, Dmitry Vlasenko

AN EFFICIENT ORDER REDUCTION STRATEGY IN EARTHQUAKE NONLINEAR RESPONSE ANALYSIS OF STRUCTURES 8552
Franz Bamer, Abbas Kazemi Amiri, Christian Bucher

ACTIVE SUBSPACES FOR THE PRELIMINARY FLUID DYNAMIC DESIGN OF UNCONVENTIONAL TURBOMACHINERY 8572
Juan S. Bahamonde, Matteo Pini, Piero Colonna

EXTENSION AND APPLICATION OF A NONLINEAR REDUCED ORDER MODEL TO GUST LOAD PREDICTION IN TIME DOMAIN 8587
Christoph Strobach, Klemens Lindhorst, Matthias Haupt, Peter Horst

CS 1200: STRUCTURAL DYNAMICS
RESEARCH ON HYSTERETIC BEHAVIORS OF A SAPARATED SHOCK ABSORBER APPLIED IN RAILWAY BRIDGE 8597
Aili Li, Ri Gao, Mingde Sun, Jilei Zhang

FINITE ELEMENT ANALYSIS OF TIMBER BEAMS WITH FLAWS 8606
Janka Kovacikova, Mats Ekevad, Olga Ivankova, Sven Berg

NONLINEAR FORCED RESPONSE OF A STATOR VANE WITH MULTIPLE FRICTION CONTACTS USING A COUPLED STATIC/DYNAMIC APPROACH 8612
Marco Lassalle, Christian Maria Firrone

COMPUTATIONAL METHOD OF DETERMINATION OF INTERNAL EFFORTS IN LINKS OF MECHANISMS AND ROBOT MANIPULATORS WITH STATICALLY DEFINABLE STRUCTURES CONSIDERING THE DISTRIBUTED DYNAMICALLY LOADINGS 8627
Zhumadil Baigunchekov, Muratulla Utenov, Nurzhan Utenov, Saltanat Zhilkibayeva

FE MODELLING OF SFRC BEAMS UNDER IMPACT LOADS 8640
Pegah Behinaein, Ali Abbas, Demetris Cotsovos

AN APROXIMATE ANALYTICAL SOLUTION FOR NONLINEAR FGM SHELL STRUCTURE WITH VARIABLE IN TIME PARAMETERS 8654
Victor Gristchak, Yuliia Fatieieva

UNCERTAINTY OF MODELS IN INTELLIGENT SYSTEMS UNDER STOCHASTIC LOADING 8665
Amalia Moutsopoulou, Georgios Stavroulakis, Tasos Pouliezos

COMPUTATIONAL MODELLING OF COLD-FORMED STEEL SCREWED CONNECTIONS AT AMBIENT AND ELEVATED TEMPERATURES 8678
Luis Mesquita, Rui Dias, Armandino Parente, Paulo Piloto

MONITORING AND ANALYSIS OF STRESS FIELD FOR ORTHOTROPIC STEEL DECK OF DASHENGGUAN YANGZTE RIVER BRIDGE 8691
Ying Wang, Y.S. Song

CS 1201: COMPUTATIONAL SOIL MECHANICS
A MORE COMPREHENSIVE MODELING OF CONTACT FORCE DURING SHEAR TESTING USING DEM 8708
Varvara Roubtsova, Mohamed Chekired

THREE-DIMENSIONAL BE-FE MODEL OF BUCKET FOUNDATIONS IN POROELASTIC SOILS 8725
Jacob D. R. Bordón, Juan J. Aznárez, Orlando F. Maeso

CS 1202: STRUCTURAL ANALYSIS AND MULTI BODY DYNAMICS
ON THE NUMERICAL INFLUENCES OF INERTIA REPRESENTATION FOR RIGID BODY DYNAMICS 8739
Xiaoming Xu, Wanxie Zhong

APPROACHES TO THE CREATION OF MULTIBODY MODELS OF THE VVER 1000 NUCLEAR REACTOR CONTROL ASSEMBLY 8756
Pavel Polach, Michal Hajžman

CS 1300: UNCERTAINTY QUANTIFICATION AND ERROR ESTIMATION
THE UPLIFT CAPACITY OF HORIZONTAL PLATE ANCHORS IN SPATIALLY VARIABLE CLAY USING SPARSE POLYNOMIAL CHAOS EXPANSIONS 8769
Tom Charlton, Mohamed Rouainia

FATIGUE RELIABILITY OF AGEING RAILWAY BRIDGES: FEASIBILITY OF PROBABILISTIC APPROACH 8778
Nirosha D. Adasooriya

AN EFFICIENT AERODYNAMIC SHAPE OPTIMIZATION FRAMEWORK FOR ROBUST DESIGN OF AIRFOILS USING SURROGATE MODELS 8787
Daigo Maruyama, Dishi Liu, Stefan Görtz

QUANTIFICATION OF SPATIAL VARIABILITY FOR TRANSVERSE ELASTIC MODULUS OF SPRUCE WOOD …….. 8801
Alireza Farajzadeh Moshtaghin, Steffen Franke, Thomas Keller, Anastasios Vassilopoulos

COMPARISON BETWEEN A POLYNOMIAL CHAOS SURROGATE MODEL AND MARKOV CHAIN MONTE CARLO FOR INVERSE UNCERTAINTY QUANTIFICATION BASED ON AN ELECTRIC DRIVE TEST BENCH …….. 8809
Philipp Glaser, Michael Schick, Kosmas Petridis, Vincent Heuveline

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NONPARAMETRIC PROBABILISTIC APPROACH OF MODEL UNCERTAINTIES INTRODUCED BY A PROJECTION-BASED NONLINEAR REDUCED-ORDER MODEL

C. Soize[1], **C. Farhat**[2]

[1]Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS,
Université Paris-Est, 5 bd Descartes, 77454 Marne-la-Vallee, France
e-mail: christian.soize@univ-paris-est.fr

[2] Department of Aeronautics and Astronautics, Department of Mechanical Engineering, and Institute for Computational and Mathematical Engineering,
Stanford University, Stanford, CA 94305, USA
e-mail: cfarhat@stanford.edu

Keywords: Modeling errors, Model uncertainties, Nonparametric probabilistic approach, Reduced-order model, Model order reduction, Uncertainty quantification.

Abstract. *The paper is devoted to model uncertainties (or model form uncertainties) induced by modeling errors in computational sciences and engineering (such as in computational structural dynamics, fluid-structure interaction, and vibroacoustics, etc.) for which a parametric high-fidelity computational model (HFM) is used for addressing optimization problems (such as a robust design optimization problem), which are solved by introducing a parametric reduced-order model (ROM) constructed using an adapted reduced-order basis (ROB) derived from the parametric HFM. Two main methodologies are available to take into account such modeling errors. The first one is the usual output-predictive error method that has been introduced for many years. This approach can induce some difficulties because the parametric HFM and ROM do not learn from data. The second one is the nonparametric probabilistic approach of model uncertainties introduced in the framework of structural dynamics fifteen years ago. This approach is adapted, but is mainly limited to linear operators of the parametric HFM. The present paper deals with this challenging problem and proposes a novel nonparametric probabilistic approach of the modeling errors for any parametric nonlinear HFM for which a parametric nonlinear ROM can be constructed from the HFM. The methodology proposed consists in substituting the deterministic ROB with a stochastic ROB for which the probability measure in constructed on a subset of a compact Stiefel manifold. The stochastic model depends on a small number of hyperparameters for which the identification is performed by solving a statistical inverse problem. An application is presented in nonlinear computational structural dynamics.*

1 INTRODUCTION

1.1 Short overview on the high-dimensional and projection-based reduced-order models

The potential of Partial Differential Equation (PDE)-based, High-Dimensional computational Models (HDMs) for enhancing system performance and predicting the unknown is recognized in almost every field of science and engineering. However, in many computational mechanics applications, nonlinear, time-dependent numerical simulations based on HDMs remain so cost-prohibitive that they cannot be used as often as needed. For this reason, nonlinear, projection-based Model Order Reduction (MOR) has recently emerged as a promising if not indispensable numerical tool for parametric applications such as, to name only a few, design, design optimization, statistical analysis, and simulation-based decision making [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. In general, a $\boldsymbol{\mu}$-parametric, high-fidelity, physics-based computational model is high-dimensional because its underlying spatial discretization is performed *a priori*. On the other hand, the projection of such a computational model of dimension N onto a subspace of low dimension $n \ll N$ represented by an associated $(N \times n)$ Reduced-Order Basis (ROB) $[V]$ leads to a Reduced-Order Model (ROM) of much lower dimension n. When $[V]$ is carefully constructed *a posteriori*, the corresponding ROM can capture the dominant behavior of the underlying $\boldsymbol{\mu}$-parametric HDM and therefore retain most of its fidelity. In general, knowledge about the system response is obtained during a *training* procedure that is performed *offline*. During this procedure, the model parameters represented here by the parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{N_p})$ belonging to the parameter space $\mathcal{C}_{\boldsymbol{\mu}}$ are sampled at a *few* points using a greedy but effective sampling strategy (for example, see [11]), and a set of problems related to the main problem of interest are solved to obtain a set of parametric solution snapshots. Then, these snapshots are compressed using, for example, the Singular Value Decomposition (SVD) to construct a *global* ROB. In general, the sampling strategy is designed so that the global ROB is reliable in a large region of the model parameter domain. Unfortunately, despite its low dimension, the resulting global (or $\boldsymbol{\mu}$-parametric) ROM does not necessarily guarantee computational feasibility. This is because the construction of this projection-based ROM scales not only with its size n, but also with that of the underlying HDM, $N \gg n$. In the deterministic setting, this issue is particularly problematic for nonlinear problems because the ROM needs to be repeatedly reconstructed to address, for example, time-dependency or Newton iterations for implicit solution strategies. This caveat is remedied by equipping a MOR method with a rigorous procedure for approximating the resulting reduced operators whose computational complexity scales only with the small size n of the ROM [1, 3, 12, 13]. Such a procedure is also known in the literature as hyper reduction [14]. It transforms the nonlinear ROM into a hyper reduced ROM that guarantees feasibility, while maintaining as much as possible a desired level of accuracy. From these reasons, a nonlinear ROM or hyper reduced ROM inherits the modeling errors and associated uncertainties of its underlying HDM, including model form uncertainties. It is also tainted by additional errors introduced by the reduction processes highlighted above. Hence, if MOR is essential for enabling simulation-based decision making, Uncertainty Quantification (UQ) for ROMs is critical for certifying the decisions they enable.

1.2 Parametric probabilistic approach of uncertainties using the $\boldsymbol{\mu}$-parametric ROM

The $\boldsymbol{\mu}$-parametric ROM has been used in the context of the parametric probabilistic approach of uncertainties, which consists in constructing prior and posterior stochastic models of the uncertain model-parameters (geometry, boundary conditions, material properties, etc) of the $\boldsymbol{\mu}$-parametric HFM and of the associated $\boldsymbol{\mu}$-parametric ROM (see for instance [15, 16, 17, 18,

19, 20, 21, 22, 23, 24, 25]). Such an approach is very well adapted and very efficient to take into account the uncertainties in the $\boldsymbol{\mu}$-parametric HFM and/or in the $\boldsymbol{\mu}$-parametric ROM (see for instance [26, 4, 27]), and for large scale statistical inverse problems, see [28, 29, 30, 31, 32, 33, 34, 35]). However, the parametric probabilistic approach has not the capability to take into account model uncertainties induced by modeling errors that are introduced during the construction of the $\boldsymbol{\mu}$-parametric HFM and of the associated $\boldsymbol{\mu}$-parametric ROM [25, 27, 36].

1.3 Nonparametric probabilistic approach of uncertainties using the μ-parametric ROM

The nonparametric probabilistic approach of model uncertainties induced by modeling errors has been introduced in 1999 [36, 37, 38, 39] for linear structural dynamic systems. It is organized around two steps. The first one consists in constructing a linear reduced-order model (ROM) of dimension n, from a linear HFM with N degrees-of-freedom (dofs), by using an adapted ROB represented by the $(N \times n)$ matrix $[V]$. The second step consists in constructing a SROM by substituting the deterministic matrices of the ROM (such as the mass, the damping, and the stiffness reduced matrices) with random matrices for which the probability distributions are constructed using the Maximum Entropy (MaxEnt) principle [40, 41, 42] under the constraints defined by available information assocaited with algebraic properties (such as lower bounds, positiveness, integrability of the inverse, etc) and statistical information (such as the mean value equals to the nominal values, etc), and for which advanced algorithms have been developed for the high dimensions [39, 43, 44]. This approach has been extended to different ensembles of random matrices (see [38, 45]), static boundary value problems [46], and has been experimentally validated and applied in many areas, including: dynamics of composite structures [47] and viscoelastic structures [48, 49], dynamic substructuring techniques [50, 51, 52, 53, 54], vibroacoustic systems [48, 55, 56, 57], soil-structure interactions and earthquake engineering [58, 59, 60], robust design and optimization [61, 62], to name only a few. More recently, this nonparametric probabilistic approach has been extended in structural dynamics to nonlinear geometrical effects [63, 64]. Such an extension is very efficient but is strongly related to the mathematical properties of the nonlinear elasticity operator. Such an extension cannot, *a priori*, be carried out for any nonlinear operator and the nonparametric probabilistic approach of model uncertainties induced by modeling errors has not received yet a solution for general nonlinear dynamical systems in computational solid mechanics and computational fluid dynamics. Hence, the objective of this paper is to propose a nonparametric probabilistic approach for such nonlinear dynamical systems, for which the framework is detailed below.

1.4 Model uncertainties induced by the use of a μ-independent ROB for constructing the μ-parametric ROM

As explained at the end of Section 1.1, despite all the mathematical analyses and the advanced methodologies that have been developed, and despite all the precautions taken to build a $\boldsymbol{\mu}$-parametric nonlinear ROM from a $\boldsymbol{\mu}$-parametric nonlinear HFM, the efficiency of the $\boldsymbol{\mu}$-parametric nonlinear ROM depends on the sampling points $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu}$ chosen for $\boldsymbol{\mu}$ in $\mathcal{C}_{\boldsymbol{\mu}}$ for constructing the ROB, $[V]$, and depends on the accuracy of the constructed approximation. In other words, it depends of the dimension n chosen for the $\boldsymbol{\mu}$-parametric ROM. Even if the $\boldsymbol{\mu}$-parametric nonlinear ROM that is constructed has the capability to give a good approximation of the $\boldsymbol{\mu}$-parametric nonlinear HFM, the CPU-time consideration leads to limiting the size m_μ of the sampling points for $\boldsymbol{\mu}$ and the dimension n of the ROB, which induces an error between the

$\boldsymbol{\mu}$-parametric nonlinear ROM and the $\boldsymbol{\mu}$-parametric nonlinear HFM when $\boldsymbol{\mu}$ runs through $\mathcal{C}_{\boldsymbol{\mu}}$.

1.5 Objective and organization of the paper

To this effect, the objective of this paper is to propose a nonparametric probabilistic approach for model uncertainties induced by two types of modeling errors:

- The first type of modeling errors is associated with the usage, for the purpose of computational efficiency, of the $\boldsymbol{\mu}$-parametric nonlinear ROM as a surrogate of the $\boldsymbol{\mu}$-parametric nonlinear HFM. The dimension $m_{\boldsymbol{\mu}}$ of the parameter $\boldsymbol{\mu}$ can be small or large. When the dimension n of the $\boldsymbol{\mu}$-parametric nonlinear ROM goes to N and the number m_{μ} of the sampling points of $\boldsymbol{\mu}$ in $\mathcal{C}_{\boldsymbol{\mu}}$ (used for constructing $[V]$) goes to infinity, the modeling error between the $\boldsymbol{\mu}$-parametric nonlinear ROM and the $\boldsymbol{\mu}$-parametric nonlinear HFM goes to zero.

- The second type of modeling errors is due to the modeling errors introduced in the construction of the $\boldsymbol{\mu}$-parametric nonlinear HFM itself. This means that the HFM does not perfectly predict the experimental data for the quantities of interest.

Consequently, the distance between the predictions of the $\boldsymbol{\mu}$-parametric ROM and the experimental data are due to the two types of modeling errors.

In this paper, we give a short presentation of a novel nonparametric probabilistic approach that is detailed in [65]. Section 2 deals with a presentation of the nonparametric probabilistic approach for model uncertainties induced by the modeling errors in the nonlinear reduced-order models. Section 3 is devoted to the probabilistic construction of the stochastic reduced-order basis (SROB). The theory is presented in a discrete form that can be immediately exploited by computational models. Section 4 is devoted to a simple application in nonlinear computational structural dynamics that is easy to reproduce by the interested reader.

2 NONPARAMETRIC PROBABILISTIC APPROACH FOR UNCERTAINTY ANALYSIS OF NONLINEAR MODEL ORDER REDUCTION METHODS

An example of a $\boldsymbol{\mu}$-parametric nonlinear HFM is chosen in order to explain what is the nonparametric probabilistic approach proposed for taking into account the two types of modeling errors:
$\diamond$ the errors induced by the use of the $\boldsymbol{\mu}$-parametric nonlinear ROM instead of the $\boldsymbol{\mu}$-parametric nonlinear HFM,
$\diamond$ the modeling errors introduced during the construction of the $\boldsymbol{\mu}$-parametric nonlinear HFM.

2.1 Description of the $\boldsymbol{\mu}$-parametric nonlinear HFM in the field of nonlinear structural dynamics

We consider a $\boldsymbol{\mu}$-parametric nonlinear computational dynamical model on $\mathbb{R}^N$ corresponding to the finite element discretization of a structure,

$$[M]\,\ddot{\mathbf{y}}(t) + \mathbf{g}(\mathbf{y}(t), \dot{\mathbf{y}}(t); \boldsymbol{\mu}) = \mathbf{f}(t; \boldsymbol{\mu}) \quad , \quad t \in]t_0, T]\,, \tag{1}$$

with the initial conditions

$$\mathbf{y}(t_0) = \mathbf{y}_0 \quad , \quad \dot{\mathbf{y}}(t_0) = \mathbf{y}_1\,, \tag{2}$$

and with $N_{\text{cd}} < N$ constraint equations that are written in a matrix form as

$$[B]^T \mathbf{y}(t) = \mathbf{0}_{N_{\text{cd}}} \quad , \quad t \in [t_0, T] \, , \tag{3}$$

in which t_0 and T are given such that $-\infty < t_0 < T < +\infty$, where $\mathbf{y}_0$ and $\mathbf{y}_1$ are two given vectors in $\mathbb{R}^N$ satisfying the constraint equation. In Eq. (1), $\boldsymbol{\mu}$ is the parameter that belongs to $\mathcal{C}_{\boldsymbol{\mu}} \subset \mathbb{R}^{m_{\boldsymbol{\mu}}}$ introduced before, t denotes the time variable, $\mathbf{y}(t) = (y_1(t), \ldots, y_N(t))$ is the $\mathbb{R}^N$ vector of the N dofs, $\dot{\mathbf{y}}(t) = d\mathbf{y}(t)/dt$ and $\ddot{\mathbf{y}}(t) = d^2\mathbf{y}(t)/dt^2$ are the velocity and the acceleration vectors, $[M]$ is the mass matrix that belongs to $\mathbb{M}_N^+$ and which is assumed to be independent of t and $\boldsymbol{\mu}$, where $\mathbf{g}(\mathbf{y}(t), \dot{\mathbf{y}}(t); \boldsymbol{\mu})$ is the $\mathbb{R}^N$ vector representing the internal forces at time t, which depends on $\mathbf{y}(t)$, $\dot{\mathbf{y}}(t)$, and $\boldsymbol{\mu}$, and finally, where $\mathbf{f}(t; \boldsymbol{\mu})$ is the $\mathbb{R}^N$ vector of the external forces at time t, which depends on parameter $\boldsymbol{\mu}$. In Eq. (3), $[B]$ is a given matrix in $\mathbb{M}_{N,N_{\text{cd}}}$, which defines the N_{cd} constraints on $\mathbf{y}$, and which is assumed to be independent of t and $\boldsymbol{\mu}$ and such that $[B]^T[B] = [I_{N_{\text{cd}}}]$. The $\mathbb{R}^N$-valued solution $\{\mathbf{y}(t; \boldsymbol{\mu}), t \in [t_0, T]\}$ of Eqs. (1) to (3) depends on $\boldsymbol{\mu}$. At time t, the quantity of interest (QoI) is a vector $\mathbf{o}(t; \boldsymbol{\mu}) = (o_1(t; \boldsymbol{\mu}), \ldots, o_{m_{\mathbf{o}}}(t; \boldsymbol{\mu}))$ with values in $\mathbb{R}^{m_{\mathbf{o}}}$, depending on $\boldsymbol{\mu}$, which is written as

$$\mathbf{o}(t; \boldsymbol{\mu}) = \mathbf{h}(\mathbf{y}(t; \boldsymbol{\mu}), \dot{\mathbf{y}}(t; \boldsymbol{\mu}), \mathbf{f}(t; \boldsymbol{\mu}), t; \boldsymbol{\mu}) \quad , \quad t \in [t_0, T] \, , \tag{4}$$

in which $\mathbf{h}$ is a given mapping.

2.2 Construction of a μ-parametric nonlinear ROM associated with the μ-parametric nonlinear HFM

Let $[V] \in \mathbb{M}_{N,n}$ be a ROB independent of $\boldsymbol{\mu}$, where $n \ll N$ is the dimension of the ROB that is assumed to be constructed for representing the solution $\{\mathbf{y}(t; \boldsymbol{\mu}), t \in [t_0, T]\}$ for all $\boldsymbol{\mu} \in \mathcal{C}_{\boldsymbol{\mu}}$ (as explained before), and satisfies $[V]^T[M]\,[V] = [I_n]$. By construction, this ROB satisfies also the constraint equation $[B]^T[V] = [0_{N_{\text{cd}},n}]$. It is assumed that the $\boldsymbol{\mu}$-parametric nonlinear ROM is constructed via the Galerkin method which yields

$$\mathbf{y}^{(n)}(t) = [V]\,\mathbf{q}(t) \quad , \quad t \in [t_0, T] \, , \tag{5}$$

$$\ddot{\mathbf{q}}(t) + [V]^T \mathbf{g}([V]\,\mathbf{q}(t), [V]\,\dot{\mathbf{q}}(t); \boldsymbol{\mu}) = [V]^T \mathbf{f}(t; \boldsymbol{\mu}) \quad , \quad t \in]t_0, T] \, , \tag{6}$$

with the initial conditions

$$\mathbf{q}(t_0) = [V]^T[M]\,\mathbf{y}_0 \quad , \quad \dot{\mathbf{q}}(t_0) = [V]^T[M]\,\mathbf{y}_1 \, . \tag{7}$$

The $\mathbb{R}^n$-valued solution $\{\mathbf{q}(t; \boldsymbol{\mu}) \, , \, t \in [t_0, T]\}$ of Eqs. (6) and (7) depends on $\boldsymbol{\mu}$, and $\{\mathbf{y}^{(n)}(t; \boldsymbol{\mu}), t \in [t_0, T]\}$ is the n-order approximation of $\{\mathbf{y}(t; \boldsymbol{\mu}) \, , \, t \in [t_0, T]\}$. The corresponding approximation $\mathbf{o}^{(n)}$ of $\mathbf{o}$ is given by

$$\mathbf{o}^{(n)}(t; \boldsymbol{\mu}) = \mathbf{h}(\mathbf{y}^{(n)}(t; \boldsymbol{\mu}), \dot{\mathbf{y}}^{(n)}(t; \boldsymbol{\mu}), \mathbf{f}(t; \boldsymbol{\mu}), t; \boldsymbol{\mu}) \quad , \quad t \in [t_0, T] \quad , \quad \boldsymbol{\mu} \in \mathcal{C}_{\boldsymbol{\mu}} \, . \tag{8}$$

For a given reduced-order n, the error induced by the use of the $\boldsymbol{\mu}$-parametric nonlinear ROM instead of the $\boldsymbol{\mu}$-parametric nonlinear HFM could, *a priori*, be estimated (in the L^2 sense) by

$$\varepsilon(n) = \int_{\mathcal{C}_{\boldsymbol{\mu}}} \int_{t_0}^{T} \|\mathbf{o}(t; \boldsymbol{\mu}) - \mathbf{o}^{(n)}(t; \boldsymbol{\mu})\|^2 \, dt \, d\boldsymbol{\mu} \, . \tag{9}$$

It should be noted that the estimation of $\varepsilon(n)$ defined by Eq. (9) must require the computation of the solution of the HFM for a very large number of sampled values of the vector-valued parameter $\boldsymbol{\mu}$ in $\mathcal{C}_{\boldsymbol{\mu}}$ in order to correctly estimate the integral in $\boldsymbol{\mu}$ over $\mathcal{C}_{\boldsymbol{\mu}}$. In practice, such a computation cannot be done and Eq. (9) must be replaced by

$$\varepsilon(n) = \sum_{i=1}^{m_\mu} \int_{t_0}^{T} \|\mathbf{o}(t;\boldsymbol{\mu}_i) - \mathbf{o}^{(n)}(t;\boldsymbol{\mu}_i)\|^2 \, dt \,, \tag{10}$$

in which $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu}$ are the sampling points used for constructing the ROB $[V]$.

2.3 Construction of a stochastic ROB (SROB) with the proposed nonparametric probabilistic approach

The proposed nonparametric probabilistic approach of model uncertainties induced by modeling errors consists in substituting the deterministic ROB $[V]$ with a stochastic ROB $[\mathbf{W}]$ that is independent of $\boldsymbol{\mu}$, which is a random matrix with values in $\mathbb{M}_{N,n}$ for which the support of its probability distribution is the subset of $\mathbb{M}_{N,n}$ corresponding to the constraints $[\mathbf{W}]^T[M]\,[\mathbf{W}] = [I_n]$ and $[B]^T[\mathbf{W}] = [0_{N_{\text{cb}},n}]$ almost surely. The probability distribution of the random matrix $[\mathbf{W}]$ depends on a vector-valued hyperparameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{m_{\boldsymbol{\alpha}}})$ belonging to a subset $\mathcal{C}_{\boldsymbol{\alpha}}$ of $\mathbb{R}^{m_{\boldsymbol{\alpha}}}$ where the dimension $m_{\boldsymbol{\alpha}}$ will be chosen small in order for the statistical inverse problem for identifying $\boldsymbol{\alpha}$ to be feasible. The construction of the SROB is presented in Section 3.

2.4 Construction of the μ-parametric nonlinear SROM associated with the μ-parametric nonlinear ROM using the proposed nonparametric probabilistic approach

The $\boldsymbol{\mu}$-parametric nonlinear SROM associated with the $\boldsymbol{\mu}$-parametric nonlinear ROM is deduced from Eqs. (5) to (8) by replacing $[V]$ with the random matrix $[\mathbf{W}]$. Consequently, $\mathbf{y}^{(n)}$, $\mathbf{q}$, and $\mathbf{o}^{(n)}$ become the stochastic processes $\mathbf{Y}^{(n)}$, $\mathbf{Q}$, and $\mathbf{O}^{(n)}$, and the SROM is written as

$$\mathbf{Y}^{(n)}(t) = [\mathbf{W}]\,\mathbf{Q}(t) \quad , \quad t \in [t_0, T] \,, \tag{11}$$

$$\ddot{\mathbf{Q}}(t) + [\mathbf{W}]^T \mathbf{g}([\mathbf{W}]\,\mathbf{Q}(t), [\mathbf{W}]\,\dot{\mathbf{Q}}(t); \boldsymbol{\mu}) = [\mathbf{W}]^T \mathbf{f}(t;\boldsymbol{\mu}) \quad , \quad t \in]t_0, T] \,, \tag{12}$$

with the initial conditions

$$\mathbf{Q}(t_0) = [\mathbf{W}]^T[M]\,\mathbf{y}_0 \quad , \quad \dot{\mathbf{Q}}(t_0) = [\mathbf{W}]^T[M]\,\mathbf{y}_1 \,. \tag{13}$$

The $\mathbb{R}^n$-valued stochastic solution $\{\mathbf{Q}(t;\boldsymbol{\mu},\boldsymbol{\alpha})\,,\, t \in [t_0,T]\}$ of Eqs. (12) and (13) depends on $\boldsymbol{\mu} \in \mathcal{C}_{\boldsymbol{\mu}}$ and $\boldsymbol{\alpha} \in \mathcal{C}_{\boldsymbol{\alpha}}$. The stochastic process $\{\mathbf{Y}^{(n)}(t;\boldsymbol{\mu},\boldsymbol{\alpha})\,,\, t \in [t_0,T]\}$ is the n-order approximation of stochastic process $\{\mathbf{Y}(t;\boldsymbol{\mu})\,,\, t \in [t_0,T]\}$. The corresponding approximation $\{\mathbf{O}^{(n)}(t;\boldsymbol{\mu},\boldsymbol{\alpha})\,,\, t \in [t_0,T]\}$ of the random quantity of interest $\{\mathbf{O}(t;\boldsymbol{\mu})\,,\, t \in [t_0,T]\}$ is given, for all $t \in [t_0,T]$, $\boldsymbol{\mu} \in \mathcal{C}_{\boldsymbol{\mu}}$, and $\boldsymbol{\alpha} \in \mathcal{C}_{\boldsymbol{\alpha}}$, by

$$\mathbf{O}^{(n)}(t;\boldsymbol{\mu},\boldsymbol{\alpha}) = \mathbf{h}(\mathbf{Y}^{(n)}(t;\boldsymbol{\mu},\boldsymbol{\alpha}), \dot{\mathbf{Y}}^{(n)}(t;\boldsymbol{\mu},\boldsymbol{\alpha}), \mathbf{f}(t;\boldsymbol{\mu}), t; \boldsymbol{\mu}) \,. \tag{14}$$

2.5 Identification of hyperparameter α of the probability distribution of $[\mathbf{W}]$

The identification of the hyperparameter $\boldsymbol{\alpha} \in \mathcal{C}_{\boldsymbol{\alpha}} \subset \mathbb{R}^{m_{\boldsymbol{\alpha}}}$ can be performed using the maximum likelihood method or a nonlinear least-squares method for the QoI. For instance, a nonlinear least-squares method can be formulated as follows for the two types of modeling errors. Let $J(\boldsymbol{\alpha})$ be the cost function defined on $\mathcal{C}_{\boldsymbol{\alpha}}$ by

$$J(\boldsymbol{\alpha}) = w_J \, J_{\text{mean}}\,(\boldsymbol{\alpha}) + (1 - w_J)\, J_{\text{std}}\,(\boldsymbol{\alpha}) \,, \tag{15}$$

in which w_J is a weight such that $0 \leq w_J \leq 1$, and where $J_{\text{mean}}(\boldsymbol{\alpha})$ and $J_{\text{std}}(\boldsymbol{\alpha})$ allow for controlling the identification of $\boldsymbol{\alpha}$ with respect to the mean value and the statistical fluctuations,

$$J_{\text{mean}}(\boldsymbol{\alpha}) = \frac{1}{c_{\text{mean}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu})} \sum_{i=1}^{m_\mu} \int_{t_0}^{T} \|\mathbf{o}^{\text{ref}}(t;\boldsymbol{\mu}_i) - E\{\mathbf{O}^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha})\}\|^2 \, dt \,, \tag{16}$$

$$J_{\text{std}}(\boldsymbol{\alpha}) = \frac{1}{c_{\text{std}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu})} \sum_{i=1}^{m_\mu} \int_{t_0}^{T} \|\mathbf{v}^{(\text{ref},n)}(t;\boldsymbol{\mu}_i) - \mathbf{v}^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha})\|^2 \, dt \,. \tag{17}$$

The positive constants $c_{\text{mean}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu})$ and $c_{\text{std}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu})$ are written as

$$c_{\text{mean}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu}) = \sum_{i=1}^{m_\mu} \int_{t_0}^{T} \|\mathbf{o}^{\text{ref}}(t;\boldsymbol{\mu}_i)\|^2 \, dt \,, \tag{18}$$

$$c_{\text{std}}(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu}) = \sum_{i=1}^{m_\mu} \int_{t_0}^{T} \|\mathbf{v}^{(\text{ref},n)}(t;\boldsymbol{\mu}_i)\|^2 \, dt \,, \tag{19}$$

where $\mathbf{v}^{(\text{ref},n)}(t;\boldsymbol{\mu}_i) = (v_1^{(\text{ref},n)}(t;\boldsymbol{\mu}_i), \ldots, v_{m_{\mathbf{o}}}^{(\text{ref},n)}(t;\boldsymbol{\mu}_i))$ is such that,

$$v_j^{(\text{ref},n)}(t;\boldsymbol{\mu}_i) = \gamma \, |o_j^{\text{ref}}(t;\boldsymbol{\mu}_i) - o_j^{(n)}(t;\boldsymbol{\mu}_i)| \quad , \quad j = 1, \ldots m_{\mathbf{o}} \,, \tag{20}$$

and where $\gamma > 0$ allows for controlling the amplitude of the target related to the statistical fluctuations. In Eq. (17), $\mathbf{v}^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha}) = (v_1^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha}), \ldots, v_{m_{\mathbf{o}}}^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha}))$ is such that

$$v_j^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha}) = \{\, E\{O_j^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha})^2\} - (E\{O_j^{(n)}(t;\boldsymbol{\mu}_i,\boldsymbol{\alpha})\})^2\} \,\}^{1/2} \quad , \quad j = 1, \ldots m_{\mathbf{o}} \,. \tag{21}$$

In Eqs. (16), (18) and (20), $\mathbf{o}^{\text{ref}}$ is defined as a function of the type of the modeling errors that are taken into account:

$\diamond$ If only the errors induced by the use of the $\boldsymbol{\mu}$-parametric nonlinear ROM instead of the $\boldsymbol{\mu}$-parametric nonlinear HFM are taken into account, then $\mathbf{o}^{\text{ref}} = \mathbf{o}$.

$\diamond$ If the two types of modeling errors are simultaneously taken into account (the modeling errors induced by the use of the $\boldsymbol{\mu}$-parametric nonlinear ROM instead of the $\boldsymbol{\mu}$-parametric nonlinear HFM and the modeling errors introduced during the construction of the $\boldsymbol{\mu}$-parametric nonlinear HFM), then $\mathbf{o}^{\text{ref}} = \mathbf{o}^{\text{exp}}$ in which $\mathbf{o}^{\text{exp}}$ corresponds to experimental data for the QoI.

With such a formulation, in the cost function,

$\diamond$ $\mathbf{o}^{\text{ref}}$ appears as the target for the mean value with a weight w_J,

$\diamond$ $\mathbf{v}^{(\text{ref},n)}$ appears as the target for the standard deviation with a weight $1 - w_J$.

The identification of hyperparameter $\boldsymbol{\alpha}$ consists in calculating $\boldsymbol{\alpha}^{\text{opt}}$ such that

$$\boldsymbol{\alpha}^{\text{opt}} = \min_{\boldsymbol{\alpha} \in \mathcal{C}_{\boldsymbol{\alpha}}} J(\boldsymbol{\alpha}) \,. \tag{22}$$

2.6 Justification

Why a random basis of the admissible set would be useful? Such a question is fully licit because it is well known that, with the Galerkin method, the solution $\mathbf{y}^{(n)}$ constructed using the ROM for an order n for which $\mathbf{y}^{(n)}$ is close to $\mathbf{y}$ (convergence of $\mathbf{y}^{(n)}$ towards the solution $\mathbf{y}$ of

the HFM when n goes to N), is independent of the choice of the vector basis of the admissible set (such a statement is perfectly true for $n = N$). For a similar reason, for any SROB of the admissible set, the stochastic solution $\mathbf{Y}^{(n)}$ constructed using the SROM converges almost surely towards the deterministic solution $\mathbf{y}$ of the HFM, which is thus, at convergence, independent of the choice of the stochastic basis. However, and as we have previously explained, for $n \ll N$, and taking into account that the ROB $[V]$ is constructed for $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m_\mu}$ in $\mathcal{C}_{\boldsymbol{\mu}}$, for all $\boldsymbol{\mu}$ fixed in $\mathcal{C}_{\boldsymbol{\mu}}$, the solution $\mathbf{y}^{(n)}$ of the ROM is only an approximation of $\mathbf{y}$, and consequently, there is an error introduced by the use of the ROM with order n (the existence of such an error is the main assumption of the method that is proposed). In such a case, for a fixed value of n for which the convergence is not reached, the value of the error depends on the choice of the basis of the admissible set (if the n vectors of the basis are extracted from another family that constitutes another vector basis of the admissible set, then the solution $\mathbf{y}^{(n)}$ of the ROM is modified). Consequently, when the basis is substituted by the SROB $[\mathbf{W}]$, a stochastic family $\mathbf{Y}^{(n)}$ is generated using the SROM. The idea is then to adapt the statistical fluctuations of $\mathbf{Y}^{(n)}$ for representing the error between the ROM and the HFM, and also, as it has been explained before, to represent the modeling errors introduced in the construction of the $\boldsymbol{\mu}$-parametric nonlinear HFM itself.

3 CONSTRUCTION OF A STOCHASTIC REDUCED-ORDER BASIS

The construction of the SROM is detailed in [65]. Only the main ideas of the construction are given below. The ROB corresponds to a compact Stiefel manifold denoted by $\mathbb{S}_{N,n}$. Since the nonparametric probabilistic approach consists in defining the SROB by a probability measure on $\mathbb{S}_{N,n}$, we have to construct a parameterization of $\mathbb{S}_{N,n}$, which must be efficient for the high dimensions.

3.1 Construction of a parameterization of the compact Stiefel manifold $\mathbb{S}_{N,n}$

Let $[M] \in \mathbb{M}_N^+$ be a given positive-definite symmetric $(N \times N)$ real matrix (possibly, $[M]$ can be $[I_N]$). The set $\mathbb{M}_{N,n}$ of all the $(N \times n)$ real matrices is considered as an Euclidean space equipped with the inner product $\ll [V_1], [V_2] \gg_M = \text{tr}\{[V_1]^T [M] [V_2]\}$ and the associated norm $\|V\|_M = \{\text{tr}\{[V]^T [M] [V]\}\}^{1/2}$. Let $[V]$ be a reduced-order basis (ROB) belonging to the compact Stiefel manifold $\mathbb{S}_{N,n}$ defined (see for instance [66]) by

$$\mathbb{S}_{N,n} = \{\ [V] \in \mathbb{M}_{N,n}\ ,\ [V]^T[M][V] = [I_n]\ \} \subset \mathbb{M}_{N,n}\,, \tag{23}$$

for which the dimension is $\nu_{\mathbb{S}} = Nn - n(n+1)/2 = n(n-1)/2 + n(N-n)$. The tangent vector space, $T_V\mathbb{S}_{N,n}$, to $\mathbb{S}_{N,n}$ at point $[V] \in \mathbb{S}_{N,n}$ is defined by

$$T_V\mathbb{S}_{N,n} = \{\ [Z] \in \mathbb{M}_{N,n} : [V]^T[M][Z] + [Z]^T[M][V] = [0_{n,n}]\ \}\,. \tag{24}$$

The dimension of $T_V\mathbb{S}_{N,n}$ is $\nu_{\mathbb{S}} = n(n-1)/2 + n(N-n)$. We are interested in constructing a non classical parameterization of $\mathbb{S}_{N,n}$, which does not require the construction of a big matrix in $\mathbb{M}_{N,N-n}$, and which consists in using the projection $[Z] = \text{Proj}_{T_V\mathbb{S}_{N,n}}([A])$ of any matrix $[A] \in \mathbb{M}_{N,n}$ onto $T_V\mathbb{S}_{N,n}$,

$$[Z] = \text{Proj}_{T_V\mathbb{S}_{N,n}}([A]) := [A] - [V][D] \quad , \quad [D] = ([V]^T[M][A] + [A]^T[M][V])/2\,, \tag{25}$$

in which $[D] \in \mathbb{M}_n^S$ is a $(n \times n)$ symmetric matrix.

Let $[Z] \mapsto R_{s,V}([Z])$ be the smooth mapping from tangent vector space $T_V\mathbb{S}_{N,n}$ of $\mathbb{S}_{N,n}$ at a given point $[V]$ into $\mathbb{S}_{N,n}$,

$$[W] = R_{s,V}([Z])\,, \tag{26}$$

which is constructed in order that

$$[V] = R_{s,V}([0_{N,n}]) \, . \tag{27}$$

Consequently, for any $[Z]$ in $T_V\mathbb{S}_{N,n}$, $[W] = R_{s,V}([Z])$ is such that $[W]^T[M]\,[W] = [I_n]$. The construction of $R_{s,V}$ is performed by using the polar decomposition (see for instance [67]), that is adapted to Eq. (23). For any $[V]$ fixed in $\mathbb{S}_{N,n}$, mapping $R_{s,V}$ is written as

$$[W] = R_{s,V}([Z]) := ([V] + s[Z])\,[H_s(Z)] \quad , \quad [Z] \in T_V\mathbb{S}_{N,n} \, , \tag{28}$$

in which $s \geq 0$ is a real number $s \geq 0$ that is used for controlling the level of fluctuations of $[Z]$ in $T_V\mathbb{S}_{N,n}$ around $[V]$ in $\mathbb{S}_{N,n}$, and where the positive-definite matrix $[H_s(Z)]$ in $\mathbb{M}_n^+$ is defined by

$$[H_s(Z)] = ([I_n] + s^2[Z]^T[M]\,[Z])^{-1/2} \, . \tag{29}$$

It should be noted that another construction could be performed in using the economy-size QR decomposition (see for instance [68]), which is also computationally efficient for $n \ll N$ with N very large.

A parameterization $[W] = \mathcal{R}_{s,V}([A])$, defined on $\mathbb{M}_{N,n}$, of $\mathbb{S}_{N,n}$ at a given point $[V]$ in $\mathbb{S}_{N,n}$ is constructed in substituting the parameterization of $[Z]$ defined by Eq. (25) into $[W] = R_{s,V}([Z])$ that is defined by Eqs. (28) and (29). Consequently, the mapping $[A] \mapsto \mathcal{R}_{s,V}([A])$ from $\mathbb{M}_{N,n}$ into $\mathbb{S}_{N,n}$ is defined, for all $[A]$ in $\mathbb{M}_{N,n}$, by

$$[W] = \mathcal{R}_{s,V}([A]) := R_{s,V}([A] - [V]\,[D]) \quad , \quad [D] = ([V]^T[M]\,[A] + [A]^T[M]\,[V])/2 \, , \tag{30}$$

and consequently, we have

$$[W]^T[M]\,[W] = [I_n] \quad , \quad [V] = \mathcal{R}_{s,V}([0_{N,n}]) \in \mathbb{S}_{N,n} \, . \tag{31}$$

3.2 Construction of a parameterization of the subset $\mathcal{S}_{N,n}$ of the compact Stiefel manifold $\mathbb{S}_{N,n}$ induced by the additional constraint $[B]^T[W] = [0_{N_{\text{CD}},n}]$

We have now to construct a parameterization $[W] = \mathcal{R}_{s,V}([A])$ in presence of an additional constraint equation $[B]^T[W] = [0_{N_{\text{CD}},n}]$ on $[W]$. Let us consider the case for which the ROB $[V]$ belongs to the subset $\mathcal{S}_{N,n}$ of $\mathbb{S}_{N,n}$ defined by

$$\mathcal{S}_{N,n} = \{\ [V] \in \mathbb{M}_{N,n}\ ,\ [V]^T[M]\,[V] = [I_n]\ ,\ [B]^T[V] = [0_{N_{\text{CD}},n}]\ \} \subset \mathbb{S}_{N,n} \, , \tag{32}$$

in which $0 < N_{\text{CD}} < N$ is the number of constraint equations and where $[B]$ is a given matrix such that

$$[B] \in \mathbb{M}_{N,N_{\text{CD}}} \quad , \quad [B]^T\,[B] = [I_{N_{\text{CD}}}] \, . \tag{33}$$

Such a parameterization is given by Eq. (30), in which $[A]$ is any matrix that belongs to the subset $\mathbb{M}_{N,n}^B$ of $\mathbb{M}_{N,n}$ defined by

$$\mathbb{M}_{N,n}^B = \{\ [A] \in \mathbb{M}_{N,n}\ ,\ [A] = ([I_N] - [B]\,[B]^T)\,[U]\ ,\ [U] \in \mathbb{M}_{N,n}\ ,\} \, . \tag{34}$$

It can easily be seen that, for all $[A]$ in $\mathbb{M}_{N,n}^B$, we have $[B]^T[A] = [0_{N_{\text{CD}},n}]$ and consequently, $[B]^T[W] = [0_{N_{\text{CD}},n}]$.

- From a numerical point of view, the parameterization of matrix $[A]$ defined by Eq. (34) can be rewritten as $[A] = [U] - [B]\,\{[B]^T\,[U]\}$ in which $[B]^T\,[U] \in \mathbb{M}_{N_{\text{CD}},n}$. Consequently, the assemblage of the big $(N \times N)$ matrix $[B]\,[B]^T$ is never done.
- If the constraint $[B]^T[W] = [0_{N_{\text{CD}},n}]$ does not exist, then Eq. (34) is replaced by $[A] = [U]$ with $[U] \in \mathbb{M}_{N,n}$.

3.3 Construction of a stochastic reduced-order basis (SROB) associated with the ROB

The construction of the SROB, associated with ROB $[V]$, is performed for $\mathbb{S}_{N,n}$ with the additional constraint equation $[B]^T[W] = [0_{N_{\text{CD}},n}]$, which means that matrix $[A]$ belongs to $\mathbb{M}^B_{N,n}$. If the constraint equation $[B]^T[W] = [0_{N_{\text{CD}},n}]$ is removed, then the stochastic modeling presented hereinafter holds (simply write that $[A] = [U]$). Consequently, the stochastic modeling of matrix $[U]$ will be the same for the two cases (with or without the constraint equation).

For a given ROB that is represented by a matrix $[V]$ given in $\mathcal{S}_{N,n} \subset \mathbb{S}_{N,n}$ (thus $[B]^T[V] = [0_{N_{\text{CD}},n}]$), the associated SROB consists in introducing a random matrix $[\mathbf{W}]$, defined on a probability space $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathcal{S}_{N,n}$, such that

$$[\mathbf{W}]^T[M]\,[\mathbf{W}] = [I_n] \ a.s. \quad , \quad [B]^T[\mathbf{W}] = [0_{N_{\text{CD}},n}] \ a.s. \,, \tag{35}$$

and possibly, verifying additional constraints that will be defined after. We then have to construct the probability measure $P_{[\mathbf{W}]}$ of the random matrix $[\mathbf{W}]$, on $\mathbb{M}_{N,n}$, for which its support is the manifold $\mathcal{S}_{N,n}$,

$$\text{supp}\, P_{[\mathbf{W}]} = \mathcal{S}_{N,n} \subset \mathbb{S}_{N,n} \subset \mathbb{M}_{N,n} \,.$$

The construction of such a probability measure on the manifold $\mathcal{S}_{N,n}$ requires the introduction of an adapted parameterization of $\mathcal{S}_{N,n}$ in order to release the difficulties induced by the support of the measure. Below, we give a construction of the SROB in three steps, and the available information is gradually introduced in the construction.

3.3.1 Step 1 of the construction of the SROB (introducing a parameterization)

The SROB is constructed by using the parameterization of $[W]$ defined by Eqs. (30) and (34). Random matrix $[\mathbf{W}]$ can then be written as

$$[\mathbf{W}] = \mathcal{R}_{s,V}([\mathbf{A}]) = R_{s,V}([\mathbf{A}] - [V]\,[\mathbf{D}]) \quad \text{with} \quad [\mathbf{D}] = ([V]^T[M]\,[\mathbf{A}] + [\mathbf{A}]^T[M]\,[V])/2 \,, \tag{36}$$

in which $[\mathbf{A}]$ is a random matrix defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in subset $\mathbb{M}^B_{N,n}$ of $\mathbb{M}_{N,n}$,

$$[\mathbf{A}] = ([I_N] - [B]\,[B]^T)\,[\mathbf{U}] = [\mathbf{U}] - [B]\,\{[B]^T\,[\mathbf{U}]\} \,, \tag{37}$$

where the random matrix $[\mathbf{U}] = [\mathbf{U}^1 \ldots \mathbf{U}^n]$ is defined on $(\Theta, \mathcal{T}, \mathcal{P})$, with values in $\mathbb{M}_{N,n}$, in which the columns of $[\mathbf{U}]$ are n random vectors $\mathbf{U}^1, \ldots, \mathbf{U}^n$ with values in $\mathbb{R}^N$. The deterministic mapping $R_{s,V}$ (from tangent vector space $T_V\mathbb{S}_{N,n}$ of $\mathbb{S}_{N,n}$ at point $[V]$ into $\mathbb{S}_{N,n}$) is defined by Eqs. (28) and (29). Note that $[\mathbf{D}]$ is a random matrix with values in $\mathbb{M}^S_n$.

3.3.2 Step 2 of the construction of the SROB (defining the available information)

Taking into account Eqs. (36) and (37), the stochastic model of random matrix $[\mathbf{W}]$ and its generator are completely defined by the stochastic model and the generator of random matrix $[\mathbf{U}]$.

(i) The construction of the SROB is performed in order that the statistical fluctuations of random matrix $[\mathbf{W}]$ are around deterministic matrix $[V]$. Taking into account that $[\mathbf{W}] = [V]$ for $[\mathbf{A}] = [0_{N,n}]$, random matrix $[\mathbf{A}]$ must be a centered random variable, which is satisfied if $[\mathbf{U}]$ is also a centered random matrix (due to Eq. (37)). Consequently, we must have,

$$E\{[\mathbf{U}]\} = [0_{N,n}] \,. \tag{38}$$

(ii) In order to minimize the number of hyperparameters in the stochastic model of the centered random matrix $[\mathbf{U}]$, the $Nn(Nn+1)/2$ components of the fourth-order symmetric covariance tensor $\{\mathbb{c}_{jkj'k'}\}_{jkj'k'}$ of $\mathbb{M}_{N,n}$-valued random matrix $[\mathbf{U}]$ are not kept because N can be very large, and consequently, the proposed stochastic model would be unusable. We therefore choose the following reduced parameterization for tensor $\mathbb{c}$, which exhibits only $1+n(n+1)/2$ hyperparameters, such that, for all j and j' in $\{1,\ldots,N\}$, and for all k and k' in $\{1,\ldots,n\}$,

$$\mathbb{c}_{jkj'k'} = E\{\mathbf{U}_{jk}\mathbf{U}_{j'k'}\} = [C_N(\beta)]_{jj'}[c_n]_{kk'}\,, \tag{39}$$

in which

- $[C_N(\beta)] \in \mathbb{M}_N^+$ is a type-covariance matrix depending on a hyperparameter β such that $0 < \beta_d \leq \beta \leq \beta_u < +\infty$, for which its construction is detailed in [65]. This covariance matrix allows for introducing a correlation between the components $U_1^k,\ldots,U_N^k$ of each random vector $\mathbf{U}^k$ such that $U_j^k = [\mathbf{U}]_{jk}$.

- $[c_n] \in \mathbb{M}_n^+$ is a type-covariance matrix and consequently, there is an upper triangular matrix $[\sigma]$ belonging to $\mathbb{M}_n^{\mathrm{u}}$ such that (Cholesky's factorization),

$$[c_n] = [\sigma]^T[\sigma]\,. \tag{40}$$

  Matrix $[c_n]$ allows for describing the correlation between the random vectors $\mathbf{U}^1,\ldots,\mathbf{U}^n$.

Using Eqs. (38), (39), and (40), the second-order $\mathbb{M}_{N,n}$-valued random matrix $[\mathbf{U}]$ can be rewritten as $[\mathbf{U}] = [\mathbf{G}]\,[\sigma]$ in which $[\mathbf{G}]$ is a second-order centered $\mathbb{M}_{N,n}$-valued random matrix defined on probability space $(\Theta,\mathcal{T},\mathcal{P})$, such that, for all j and j' in $\{1,\ldots,N\}$, and for all k and k' in $\{1,\ldots,n\}$,

$$E\{\mathbf{G}_{jk}\} = 0 \quad , \quad E\{\mathbf{G}_{jk}\mathbf{G}_{j'k'}\} = [C_N(\beta)]_{jj'}\,\delta_{kk'}\,. \tag{41}$$

It can be seen that

$$E\{[\mathbf{G}]\,[\mathbf{G}]^T\} = n\,[C_N(\beta)] \quad , \quad E\{[\mathbf{G}]^T[\mathbf{G}]\} = (\mathrm{tr}[C_N(\beta)])\,[I_n]\,.$$

Consequently, for the construction proposed, random matrix $[\mathbf{U}]$ is parameterized as

$$[\mathbf{U}] = [\mathbf{G}]\,[\sigma]\,, \tag{42}$$

and is such that

$$E\{[\mathbf{U}]\,[\mathbf{U}]^T\} = (\mathrm{tr}[c_n])\,[C_N(\beta)] \quad , \quad E\{[\mathbf{U}]^T[\mathbf{U}]\} = (\mathrm{tr}[C_N(\beta)])\,[c_n]\,. \tag{43}$$

3.3.3 Step 3 of the construction of the SROB (constructing the stochastic model of $[\mathbf{W}]$)

(i) *About the stochastic model of random matrix* $[\boldsymbol{G}]$. The stochastic model of the second-order centered random matrix $[\mathbf{G}]$ with values in $\mathbb{M}_{N,n}$ and its generator of independent realizations is detailed in [65] for which $[\mathbf{G}]$ is a non-Gaussian random matrix. This construction is based on the finite element discretization of a non-Gaussian random field whose trajectories are indefinitely continuously differentiable functions almost surely in order to preserve the regularity properties of the solution of the HFM. The stochastic model proposed is chosen in order that it exhibits only a scalar hyperparameter β and does not require the explicit construction of the big matrix $[C_N(\beta)]$ for which the number of entries is $N(N+1)/2$, that is unusable for a large

value of N.

(ii) *Stochastic model and hyperparameters of random matrix* $[\boldsymbol{W}]$. The stochastic representation of the second-order non-Gaussian and not centered random matrix $[\mathbf{W}]$ with values in the manifold $\mathcal{S}_{N,n} \subset \mathbb{S}_{N,n} \subset \mathbb{M}_{N,n}$ is defined by Eqs. (36), (37), and (42), which are rewritten (using Eqs. (28) and (29)) as

$$[\mathbf{W}] = R_{s,V}([\mathbf{Z}]) = ([V] + s\,[\mathbf{Z}])\,[H_s(\mathbf{Z})]\,, \tag{44}$$

$$[H_s(\mathbf{Z})] = ([I_n] + s^2\,[\mathbf{Z}]^T[M]\,[\mathbf{Z}])^{-1/2}\,, \tag{45}$$

$$[\mathbf{Z}] = [\mathbf{A}] - [V]\,[\mathbf{D}]\,, \tag{46}$$

$$[\mathbf{D}] = ([V]^T[M]\,[\mathbf{A}] + [\mathbf{A}]^T[M]\,[V])/2\,, \tag{47}$$

$$[\mathbf{A}] = [\mathbf{U}] - [B]\,\{[B]^T\,[\mathbf{U}]\}\,, \tag{48}$$

$$[\mathbf{U}] = [\mathbf{G}(\beta)]\,[\sigma]\,. \tag{49}$$

in which

- $[\mathbf{G}(\beta)]$ is the second-order non-Gaussian centered random matrix with values in $\mathbb{M}_{N,n}$ detailed in [65], and for which the covariance tensor is defined by Eq. (41).
- $[\sigma]$ is a given upper triangular matrix in $\mathbb{M}_n^{\mathrm{u}}$ (positive diagonal entries).
- $[B]$ is a given matrix in $\mathbb{M}_{N,N_{\mathrm{cb}}}$ such that $[B]^T\,[B] = [I_{N_{\mathrm{cb}}}]$.
- $[V]$ is a given matrix in $\mathcal{S}_{N,n}$.

For $[V]$ fixed in $\mathcal{S}_{N,n}$, the $2 + n(n+1)/2$ *hyperparameters* of the stochastic model of random matrix $[\mathbf{W}]$ with values in $\mathcal{S}_{N,n}$ are:

- the deterministic real parameter s is such that $\varepsilon_0 \leq s \leq 1$ in which ε_0 is given and such that $0 \leq \varepsilon_0 < 1$ (if $s = 0$, then $[\mathbf{W}] = [V]$ is deterministic and there are no statistical fluctuations),
- the deterministic real parameter β such that $0 < \beta_d \leq \beta \leq \beta_u < +\infty$ in which β_d and β_u are given,
- the upper triangular matrix $[\sigma]$ in $\mathbb{M}_n^{\mathrm{u}}$ (positive diagonal entries), which is parameterized by $n(n+1)/2$ parameters, and such that the diagonal entries satisfy the constraints $\varepsilon_0 \leq [\sigma]_{11}, \ldots, [\sigma]_{nn} \leq \sigma_u < +\infty$, in which σ_u is given.

The hyperparameter is thus $\boldsymbol{\alpha} = (s, \beta, \{[\sigma]_{kk'}, 1 \leq k \leq k' \leq n\})$ with length $m_{\boldsymbol{\alpha}} = 2 + n(n+1)/2$, which belongs to the admissible set $\mathcal{C}_{\boldsymbol{\alpha}}$ defined by

$$\mathcal{C}_{\boldsymbol{\alpha}} = \{s \in [\varepsilon_0\,, 1]\,,\ \beta \in [\beta_d, \beta_u]\,,\ \varepsilon_0 \leq [\sigma]_{11}, \ldots, [\sigma]_{nn} \leq \sigma_u\,,\ [\sigma]_{kk'} \in \mathbb{R}\,, k < k'\}\,. \tag{50}$$

4 APPLICATION IN NONLINEAR COMPUTATIONAL STRUCTURAL DYNAMICS

In this section, a verificaton problem for the theory exposed above is presented. The chosen HFM corresponds to a finite element model of a three-dimensional slender damped elastic bounded medium with nonlinear barriers that induce nonlinearities in the dynamical system. Additional applications for a more deeper validation of the proposed theory can be found in

[65]. The problem is intentionally chosen to be sufficiently simple so that it can be easily reproduced by the interested reader. Nevertheless, it is representative of difficult problems as far as ROM model form uncertainties are concerned. In the frequency domain, the excitation has an energy located in a narrow frequency band within in a broad frequency band of analysis. The differences between the HFM and the ROM are very small in the frequency band of excitation (small uncertainties), but are large outside this frequency band (large uncertainties). These large uncertainties are due to the transfer of the energy outside the frequency band of excitation due to the nonlinearities in the dynamical system. Consequently, they correspond to second-order contributions. This application shows how such second-order contributions, which are located outside the frequency band of excitation, can be predicted using the nonparametric probabilistic approach of model uncertainties.

4.1 Description of the mechanical system

The mechanical system is made up of a 3D linear elastic structure with two elastic barriers that induce impact non-linearities. It is defined in a cartesian coordinate system $Ox_1x_2x_3$ (see Fig. 1). The cylinder has length $L_1 = 1.20\,m$ and a rectangular section with height $L_2 = 0.12\,m$ and width $L_3 = 0.24\,m$. The two end sections are located at $x_1 = 0$ and $x_1 = L_1$. The elastic medium is made of a homogeneous and isotropic elastic material for which the Young modulus is $10^{10}\,N/m^2$, the Poisson coefficient is 0.15 and the mass density is $1500\,Kg/m^3$. A damping term is added and is described by a global damping rate of $\xi_d = 0.01$ for each elastic mode of the structure without the elastic barriers, and will be introduced at the ROM level. The nonlinear

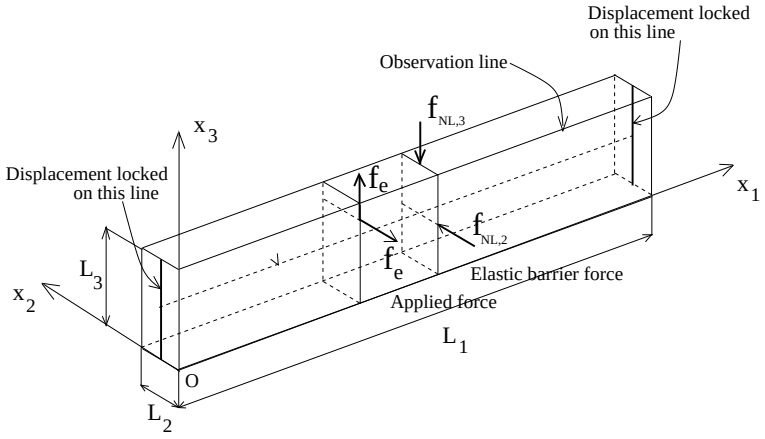


Figure 1: Scheme of the mechanical system: slender elastic structure with elastic barriers.

forces are due to elastic barriers that induce two nonlinear point forces on the structure, one in direction x_2 applied to the point $(x_1 = 0.66, x_2 = 0, x_3 = L_3/2)$, and another one in direction x_3 applied to the point $(x_1 = 0.66, x_2 = L_2/2, x_3 = L_3)$ (see Fig. 1) of intensity $-f_{\text{NL},2}$ and $-f_{\text{NL},3}$ that are assumed to be independent of the velocity and such that

$$f_{\text{NL},2}(\eta) = k_{b,2}\,(\eta + \varepsilon_{b,2})\,\mathbb{1}_{\mathbb{R}^+}(-\eta - \varepsilon_{b,2}) \quad , \quad \eta \in \mathbb{R}\,, \tag{51}$$

$$f_{\text{NL},3}(\eta) = k_{b,3}\,(\zeta - \varepsilon_{b,3})\,\mathbb{1}_{\mathbb{R}^+}(\zeta - \varepsilon_{b,3}) \quad , \quad \zeta \in \mathbb{R}\,, \tag{52}$$

where $k_{b,2} = k_{b,3} = 2 \times 10^9\,N/m$ is the elasticity constant of the barriers and $\varepsilon_{b,2} = \varepsilon_{b,3} = 2 \times 10^{-4}\,m$ are the two gaps (positive values). A time-dependent point force is applied at the

boundary of the slender structure at the point $(x_1 = 0.46, x_2 = 0, x_3 = 0.2)$ (see Fig. 1) in the directions x_2 and x_3. The x_2- and x_3-components of this point force are equal to a square integrable real-valued function f_e defined by

$$f_e(t) = \frac{f_0}{\pi\, t} \left\{ \sin(t(\omega_c + \Delta\omega_c/2)) - \sin(t(\omega_c - \Delta\omega_c/2) \right\} \quad , \quad t \in [t_0, T]\,, \tag{53}$$

where $f_0 = 100\ N$ is the amplitude, $\omega_c = 2\,\pi \times 470\ rad/s$ is the central frequency band, and $\Delta\omega_c = 2\,\pi \times 300\ rad/s$ is the bandwidth. The signal energy of this excitation is mainly concentrated in the frequency band $[320\,, 620]\ Hz$. An observation line of equation $\{0 \leq x_1 \leq L_1\,;\, x_2 = 0\,;\, x_3 = L_3\}$ for $t \in [t_0, T]$ is introduced. In order to limit the number of figures, o nly the x_2- and x_3-displacements of 2 observation points belonging to the observation line (see Fig. 1) are considered and are denoted by Obs_{34} and Obs_{51} for which the x_1-coordinates are 0.66 and 1.00. At initial time $t_0 = -0.0403\ s$, the system is at rest. The final time is $T = 0.3790\ s$. The *frequency band of observation* is chosen as $\mathcal{B}_o = [0\,, \omega_o]$ with $\omega_o = 2\,\pi \times 1550\ rad/s$. The time sampling is defined by the frequency $\omega_{\text{max}} = 2\,\pi \times 12,400\ rad/s$. There are $10,400$ time steps and the sampling time step is $\Delta t = \pi/\omega_{\text{max}} = 4.032 \times 10^{-5}\ s$. There are $10,400$ frequency steps in the frequency band $[-\omega_{\text{max}}, \omega_{\text{max}}]$ and the sampling frequency step is $\Delta\omega = 2\,\pi \times 2.38\ rad/s$.

4.2 HFM, numerical solver, and results

(i) *High fidelity computational model*. A 3D computational model is constructed with a finite element mesh made up of $60 \times 6 \times 12 = 4\,320$ three-dimensional 8-nodes solid elements. There are $5\,551$ nodes and $N = 16,653$ dofs. The number of zero Dirichlet conditions is $N_{\text{CD}} = 78$ (the displacements are zero for 2×13 nodes). For this HFM, the equations Eqs. (1) to (4) are rewritten as follows:

$$[M]\,\ddot{\mathbf{y}}(t) + \mathbf{g}(\mathbf{y}(t), \dot{\mathbf{y}}(t)) = \mathbf{f}(t) \quad , \quad t \in]t_0, T]\,, \tag{54}$$

$$\mathbf{g}(\mathbf{y}(t), \dot{\mathbf{y}}(t)) = [\mathbb{D}]\,\dot{\mathbf{y}}(t) + [K]\,\mathbf{y}(t) + \mathbf{f}_{\text{NL}}(\mathbf{y}(t))\,, \tag{55}$$

with the zero initial conditions at time t_0,

$$\mathbf{y}(t_0) = \mathbf{0}_N \quad , \quad \dot{\mathbf{y}}(t_0) = \mathbf{0}_N\,, \tag{56}$$

and the $N_{\text{CD}} < N$ constraint equations written in matrix form as

$$[B]^T \mathbf{y}(t) = \mathbf{0}_{N_{\text{CD}}} \quad , \quad t \in [t_0, T]\,, \tag{57}$$

where $[B]$ is a given matrix in $\mathbb{M}_{N,N_{\text{CD}}}$ such that $[B]^T[B] = [I_{N_{\text{CD}}}]$, which is constructed by using the N_{CD} zero Dirichlet conditions defined in Section 4.1. A construction of $[\mathbb{D}]$ is detailed in [65] for which the numerical complexity of the product of $[\mathbb{D}]$ by a vector in $\mathbb{R}^N$ is of the order of the one corresponding to the product of $[M]$ by a vector in $\mathbb{R}^N$, which is much less than N^2 because $[M]$ is a sparse finite element matrix. In the frequency band $\mathcal{B}_o = [0\,, \omega_o]$, the QoI is the vector $\widehat{\mathbf{o}}(\omega) = (\widehat{o}_1(\omega), \ldots, \widehat{o}_{m_{\mathbf{o}}}(\omega)) \in \mathbb{C}^{m_{\mathbf{o}}}$ defined as follows:

- For plotting and analyzing the responses of the SROM, $m_{\mathbf{o}} = 2 \times 2 = 4$ dofs in direction x_2 and x_3 of the 2 observation nodes Obs_{34} and Obs_{51}.

- For the identification of the hyperparameter $\boldsymbol{\alpha}$ of the SROB, $m_{\mathbf{o}} = 2 \times 61 = 122$ dofs in direction x_2 and x_3 of all the 61 nodes belonging to the observation line.

For all ω in $\mathcal{B}_o$, the complex vector vector $\widehat{\mathbf{o}}(\omega)$ is written as

$$\widehat{\mathbf{o}}(\omega) = \mathbf{h}(-\omega^2\, \widehat{\mathbf{y}}(\omega)) \quad , \quad \omega \in \mathcal{B}_o \, , \tag{58}$$

in which $\mathbf{h}$ is a linear mapping from $\mathbb{C}^N$ into $\mathbb{C}^{m_{\mathbf{o}}}$, which extracts the dofs from vector $-\omega^2\, \widehat{\mathbf{y}}(\omega) \in \mathbb{C}^N$, and where

$$\widehat{\mathbf{y}}(\omega) = \int_{t_0}^{T} e^{-i\omega\, t} \mathbf{y}(t)\, dt \quad , \quad \omega \in \mathcal{B}_o \, . \tag{59}$$

(ii) *Numerical solver*. The implicit Newmark time-integration scheme [69] is used (with parameters $\delta = 0.5$ and $\alpha = 0.25(0.5 + \delta)^2 = 0.25$) and with a fixed point method (without relaxation) at each sampling time with a relative precision of 10^{-6}. In order to guaranty the convergence of the fixed point method, a local adaptive time step is implemented (time step Δt is locally decreased). Such a numerical method has been preferred to the other possible ones due to the presence of contacts induced by the barriers, which require an adaptive time steps.

(iii) *Results and quantification of the effects of the nonlinearities*. In Fig. 2, the four figures compare the graphs $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM, with the graphs computed with the linear HFM (that is to say, in removing the nonlinear elastic barriers in the HFM), for the x_2- and x_3-accelerations of Obs_{51} and Obs_{34}. These figures show the effects of the nonlinear elastic barriers on the responses, in particular, it can be seen an important transfer of the energy in the frequency band that is outside the main frequency band $[320\,, 620]\ Hz$ of the excitation.

4.3 ROM, numerical solver, and results

(i) *Reduced-order computational model*. Let $\{\boldsymbol{\varphi}^1, \ldots, \boldsymbol{\varphi}^n\}$ be the first n elastic modes associated with the first n eigenfrequencies $0 < \omega_1 < \ldots < \omega_n$ of the linear undamped structure associated with the nonlinear damped dynamical system, which are such that

$$[K]\, \boldsymbol{\varphi}^k = \lambda_k\, [M]\, \boldsymbol{\varphi}^k \, , \tag{60}$$

with the constraint equation

$$[B]^T \boldsymbol{\varphi}^k = \mathbf{0}_{N_{\text{CD}}} \quad , \quad k = 1, \ldots, n \, , \tag{61}$$

and where $\lambda_1 = \omega_1^2, \ldots, \lambda_n = \omega_n^2$. The elastic modes satisfy the usual orthogonality properties,

$$< [M]\, \boldsymbol{\varphi}^k \, , \boldsymbol{\varphi}^{k'} >= \delta_{kk'} \quad , \quad < [K]\, \boldsymbol{\varphi}^k \, , \boldsymbol{\varphi}^{k'} >= \lambda_k\, \delta_{kk'} \, . \tag{62}$$

For the nonlinear dynamical system, the ROB is chosen as $[V] = [\boldsymbol{\varphi}^1 \ldots \boldsymbol{\varphi}^n] \in \mathbb{M}_{N,n}$, which is such that

$$[V]^T\, [M]\, [V] = [I_n] \quad , \quad [B]^T\, [V] = [0_{N_{\text{CD}},n}] \, . \tag{63}$$

Using the ROB $[V] \in \mathbb{M}_{N,n}$ and Eq. (63), the ROM associated with the HFM defined by Eqs. (54) to (57) is written as

$$\mathbf{y}^{(n)}(t) = [V]\, \mathbf{q}(t) \quad , \quad t \in [t_0, T] \, , \tag{64}$$

$$\ddot{\mathbf{q}}(t) + [V]^T[\mathbb{D}][V]\, \dot{\mathbf{q}}(t) + [V]^T[K][V]\mathbf{q}(t) + [V]^T \mathbf{f}_{\text{NL}}([V]\mathbf{q}(t)) = [V]^T \mathbf{f}(t) \ , \ t \in]t_0, T] \, , \tag{65}$$

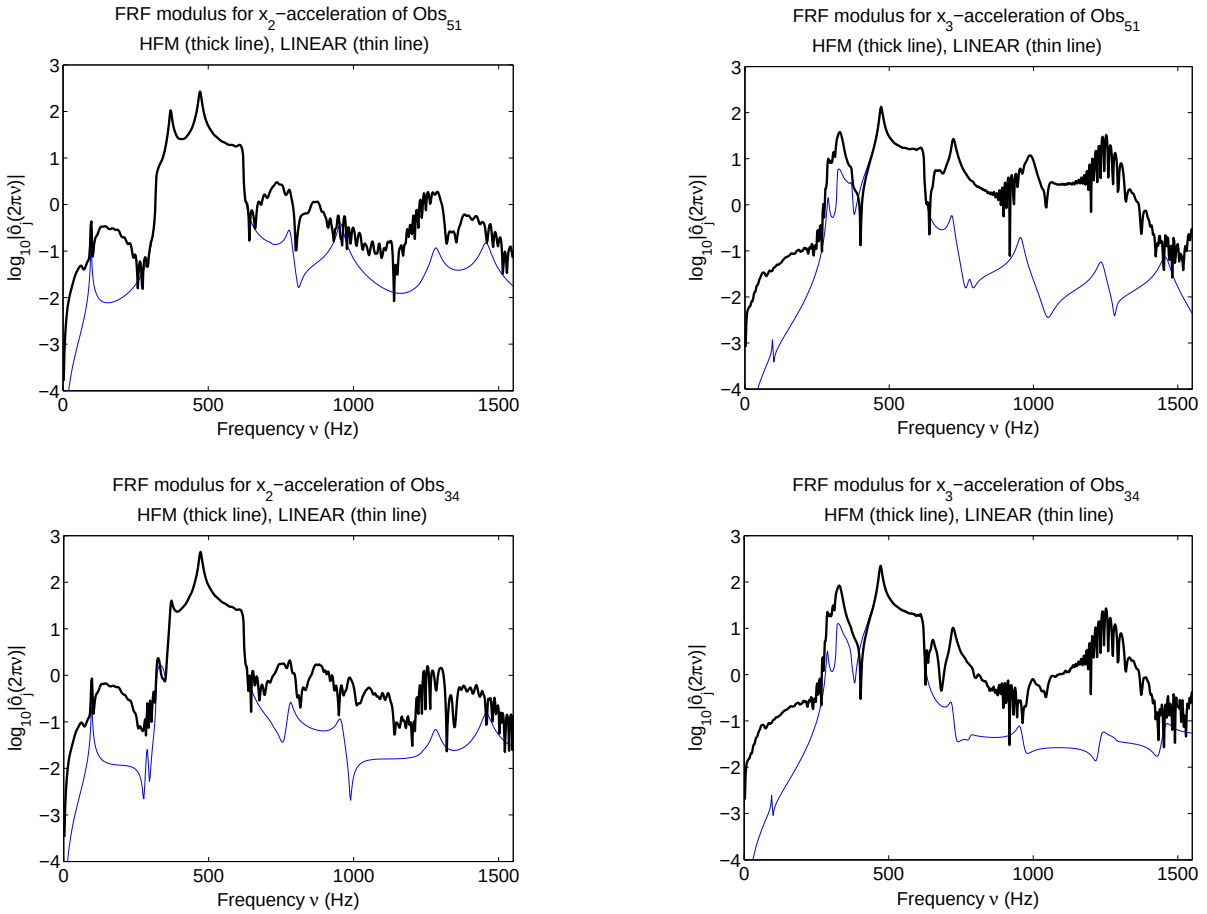


Figure 2: Graphs $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM (thick lines) and with the LINEAR HFM (thin lines) for x_2- and x_3-accelerations of Obs_{51} (up left and up right figures), and for x_2- and x_3-accelerations of Obs_{34} (down left and down right figures).

with the initial conditions

$$\mathbf{q}(t_0) = \mathbf{0}_n \quad , \quad \dot{\mathbf{q}}(t_0) = \mathbf{0}_n \, . \tag{66}$$

The approximation $\widehat{\mathbf{o}}^{(n)}(\omega)$ of the QoI $\widehat{\mathbf{o}}(\omega)$ defined by Eqs. (58) and (59) is written as

$$\widehat{\mathbf{o}}^{(n)}(\omega) = \mathbf{h}(-\omega^2 \, \widehat{\mathbf{y}}^{(n)}(\omega)) \quad , \quad \omega \in \mathcal{B}_o \, , \tag{67}$$

$$\widehat{\mathbf{y}}^{(n)}(\omega) = \int_{t_0}^{T} e^{-i\omega t} \mathbf{y}^{(n)}(t) \, dt \quad , \quad \omega \in \mathcal{B}_o \, . \tag{68}$$

(ii) *Numerical solver*. The algorithm described in Section 4.2-(ii) is used but is adapted to Eq. (65) that is rewritten as $\ddot{\mathbf{q}}(t) + 2\,\xi_d\,[\lambda^{(n)}]^{1/2}\,\dot{\mathbf{q}}(t) + [\lambda^{(n)}]\,\mathbf{q}(t) = [V]^T\mathbf{f}(t) - [V]^T\mathbf{f}_{\text{NL}}([V]\,\mathbf{q}(t))$ in which $[\lambda^{(n)}]$ is the positive-definite diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_n$.

(iii) *Choice of the reduced-order dimension n, results, and quantification of the errors induced by the use of the ROM instead of the HFM*. Taking into account that a significative difference between the responses computed with ROM and with the HFM must be generated (in order to validate the capability of the nonparametric stochastic method proposed to take into account this type of modeling errors), a good compromise between the numerical cost and

the quality of the ROM leads us to choose $n = 20$ as the dimension of the ROM. We then have $\omega_1 = 2\,\pi \times 96.69\,rad/s$, $\omega_4 = 2\,\pi \times 472\,rad/s$, $\omega_5 = 2\,\pi \times 720\,rad/s$, $\omega_{11} = 2\,\pi \times 1474\,rad/s$, $\omega_{12} = 2\,\pi \times 1754\,rad/s$, and $\omega_{20} = 2\,\pi \times 2936\,rad/s$. Consequently, there are 4 elastic modes in the frequency band $[0\,,620]\;Hz$ containing the main part of the excitation, 11 elastic modes in the frequency band of observation $[0\,,1550]\;Hz$ and 9 elastic modes in the frequency band $[1550\,,3100]\;Hz$. In Fig. 3, the four figures show a comparison of the graphs $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM, with the graphs $\nu \mapsto \log_{10}(|\widehat{o}_j^{(n)}(2\pi\nu)|)$ computed with the ROM, for the x_2- and x_3-accelerations of Obs_{51} and Obs_{34}. These figures show that the differences between the HFM and the ROM are very small in the frequency band $[320\,,620]\;Hz$ of the excitation, but are significant outside this frequency band. Such differences could be reduced in increasing dimension n of the ROM, but as we have explained above, the reduced-order dimension n is chosen in order that significant differences exist between the ROM and the HFM.

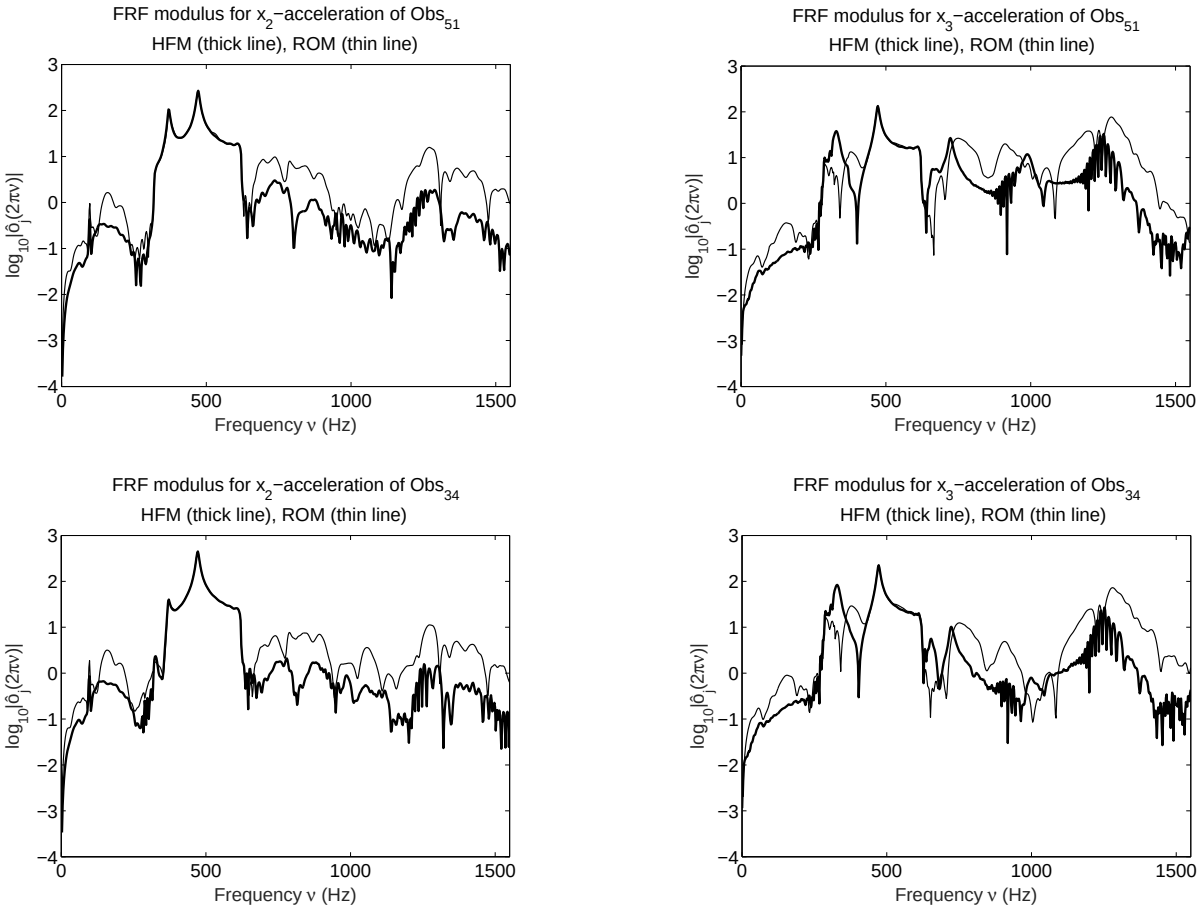


Figure 3: Graphs $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM (thick lines) and graphs $\nu \mapsto \log_{10}(|\widehat{o}_j^{(n)}(2\pi\nu)|)$ computed with the ROM (thin lines) for x_2- and x_3-accelerations of Obs_{51} (up left and up right figures), and for x_2- and x_3-accelerations of Obs_{34} (down left and down right figures).

4.4 SROM, stochastic solver, and results

(i) *Stochastic reduced-order computational model.* From Eqs. (11) to (14), the SROM associated with Eqs. (64) to (68), is obtained in replacing $[V]$ by the random matrix $[\mathbf{W}]$. Conse-

quently, $\mathbf{y}^{(n)}(t)$, $\mathbf{q}(t)$, and $\widehat{\mathbf{o}}^{(n)}(\omega)$, become the random vectors $\mathbf{Y}^{(n)}(t)$, $\mathbf{Q}(t)$, and $\widehat{\mathbf{O}}^{n}(\omega)$, and the SROM is written as

$$\mathbf{Y}^{(n)}(t) = [\mathbf{W}]\,\mathbf{Q}(t) \quad , \quad t \in [t_0, T]\,, \tag{69}$$

$$\ddot{\mathbf{Q}}(t)+[\mathbf{W}]^T[\mathbb{D}][\mathbf{W}]\,\dot{\mathbf{Q}}(t)+[\mathbf{W}]^T[K][\mathbf{W}]\mathbf{Q}(t)+[\mathbf{W}]^T\mathbf{f}_{\text{NL}}([\mathbf{W}]\mathbf{Q}(t))=[\mathbf{W}]^T\mathbf{f}(t)\,,\ t \in]t_0, T]\,, \tag{70}$$

with the initial conditions

$$\mathbf{Q}(t_0) = \mathbf{0}_n \quad , \quad \dot{\mathbf{Q}}(t_0) = \mathbf{0}_n \quad , a.s \quad . \tag{71}$$

Let $\widehat{\mathbf{O}}^{(n)}(\omega;\boldsymbol{\alpha})$ be the random QoI with values in $\mathbb{C}^{m_o}$ with $m_o = 122$, depending on hyperparameter $\boldsymbol{\alpha}$, such that

$$\widehat{\mathbf{O}}^{(n)}(\omega;\boldsymbol{\alpha}) = \mathbf{h}(-\omega^2\,\widehat{\mathbf{Y}}^{(n)}(\omega;\boldsymbol{\alpha})) \quad , \quad \omega \in \mathcal{B}_o\,, \tag{72}$$

$$\widehat{\mathbf{Y}}^{(n)}(\omega;\boldsymbol{\alpha}) = \int_{t_0}^{T} e^{-i\omega\,t}\mathbf{Y}^{(n)}(t;\boldsymbol{\alpha})\,dt \quad , \quad \omega \in \mathcal{B}_o\,. \tag{73}$$

(ii) *Identification of the hyperparameter*. The problem is the identification of hyperparameter $\boldsymbol{\alpha} = (s, \beta, \sigma)$ that is defined in Section 3.3.3-(ii), with length $m_{\boldsymbol{\alpha}} = 2 + n(n+1)/2 = 212$, and which belongs to the admissible set $\mathcal{C}_{\boldsymbol{\alpha}} = \mathbb{R}^+ \times [0.01\,,0.1] \times \mathbb{M}_n^{\text{u}}$. For defining the cost function $J(\boldsymbol{\alpha})$, we introduce the random function $\omega \mapsto \mathbf{dB}^{(n)}(\omega;\boldsymbol{\alpha}) = (\text{dB}_1^{(n)}(\omega;\boldsymbol{\alpha}), \ldots, \text{dB}_{m_o}^{(n)}(\omega;\boldsymbol{\alpha}))$ defined on $\mathcal{B}_o$ with valued in $\mathbb{R}^{m_o}$ such that, for all $j = 1, \ldots, m_o$ (with $m_o = 122$),

$$\text{dB}_j^{(n)}(\omega;\boldsymbol{\alpha}) = \log_{10}(|\widehat{O}_j^{(n)}(\omega;\boldsymbol{\alpha})|)\,, \tag{74}$$

in which $\widehat{\mathbf{O}}^{(n)}(\omega;\boldsymbol{\alpha}) = (\widehat{O}_1^{(n)}(\omega;\boldsymbol{\alpha}), \ldots, \widehat{O}_{m_o}^{(n)}(\omega;\boldsymbol{\alpha}))$ is defined by Eq. (72). In order to define the target functions for constructing the cost function, we introduce the functions $\omega \mapsto \mathbf{db}^{\text{ref}}(\omega) = (\text{db}_1^{\text{ref}}(\omega), \ldots, \text{db}_{m_o}^{\text{ref}}(\omega))$ and $\omega \mapsto \mathbf{db}^{(n)}(\omega) = (\text{db}_1^{(n)}(\omega), \ldots, \text{db}_{m_o}^{(n)}(\omega))$, defined on $\mathcal{B}_o$ with values in $\mathbb{R}^{m_o}$ such that, for all $j = 1, \ldots, m_o$,

$$\text{db}_j^{\text{ref}}(\omega) = \log_{10}(|\widehat{o}_j(\omega)|) \quad , \quad \text{db}_j^{(n)}(\omega) = \log_{10}(|\widehat{o}_j^{(n)}(\omega)|)\,, \tag{75}$$

in which $\widehat{\mathbf{o}}(\omega) = (\widehat{o}_1(\omega), \ldots, \widehat{o}_{m_o}(\omega))$ is defined by Eq. (58) and where $\widehat{\mathbf{o}}^{(n)}(\omega) = (\widehat{o}_1^{(n)}(\omega), \ldots, \widehat{o}_{m_o}^{(n)}(\omega))$ is defined by Eq. (67). The cost function is defined by Eq. (15) with

$$J(\boldsymbol{\alpha}) = w_J\,J_{\text{mean}}\,(\boldsymbol{\alpha}) + (1 - w_J)\,J_{\text{std}}\,(\boldsymbol{\alpha})\,, \tag{76}$$

in which

$$J_{\text{mean}}\,(\boldsymbol{\alpha}) = \frac{1}{c_{\text{mean}}} \sum_{j=1}^{m_o} \int_{\mathcal{B}_o} |\text{db}_j^{\text{ref}}(\omega) - E\{\text{dB}_j^{(n)}(\omega;\boldsymbol{\alpha})\}|^2\,w_j(\omega)\,d\omega\,, \tag{77}$$

$$J_{\text{std}}\,(\boldsymbol{\alpha}) = \frac{1}{c_{\text{std}}} \sum_{j=1}^{m_o} \int_{\mathcal{B}_o} |v_j^{(\text{ref},n)}(\omega) - v_j^{(n)}(\omega;\boldsymbol{\alpha})|^2\,w_j(\omega)\,d\omega\,, \tag{78}$$

in which the positive constants c_{mean} and c_{std} are defined by

$$c_{\text{mean}} = \sum_{j=1}^{m_o} \int_{\mathcal{B}_o} |\text{db}_j^{\text{ref}}(\omega)\}|^2\,w_j(\omega)\,d\omega \quad , \quad c_{\text{std}} = \sum_{j=1}^{m_o} \int_{\mathcal{B}_o} |v_j^{(\text{ref},n)}(\omega)|^2\,w_j(\omega)\,d\omega\,. \tag{79}$$

In these equations, $\mathbf{v}^{(\mathrm{ref},n)}(\omega) = (v_1^{(\mathrm{ref},n)}(\omega), \ldots, v_{m_o}^{(\mathrm{ref},n)}(\omega))$ and $\mathbf{v}^{(n)}(\omega;\boldsymbol{\alpha}) = (v_1^{(n)}(\omega;\boldsymbol{\alpha}), \ldots, v_{m_o}^{(n)}(\omega;\boldsymbol{\alpha}))$ are defined, for $j = 1, \ldots m_o$, by

$$v_j^{(\mathrm{ref},n)}(\omega) = \gamma \, |\mathrm{db}_j^{\mathrm{ref}}(\omega) - \mathrm{db}_j^{(n)}(\omega)| \, , \tag{80}$$

in which $\gamma > 0$ is an amplitude factor and where

$$v_j^{(n)}(\omega;\boldsymbol{\alpha}) = \{ \, E\{dB_j^{(n)}(\omega;\boldsymbol{\alpha})^2\} - (E\{dB_j^{(n)}(\omega;\boldsymbol{\alpha})\})^2\} \, \}^{1/2} \, . \tag{81}$$

In Eqs. (77) to (79), for $j = 1, \ldots m_o$, the functions $\omega \mapsto w_j(\omega)$ are bounded on $\mathcal{B}_o$ with values in $\mathbb{R}^+$ and are chosen such that $w_j(\omega) = |\mathrm{db}_j^{\mathrm{ref}}(\omega) - \mathrm{db}_j^{(n)}(\omega)|$ for all ω in $\mathcal{B}_o$. The weight (in the cost function) is $w_J = 0.9$, and the amplitude factor (for the target) is $\gamma = 0.3$. The optimization problem

$$\boldsymbol{\alpha}^{\mathrm{opt}} = (s^{\mathrm{opt}}, \beta^{\mathrm{opt}}, [\sigma^{\mathrm{opt}}]) = \min_{\boldsymbol{\alpha} \in \mathcal{C}_{\boldsymbol{\alpha}}} J(\boldsymbol{\alpha}) \, , \tag{82}$$

is solved using the algorithm detailed in [65], which is based on the use of the interior-point algorithm with constraints. The Monte Carlo solver is used with $1\,000$ independent realizations (mean-square convergence reached). The optimal value of s and β are $s^{\mathrm{opt}} = 0.0103$ and $\beta^{\mathrm{opt}} = 0.0181$ and the optimal value $[\sigma^{\mathrm{opt}}]$ of $[\sigma]$ can be found in [65].

(iii) *Results*. In Fig. 4, the four figures that are displayed summarize the results obtained with the SROM. These four figures are related to the x_2- and x_3-accelerations of Obs_{51} and Obs_{34}. Each figure displays the graph $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM (the target for the mean), the graph $\nu \mapsto \log_{10}(|\widehat{o}_j^{(n)}(2\pi\nu)|)$ computed with the linear ROM, and the confidence region (with a probability $p_c = 0.98$) of the frequency sampled stochastic process $\nu \mapsto \log_{10}(|\widehat{O}_j^{(n)}(2\pi\nu)|)$ constructed with the SROM. The upper envelope of the confidence region corresponds to the quantile for the probability p_c and the lower envelope to the quantile for the probability $1 - p_c$. It can be seen that the results obtained are very good, except for the x_3-accelerations in the small part $[270 \, , 438] \, Hz$ of the frequency band $[0 \, , 1550] \, Hz$ of analysis. Such a relatively bad prediction could certainly be improved using a much more sophisticated optimization algorithm for the identification of the hyperparameters. Nevertheless, it can be seen that the SROM allows for generating a confidence region, which is not centered around the responses computed with the ROM, but which is approximatively well centered around the responses computed with the HFM, which is a relatively difficult problem for taking into account contributions of second-order. Such a result demonstrates the capability of method proposed.

5 CONCLUSIONS

In this paper, a novel nonparametric probabilistic approach has been presented for taking into account modeling errors in any nonlinear high-fidelity model (HFM) for which a nonlinear ROM can be constructed. The nonparametric probabilistic model is implemented in the nonlinear ROM. The proposed stochastic model exhibits a small number of hyperparameters, which allows their identification by solving a statistical inverse problem. The cost function is formulated with respect to a given target related to given observations, which allows for specifying the level of uncertainties induced by the use of the ROM instead of the HFM and/or by the modeling errors (model form uncertainties) introduced in the HFM with respect to experimental data. The least-square method that is used in the paper can be replaced by another statistical inverse method such as the maximum likelihood method. A first validation of the proposed method

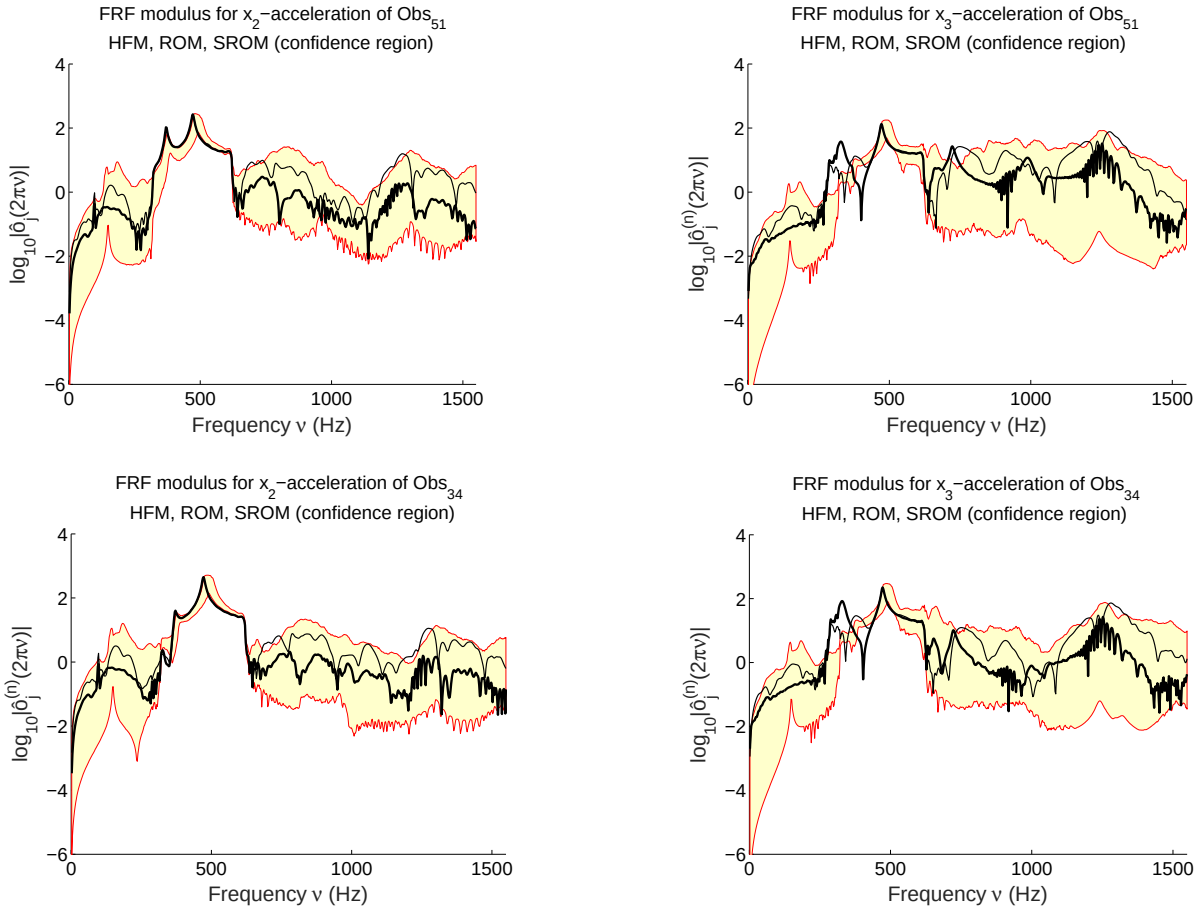


Figure 4: Graphs $\nu \mapsto \log_{10}(|\widehat{o}_j(2\pi\nu)|)$ computed with the HFM (thick lines), graphs $\nu \mapsto \log_{10}(|\widehat{o}_j^{(n)}(2\pi\nu)|)$ computed with the ROM (thin lines), and confidence region for $p_c = 0.98$ (yellow region with red upper and red lower envelopes) computed for $\nu \mapsto \log_{10}(|\widehat{O}_j^{(n)}(2\pi\nu)|)$ computed with the SROM, for x_2- and x_3-accelerations of Obs_{51} (up left and up right figures), and for x_2- and x_3-accelerations of Obs_{34} (down left and down right figures).

has been obtained by analyzing a nonlinear dynamical problem. The proposed nonparametric probabilistic approach of modeling errors can simultaneously be used with the parametric probabilistic approach of the uncertain parameters of the computational model, because, if the construction of the ROB can depend on these random parameters, the construction of the SROB does not directly depend on them (but depends indirectly through the ROB). In addition, if the Monte Carlo method is used as stochastic solver, the realizations of the SROB and those of the random parameters are constructed simultaneously. Consequently, the speed of convergence (that does not depends on the dimension) is not altered if the parametric uncertainties are taken into account in the computational model. The method can be extended to the case of a HFM for which the level of model uncertainties is not the same in the different parts of the system.

ACKNOWLEDGEMENTS

The second author acknowledges partial support by the Army Research Laboratory through the Army High Performance Computing Research Center under Cooperative Agreement W911NF-07-2-0027, partial support by The Boeing Company under Contract Sponsor Ref. 45047, and partial support by DARPA under the Enabling Quantification of Uncertainty in Physical Sys-

tems (EQUiPS) program. This document does not necessarily reflect the position of these institutions, and no official endorsement should be inferred.

REFERENCES

[1] M.A. Grepl, Y. Maday, N.C. Nguyen, A. Patera, Efficient reduced-basis treatment of non-affine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, **41**(03), 575-605, 2007.

[2] N. Nguyen, J. Peraire, An efficient reduced-order modeling approach for non-linear parametrized partial differential equations. *International Journal for Numerical Methods in Engineering*, **76**(1), 27-55, 2008.

[3] S. Chaturantabut, D.C. Sorensen, Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific and Statistical Computing*, **32**(5), 2737-2764, 2010.

[4] J. Degroote, J. Virendeels, K. Willcox, Interpolation among reduced-order matrices to obtain parameterized models for design, optimization and probabilistic analysis. *International Journal for Numerical Methods in Fluids*, **63**, 207-230, 2010.

[5] K. Carlberg, C. Bou-Mosleh, C. Farhat, Efficient non-linear model reduction via a least-squares Petrov-Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, **86**(2), 155-181, 2011.

[6] K. Carlberg, C. Farhat, A low-cost, goal-oriented compact proper orthogonal decomposition basis for model reduction of static systems. *International Journal for Numerical Methods in Engineering*, **86**(3), 381-402, 2011.

[7] D. Amsallem, M.J. Zahr, C. Farhat, Nonlinear model order reduction based on local reduced-order bases. *International Journal for Numerical Methods in Engineering*, **92**(10), 891-916, 2012.

[8] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, **242**, 623-647, 2013.

[9] M. Zahr, C. Farhat, Progressive construction of a parametric reduced-order model for PDE-constrained optimization. *International Journal for Numerical Methods in Engineering*, **102**(5), 1077-1110, 2015.

[10] D. Amsallem, M. Zahr, Y. Choi, C. Farhat, Design optimization using hyper-reduced-order models. *Structural and Multidisciplinary Optimization*, **51**(4), 919-940, 2015.

[11] A. Paul-Dubois-Taine, D. Amsallem, An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models. *International Journal for Numerical Methods in Engineering*, **102**(5), 1262-1292, 2015.

[12] C. Farhat, P. Avery, T. Chapman, J. Cortial, Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency. *International Journal for Numerical Methods in Engineering*, **98**(9), 625-662, 2014.

[13] C. Farhat, T. Chapman, P. Avery, Structure-preserving, stability, and accuracy properties of the Energy-Conserving Sampling and Weighting (ECSW) method for the hyper reduction of nonlinear finite element dynamic models. *International Journal for Numerical Methods in Engineering*, **102**(5), 1077-1110, 2015.

[14] D. Ryckelynck, A priori hyperreduction method: an adaptive approach. *Journal of Computational Physics*, **202**, 346-366, 2005.

[15] R. Ghanem, P.D. Spanos, *Stochastic Finite Elements: a Spectral Approach*. Springer-Verlag, New York, 1991.

[16] R. Ghanem, P.D. Spanos, *Stochastic Finite Elements: A spectral Approach (revised edition)*. Dover Publications, New York, 2003.

[17] C. Soize, R. Ghanem, Physical systems with random uncertainties: Chaos representation with arbitrary probability measure. *SIAM Journal On Scientific Computing*, **26**(2), 395-410, 2004.

[18] R. Mace, W. Worden, G. Manson, Uncertainty in Structural Dynamics, *Special issue of the Journal of Sound and Vibration*, **288**(3), 431-790, 2005.

[19] G.I. Schueller, Computational Methods in Stochastic Mechanics and Reliability Analysis. *Special issue of Computer Methods in Applied Mechanics and Engineering*, **194**(12-16), 1251-1795, 2005.

[20] G.I. Schueller, Uncertainties in Structural Mechanics and Analysis-Computational Methods. *Special issue of Computer and Structures*, **83**(14), 1031-1150, 2005.

[21] G.I. Schueller, Developments in stochastic structural mechanics. *Archive of Applied Mechanics*, **75**(10-12), 755-773, 2006.

[22] G. Deodatis, P.D. Spanos, 5th International Conference on Computational Stochastic Mechanics. *Special issue of the Probabilistic Engineering Mechanics*, **23**(2-3), 103-346, 2008.

[23] G.I. Schueller, H.J. Pradlwarter, Uncertain linear systems in dynamics: Retrospective and recent developments by stochastic approaches. *Engineering Structures*, **31**(11), 2507-2517, 2009.

[24] O.P. Le Maitre, O.M. Knio, *Spectral Methods for Uncerainty Quantification with Applications to Computational Fluid Dynamics*. Springer, Heidelberg, 2010.

[25] C. Soize, *Stochastic Models of Uncertainties in Computational Mechanics*. American Society of Civil Engineers (ASCE), Reston, 2012.

[26] T. Bui-Thanh, K. Willcox, O. Ghattas, Parametric reduced-order models for probabilistic analysis of unsteady aerodynamic applications. *AIAA Journal*, **46**(10), 2520-2529, 2008.

[27] C. Soize, Stochastic modeling of uncertainties in computational structural dynamics - Recent theoretical advances. *Journal of Sound and Vibration*, **332**(10), 2379-2395, 2013.

[28] Y.M. Marzouk, H.N. Najm, L.A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, **224**(2), 560-586, 2007.

[29] D. Galbally, K. Fidkowski, K. Willcox, O. Ghattas, Non-linear model reduction for uncertainty quantification in large scale inverse problems. *International Journal for Numerical Methods in Engineering*, **81**(12), 1581-1608, 2010.

[30] C. Lieberman, K. Willcox, O. Ghattas, Parameter and state model reduction for large scale statistical inverse problems. *SIAM Journal on Scientific Computing*, **32**, 2523-2542 , 2010.

[31] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data. *Computer Methods in Applied Mechanics and Engineering*, **199**(33-36), 2150-2164, 2010.

[32] C. Soize, A computational inverse method for identification of non-Gaussian random fields using the Bayesian approach in very high dimension. *Computer Methods in Applied Mechanics and Engineering*, **200**(45-46), 3083-3099, 2011.

[33] A. Nouy, C. Soize, Random fields representations for stochastic elliptic boundary value problems and statistical inverse problems. *European Journal of Applied Mathematics*, **25**(3), 339-373, 2014.

[34] T. Cui, Y.M. Marzouk, K. Willcox, Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering*, **102**(5), 966-990, 2015.

[35] C. Soize, Random vectors and random fields in high dimension. Parametric model-based representation, identification from data, and inverse problems. pp:1-65, in *Handbook for Uncertainty Quantification*, edited by Ghanem R, Higdon D, Owhadi H., Springer, Heidelberg, 2016.

[36] C. Soize A nonparametric model of random uncertainties for reduced matrix models in structural dynamics. *Probabilistic Engineering Mechanics*, **15**(3), 277-294, 2000.

[37] C. Soize, Maximum entropy approach for modeling random uncertainties in transient elastodynamics. *Journal of the Acoustical Society of America*, **109**(5), 1979-1996, 2001.

[38] C. Soize, Random matrix theory for modeling random uncertainties in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, **194**(12-16), 1333-1366, 2005.

[39] C. Soize, Random Matrix Models and Nonparametric Method for Uncertainty Quantification. pp:1-84, *Handbook for Uncertainty Quantification*, edited by Ghanem R, Higdon D, Owhadi H., Springer, Heidelberg, 2016.

[40] C.E. Shannon, A mathematical theory of communication. *Bell System Technical Journal*, **27**,379-423 and 623-659, 1948.

[41] E.T. Jaynes, Information theory and statistical mechanics. *Physical Review*, **106**(4), 620-630 and **108**(2), 171-190, 1957.

[42] J.N. Kapur, H.K. Kesavan, *Entropy Optimization Principles with Applications*. Academic Press, San Diego, 1992.

[43] C. Soize, Construction of probability distributions in high dimension using the maximum entropy principle. Applications to stochastic processes, random fields and random matrices. *International Journal for Numerical Methods in Engineering*, **76**(10), 1583-1611, 2008.

[44] A. Batou, C. Soize, Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension. *SIAM/ASA Journal on Uncertainty Quantification*, **1**(1),431-451, 2013.

[45] M.P. Mignolet, C. Soize, Nonparametric stochastic modeling of linear systems with prescribed variance of several natural frequencies. *Probabilistic Engineering Mechanics*, **23**(2-3), 267-278, 2008.

[46] C. Soize, Nonparametric probabilistic approach of uncertainties for elliptic boundary value problem. *International Journal for Numerical Methods in Engineering*, **80**(6-7), 673-688, 2009.

[47] C. Chen, D. Duhamel, C. Soize, Probabilistic approach for model and data uncertainties and its experimental identification in structural dynamics: Case of composite sandwich panels. *Journal of Sound and Vibration*, **294**(1-2),64-81, 2006.

[48] R. Ohayon, C. Soize, *Advanced Computational Vibroacoustics. Reduced-Order Models and Uncertainty Quantification*. Cambridge University Press, Cambridge, 2014.

[49] C. Soize, I.E. Poloskov, Time-domain formulation in computational dynamics for linear viscoelastic media with model uncertainties and stochastic excitation. *Computers and Mathematics with Applications*, **64**(11), 3594-3612, 2012.

[50] C. Soize, H. Chebli, Random uncertainties model in dynamic substructuring using a nonparametric probabilistic model. *Journal of Engineering Mechanics*, **129**(4),449-457, 2003.

[51] H. Chebli, C. Soize, Experimental validation of a nonparametric probabilistic model of non homogeneous uncertainties for dynamical systems. *Journal of the Acoustical Society of America*, **115**(2), 697-705, 2004.

[52] E. Capiez-Lernout, C. Soize, J.P. Lombard, C. Dupont, E. Seinturier, Blade manufacturing tolerances definition for a mistuned industrial bladed disk. *Journal of Engineering for Gas Turbines and Power*, **127**(3), 621-628, 2005.

[53] J. Duchereau, C. Soize, Transient dynamics in structures with nonhomogeneous uncertainties induced by complex joints. *Mechanical Systems and Signal Processing*, **20**(4), 854-867, 2006.

[54] M.P. Mignolet, C. Soize, J. Avalos, Nonparametric stochastic modeling of structures with uncertain boundary conditions / coupling between substructures. *AIAA Journal*, **51**(6), 1296-1308, 2013.

[55] J.F. Durand, C. Soize, L. Gagliardini, Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation. *Journal of the Acoustical Society of America*, **124**(3), 1513-1525, 2008.

[56] C. Fernandez, C. Soize, L. Gagliardini, Sound-insulation layer modelling in car computational vibroacoustics in the medium-frequency range. *Acta Acustica United with Acustica (AAUWA)*, **96**(3), 437-444, 2010.

[57] R. Ohayon, C. Soize, Advanced computational dissipative structural acoustics and fluid-structure interaction in low- and medium-frequency domains. Reduced-order models and uncertainty quantification. *International Journal of Aeronautical and Space Sciences*, **13**(2),127-153, 2012.

[58] C. Desceliers, C. Soize, S. Cambier, Non-parametric-parametric model for random uncertainties in nonlinear structural dynamics - Application to earthquake engineering. *Earthquake Engineering and Structural Dynamics*, **33**(3), 315-327, 2004.

[59] R. Cottereau, D. Clouteau, C. Soize, Construction of a probabilistic model for impedance matrices. *Computer Methods in Applied Mechanics and Engineering*, **196**(17-20), 2252-2268, 2007.

[60] R. Cottereau, D. Clouteau, C. Soize, Probabilistic impedance of foundation: impact of the seismic design on uncertain soils. *Earthquake Engineering and Structural Dynamics*, **37**(6), 899-918, 2008.

[61] E. Capiez-Lernout, C. Soize, Design optimization with an uncertain vibroacoustic model. *Journal of Vibration and Acoustics*, **130**(2), 1-8, 2008.

[62] E. Capiez-Lernout, C. Soize, Robust design optimization in computational mechanics. *Journal of Applied Mechanics - Transactions of the ASME*, **75**(2),1-11, 2008.

[63] M.P. Mignolet, C. Soize, Stochastic reduced order models for uncertain nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, **197**(45-48), 3951-3963, 2008.

[64] E. Capiez-Lernout, C. Soize, M.P. Mignolet, Post-buckling nonlinear static and dynamical analyses of uncertain cylindrical shells and experimental validation. *Computer Methods in Applied Mechanics and Engineering*, **271**(1), 210-230, 2014.

[65] C. Soize, C. Farhat, A nonparametric probabilistic approach for quantifying uncertainties in low- and high-dimensional nonlinear models. *International Journal for Numerical Methods in Engineering*, submitted 1st February, 2016.

[66] A. Edelman, T.A. Arias, S.T. Smith, The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, **20**(2), 303-353, 1998.

[67] P.A. Absil, R. Mahony, R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, 2008.

[68] G.H. Golub, C.F. Van Loan, *Matrix Computations (Third Edition)*. The Johns Hopkins University Press, Baltimore, 1996.

[69] K.J. Bathe, E.L. Wilson, *Numerical Methods in Finite Element Analysis*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

THE INFLUENCE OF MECHANICAL STIMULUS ON NUTRIENT TRANSPORT AND CELL GROWTH IN ENGINEERED CARTILAGE: A FINITE ELEMENT APPROACH

S. Cortez[1], **A. Completo**[2], **and J. L. Alves**[1]

[1] Center for MicroElectroMechanical Systems (CMEMS)
University of Minho
Campus of Azurém 4800-058 Guimarães, Portugal
{scortez, jlalves}@dem.uminho.pt

[2] Department of Mechanical Engineering
University of Aveiro
Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
completo@ua.pt

Keywords: tissue engineering, cartilage, nutrient transport, cell growth, dynamic compression

Abstract. *Osteoarthritis is a degenerative joint disease characterized by cartilage degeneration, which affects more people than any other joint disease. To overcome this problem, tissue engineering has been improving to restore tissue functionality developing new implantable cartilage. Bioreactors can provide an adequate nutrient transport to cells in the scaffold, a good mechanical stimuli and a hydrodynamic environment, which imposes a fluid flow to promote the cell proliferation. Finite element analysis of cartilage growth can be used to guide these tissue engineering experiments in order to produce cell scaffold constructs with specific biomechanical properties. In this work, a 3D finite element model was developed to simulate the diffusion and transport of nutrients and the cell growth kinetics in a porous scaffold when it is subject to different mechanical stimuli. The mass transport was defined by convection-diffusion equation, where the nutrient uptake was represented through the Michaelis-Menten kinetics and the fluid dynamics within the construct was modelled by Brinkman's equation. Cell growth kinetics was modelled by the Contois equation, which includes the effect of glucose concentration and the cell density saturation. The objective of this study was to understand the effect of mechanical loading in nutrition and cell growth during culture. Different mechanical stimuli (5%, 10% and 15% of compressive strain with frequencies of 0.5Hz, 1Hz and 2Hz) were performed during a period of 48 hours of cell culture. The spatial-temporal evolution of the local glucose, oxygen and lactate concentrations, the pH level and the cell density within the scaffold was analysed. The numerical simulations of solutes concentrations and cell growth show a good agreement with experimental results. These results show that fluid shear stress promoted by fluid dynamics inside the scaffold is influenced by compression and consequently the solute transport and the chondrocyte activity is affected for altered levels of stimulation.*

1 INTRODUCTION

In tissue engineering, development of cartilage depends on the exchange of a suitable amount of nutrients, including oxygen and glucose, to cells during the culture process [1-3]. The chondrocytes during the culture process produces a large amount of lactate and its production rate is related to the glucose and oxygen metabolism [4]. The diffusion process and metabolism of cells develop concentration gradients of metabolites that affect the cell viability [2, 5-7].

The culture process under dynamic flow conditions has shown significant potential for the development of cartilage tissue [8]. In this culture process, cells are seeded into three-dimensional scaffolds that are placed inside the bioreactor to provide nutrients and mechanical stimuli. When compared with the control constructs kept in a loading-free culture, the dynamic stimulation culture not only improves the mass exchange between cells and medium, but also exposes cells to physical stimulations [9]. The effects of dynamic compression may be due to fluid flow, tissue and cell deformation or hydrostatic pressurization. Compression-induced fluid flow increases the shear stress, the rate of transport of nutrients and the biomass production inside the scaffold [10].

Applying dynamic compression loading at moderate levels as 2-15% strains [8,11,12], 0.5-2.5 MPa stresses [13] and physiological frequencies between 0.001 to 1.0 Hz can stimulate the cell differentiation and the development of a functional tissue [9].

Despite the loading parameters to maximize the development of tissues in culture, other factors should be taken into account in tissue engineering, such as the type of the scaffold, the nutrient medium, and cellular factors, such as initial cell seeding density [14].

In this study, the effects of dynamic compression on metabolites transport and cellular biosynthesis in a chondrocyte-seeded agarose scaffold were investigated. This numerical model includes the consumption of glucose and oxygen, the production of lactate by cells, and the pH and shear stress effects on the cellular differentiation.

2 METHODOLY

2.1 Nutrient transport formulation

A mathematical formulation to describe the supply of metabolites to cells is presented. The time varying evolution of metabolites content in the three-dimensional scaffold is represented by diffusion-reaction and cell kinetics equations.

This numerical modelling assumes the diffusion of nutrients from a culture medium boundary condition and a consumption by cells within the scaffold matrix. The constitutive equations for each domain are follow detailed. Equations are written in the generic form for each metabolite i (oxygen, glucose and lactate). The concentration in time for each specie is governed by the Fick's law of diffusion:

$$\frac{\partial C^i}{\partial t} = \mathbf{D}^i \frac{\partial^2 C^i}{\partial x^2} + \dot{q} \tag{1}$$

and the oxygen and glucose uptake is assumed by Michaelis-Menten kinetics:

$$\dot{q}^i = -\frac{V^i_{\max} C^i}{K^i_m + C^i} \rho_{cell} \tag{2}$$

where C^i is the concentration of the solute i in the fluid phase, $\mathbf{D}^i$ the diffusion effective tensor, $V^i_{\max}$ the maximum uptake rate, K^i_m the half-maximum rate concentration and ρ_{cell} the cellular density.

Glucose and oxygen are consumed by cells, which synthesizes lactate into the scaffold. The content of lactate produced by chondrocytes is twice the number of moles of glucose consumption and a third of the number of moles of oxygen consumption [4] given by the following equation:

$$\dot{q}^{lact} = -2\dot{q}^{glu} + \frac{1}{3}\dot{q}^{ox} \tag{3}$$

The pH is linearly dependent on the concentration of lactate and its control can reduces the factor of two between the absorption of glucose and the lactate production. The pH less than 7 becomes a more acidic medium, causing an adverse effect on cell viability.

2.2 Cell balance

The cell kinetics includes cell migration, growth and death, and it is modelled by the Monod growth equation [15]. In this paper, the glucose concentration is used as the only limiting factor for cell growth, where a glucose concentration of 3.0×10^{-10} mol/mm^3 is considered critical for chondrocyte viability [7].

To incorporate the influence of the shear stress acting on the chondrocytes, a modulation function $g(|\tau|)$ was assumed, as described by the following relation [16]:

$$g(|\tau|) = \begin{cases} 0.6 + 4|\tau| & |\tau| \in [0,0.1[\text{Pa}, \\ 1 & |\tau| \in [0.1,0.6]\text{Pa}, \\ 2.5(1-|\tau|) & |\tau| \in]0.6,1]\text{Pa}, \\ 0 & |\tau| > 1\text{Pa}, \end{cases} \tag{4}$$

where $|\tau|$ is the normalized fluid shear stress.

In this model, only the beneficial effect of shear stress on cell growth has been included. As the cell growth can influences the porosity due to increase of the solid volume fraction, the Mackie-Meares equation is used to include the porosity effect on the diffusion coefficients [15]. As consequence, the metabolites diffusion will be affected by the porosity of the scaffold. The decrease of the scaffold permeability with an increase of solid volume fraction was modelled by the Carman-Kozeny equation [10,17] as:

$$k = k_0 \frac{n_f^3}{(1-n_f^2)} \tag{5}$$

where n_f is the fluid volume fraction and k_0 the initial permeability of the scaffold.

3 FINITE ELEMENT IMPLEMENTATION

The governing equations subject to initial boundary conditions were implemented using a computation tool (V-Biomech) [18] based on a finite element method (FEM) to solve the distribution of glucose, oxygen and lactate, and the cell differentiation in the scaffold in relation

to time. A quarter of a 3D disc with 10 mm diameter and 3 mm height was meshed with 512 hexahedral finite elements. (Fig.1).

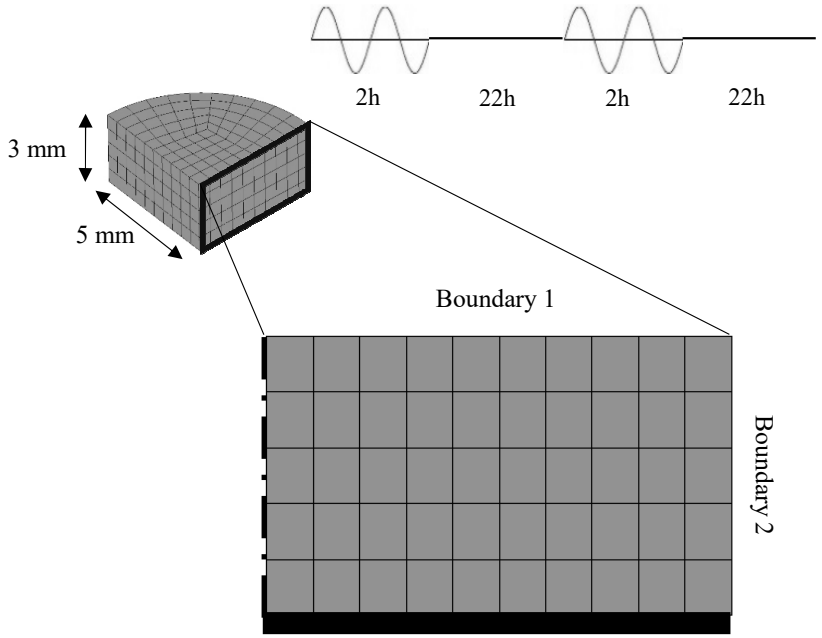


Figure 1: 3D mesh of the agarose disc submerged into a culture medium. Boundaries 1 and 2 are the surfaces in contact with the fluid. The cells are seeded into scaffold.

The model was based on a scaffold disc of 4% w/w agarose (n_f = 0.96 and k_0=0.661 mm^4 N^{-1}s^{-1}) with a uniform cell seeding of 40 000 cells/mm^3 and an initial oxygen, glucose and lactate concentration of 2.05×10^{-10} mol/mm^3, 5.10×10^{-9} mol/mm^3 and 1.20×10^{-9} mol/mm^3, respectively. The initial pH value of the scaffold was set at 7.4 and the diffusivity coefficients for each species are presented in Table 1.

The cell-seeded scaffold was considered to be submerged in a standard culturing environment of 5% CO_2 and 37ºC, with a boundary oxygen, glucose and lactate concentration of 2.05×10^{-10} mol/mm^3, 2.50×10^{-8} mol/mm^3 and 2.30×10^{-9} mol/mm^3, respectively. At the interface of the medium and the scaffold (boundaries 1 and 2), the concentrations are assumed to be continuous during the simulation and cells are not able to move out of the scaffold. Due to double symmetry about the centre line, only one quarter of the domain was computed.

Parameters	Value	Units	Reference
D^{ox}_{medium}	2.10×10^{-3}	mm^2/s	[19]
D^{glu}_{medium}	1.00×10^{-3}	mm^2/s	[20]
D^{lact}_{medium}	1.40×10^{-3}	mm^2/s	[20]
D_{cell}	1.70×10^{-8}	mm^2/s	[17]
K^{ox}_{m}	6.00×10^{-12}	mol/mm^3	[4]
K^{glu}_{m}	3.50×10^{-10}	mol/mm^3	[6]
$V^{ox}_{\max}$	1.18×10^{-18}	mol/(cell·s)	[4]
$V^{glu}_{\max}$	1.70×10^{-17}	mol/(cell·s)	[4]

Table 1: Diffusion parameters values used in the simulations.

In this study, two simulations were performed: (*i*) free-swelling culture simulation and (*ii*) dynamic compressive loading simulation, with compressive strains of 5%, 10% and 15% for frequencies of 0.5 Hz, 1 Hz and 2 Hz. Both simulations were tested during a period of 48 hours of cell culture. In the cyclic dynamic simulation, the displacement was applied on the top surface (boundary 1) of the construct during 2 hours and then a free-swelling was preserved during the next 22 hours. After this time, the stimulation was repeated during 2 hours and again, the free-swelling state was kept until reaches the 48 hours of the culture process. The average evolution within the scaffold of the local glucose, oxygen and lactate concentrations and the influence of the shear stress on cell kinetics was studied.

4 RESULTS AND DISCUSSION

A free-swelling (static) experiment and a dynamic compression with different combinations of strain (5%, 10% and 15%) and frequency (0.5 Hz, 1 Hz and 2 Hz) were simulated for a time culture of 48 hours and the evolution in time of the average glucose, oxygen and lactate concentrations and the cell density within the scaffold was analysed.

4.1 Influence on nutrient transport

The average oxygen concentration versus time curves at frequencies of 0.5 Hz, 1.0 Hz and 2 Hz are illustrated in Figure 2a, b and c, respectively. For each case, three different compressive strains, as 5%, 10% and 15% of the scaffold height were applied.

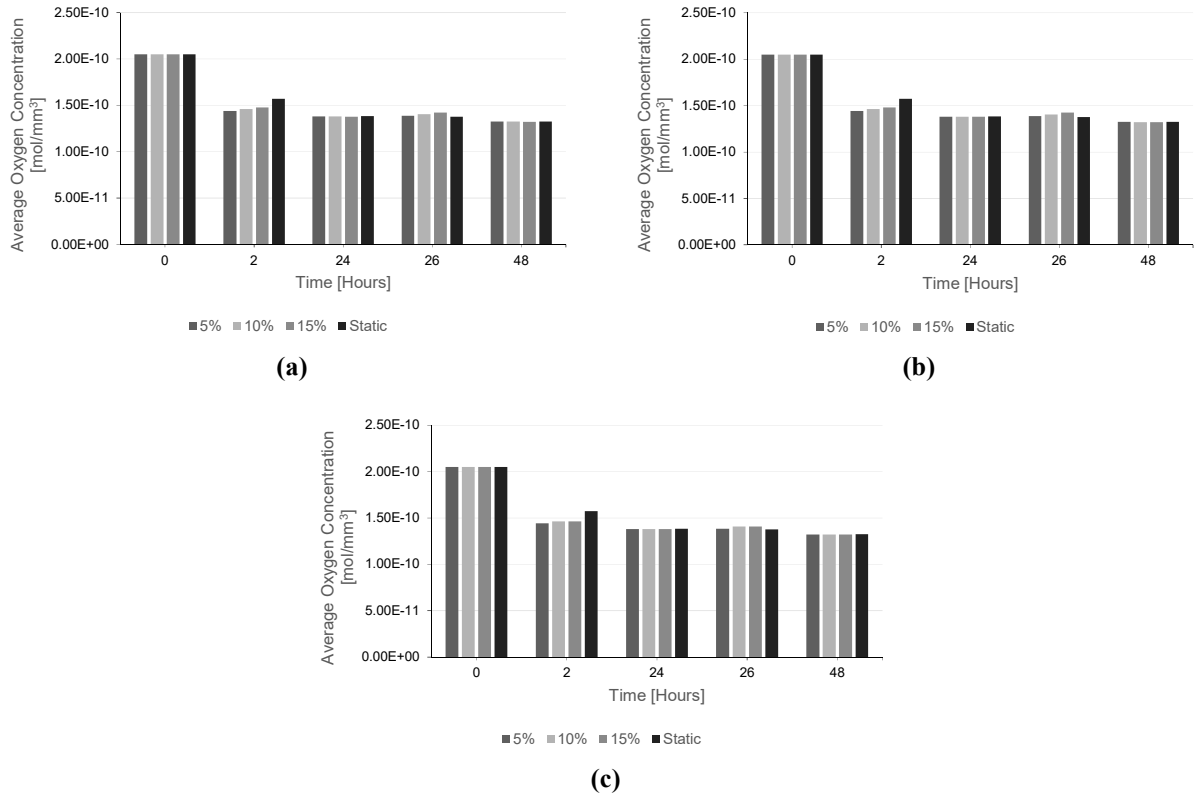


Figure 2: Average oxygen concentration for free-swelling conditions and for different dynamic compressive stimulations (5%, 10% and 15%) with frequencies of (a) 0.5 Hz, (b) 1 Hz and (c) 2 Hz.

All the three plots show a decrease on the average oxygen concentration inside the scaffold during the culture time, associated with the consumption of the oxygen inside the scaffold by chondrocytes. These results follow the expected trend showed in [2,4,6]. As in all simulations

the medium is constantly updated, the oxygen diffuses into the scaffold and tends to reach a steady state. The minimum value of oxygen concentration was 1.32×10^{-10} mol/mm^3. However, in cyclic compression stimulations the equilibrium of the oxygen concentration is reached faster (after 2 hours of incubation) than unloaded case, which can be associated with fluid velocities. After 24 hours of culture time, the chondrocytes consume oxygen at a slower rate. The 15% compression stimulations show to influence the oxygen concentrations due to the oxygen replacement rate induced by compression is larger than consumption rate by cells and consequently, the oxygen diffuses faster into the scaffold. At frequencies of 0.5 Hz and 1 Hz (Figure 2a and b), the 15% compression stimulation showed to be slightly higher than the results for 2 Hz cyclic compression simulation. Low frequencies results in low velocities and consequently a higher fluid flow infiltration depth.

Figure 3 shows the evolution of the total glucose content in the scaffold during the cultivation time of 48 hours. The average glucose concentration under the 0.5 Hz, 1.0 Hz and 2 Hz frequency is illustrated in Figure 3a, b and c, respectively.

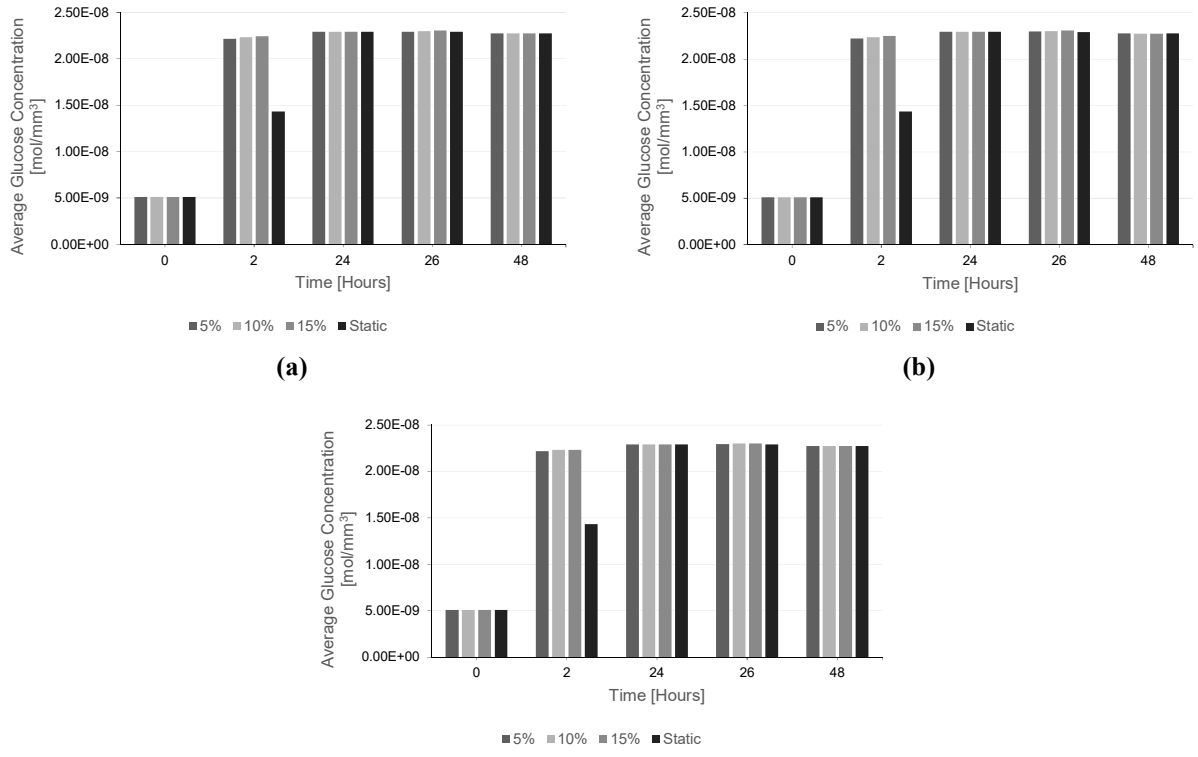


Figure 3: Average glucose concentration for free-swelling conditions and for different dynamic compressive stimulations (5%, 10% and 15%) with frequencies of (a) 0.5 Hz, (b) 1 Hz and (c) 2 Hz.

As the initial concentration inside the scaffold is lower than the medium, during the incubation time the glucose diffuses into the scaffold. As consequence, a softly increase trend of the average glucose concentration is observed for the unloaded simulation (free-swelling). Contrary to what occurs in oxygen plots, after 2 hours of cultivation the amount of glucose into the scaffold is higher in the stimulated case than for load-free simulation. Here, the steady state of glucose concentration is reached faster due to the fluid velocities imposed by compression. During the time, the glucose remained well above 1.0×10^{-8} mol/mm^3 reaching a maximum value of 2.29×10^{-8} mol/mm^3 at 26 hours. However, this value decreases slightly to

2.27×10^{-8} mol/mm^3 at the end of the culture period due to the cell consumption. Even at high cell densities when the consumption is higher, glucose remains constant due medium update (e.g., at 48 hours). Despite 15% compression stimulation showed to improve the glucose diffusion depth, the trend under different compressive stimulations remains identical. For this metabolite transport, frequency was not significant.

Figure 4a, b and c shows the average lactate concentration produced inside the scaffold under the 0.5 Hz, 1.0 Hz and 2 Hz frequency, respectively. Lactate is produced by cells and then diffuses to the medium. Its content increases gradually during incubation time as computed using Equation 3. The obtained results agree well with experimental data showed in some studies [2,4,6]. For the first two hours, the lactate content much higher in cyclic compression simulations than free-swelling case. It is associated with the cell density and the glucose inside the scaffold, which is higher for compressive stimulation. When incubation time reaches 24 hours, lactate secretion slows down. As glucose in the culture environment is mostly consumed and reaches an equilibrium state (Figure 3), the rate of lactate production gradually decrease on time.

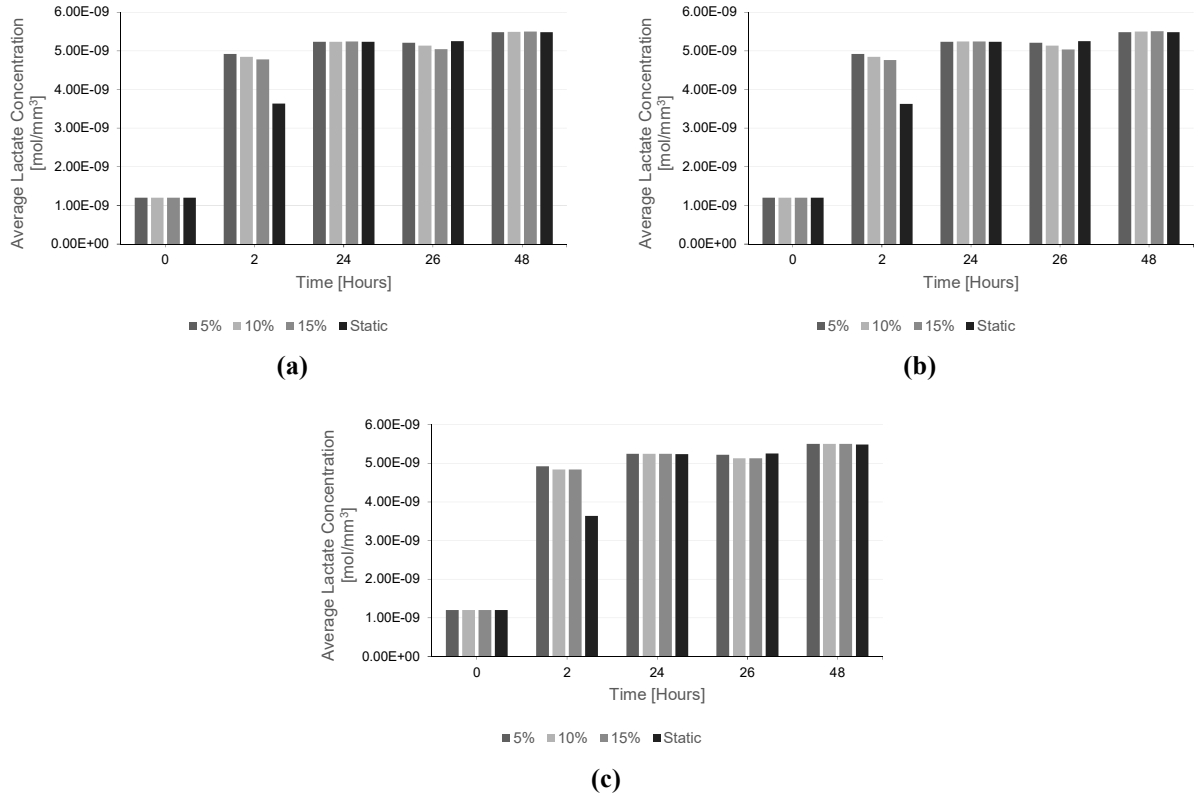


Figure 4: Average lactate concentration for free-swelling conditions and for different dynamic compressive stimulations (5%, 10% and 15%) with frequencies of (a) 0.5 Hz, (b) 1 Hz and (c) 2 Hz.

In cyclic compression simulations, when the compressive strain increases, the glucose content decreases. This phenomenon is associated with the increase of fluid velocities necessary to expel the lactate from the scaffold to the medium. The lactate removal is largest for 15% compression when the fluid velocity is higher.

4.2 Influence on cell proliferation

Figure 5a, b and c shows the average cell density inside the scaffold under the 0.5 Hz, 1.0 Hz and 2 Hz frequency, respectively. In all stimulation regimes of 5%, 10% and 15% compression loading, there is a higher proliferation of cells [2,4,6] comparing with the cell density obtained by free-swelling simulation. Observing Figure 2, the high cell growth complies with the low oxygen concentration.

When the scaffold is under compression, fluid culture medium is forced into the construct promoting the transport of nutrients to the cells but also generating a shear stress, which influence the cell kinetics. Compared to free-swelling culture outcome in Figure 4, more cells are resulted from dynamic loading cases, indicating the important roles of shear in promoting cell proliferation when nutrient supply is relatively constant.

The comparison between the simulations under the three different cyclic compressions shows that the fluid shear effect on cell growth is more significant at higher strains. For lower frequencies (Figure 4a and b) the 15% compressive stimulation significantly enhanced the proliferation of cells. However, the maximum value of 47379 cell/mm^3 was reached by 1 Hz frequency simulation. For a high frequency of 2 Hz (Figure 5c) the average cell density was almost the same for all stimulations.

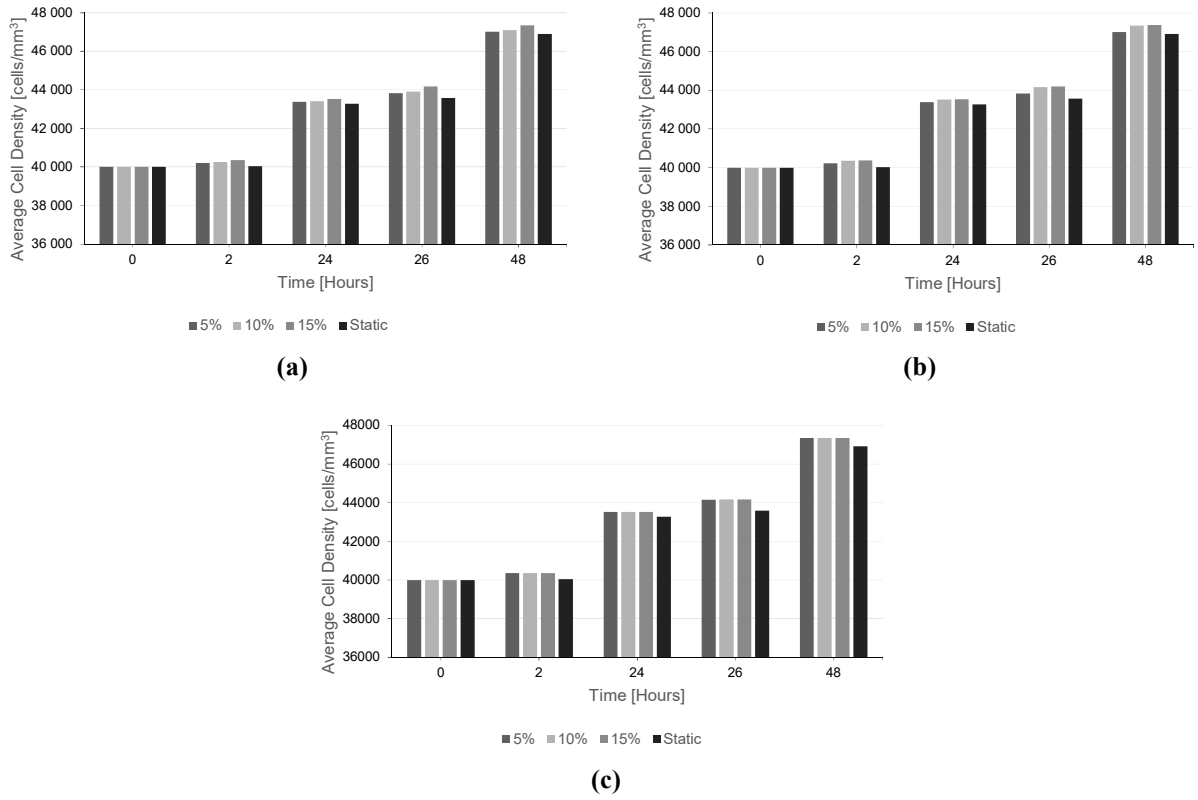


Figure 5: Average cell density for free-swelling conditions and for different dynamic compressive stimulations (5%, 10% and 15%) with frequencies of (a) 0.5 Hz, (b) 1 Hz and (c) 2 Hz.

Accompanying lactate accumulation in the scaffold, the culture environment becomes more acidic and the average pH decreases from 7.4. However, as the simulation time is too short (only 48 hours) a higher cell density was not achieved and the lactate production inside the scaffold was not sufficient to see the pH inhibition cell proliferation effects (when the pH is lower than 6.8).

5 CONCLUSIONS

In summary, this work numerically analysed the effects of dynamic compression on nutrient transport and cellular proliferation in a chondrocyte-seeded agarose scaffold. The results show that fluid shear stress promoted by fluid velocities inside the scaffold is influenced by compression and consequently the solute transport and the chondrocyte activity is affected for altered levels of stimulation. In this study, the simulation of 15% compressive strain with 1 Hz compressive loading for 2 days seemed to be beneficial for chondrocytes proliferation.

The present numerical model is a helpful tool to clarify experimental studies of cell viability and transport of nutrients over the development of tissue-engineered constructs during bioreactor culture under nutrition and mechanical loading conditions.

6 ACKNOWLEDGEMENTS

Sara Cortez is grateful to FCT (Fundação para a Ciência e Tecnologia) for the PhD grant (SFRH/BD/87933/2012). This work is supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 – Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941.

REFERENCES

[1] J. Malda, J. Rouwkema, D. E. Martens, E.P. le Comte, F.K. Kooy, J. Tramper, C.A. van Blitterswijk and J. Riesle, Oxygen gradients in tissue-engineered Pegt/Pbt cartilaginous constructs: Measurement and modelling, *Biotechnology and bioengineering*, **86**(1), 9-18, 2004.

[2] S. Li, R.O. Oreffo, B.G. Sengers, and R.S. Tare, The effect of oxygen tension on human articular chondrocyte matrix synthesis: Integration of experimental and computational approaches. *Biotechnology and bioengineering*, **111**(9), 1876-1885, 2014.

[3] B.G. Sengers, C.C. Van Donkelaar, C.W.J. Oomens, and F.P.T. Baaijens, Computational study of culture conditions and nutrient supply in cartilage tissue engineering. *Biotechnology Progress*, **21**(4), 1252-1261, 2004.

[4] B.G. Sengers, H.K. Heywood, D.A. Lee, C.W. Oomens and D.L. Bader, Nutrient utilization by bovine articular chondrocytes: a combined experimental and theoretical approach. *Journal of biomechanical engineering*, **127**(5), 758-766, 2005.

[5] C. Bandeiras, A. Completo, A. Ramos, A.T. Rufino, M. Ribeiro, J.P. Ferreira, and A.F. Mendes, Tissue Engineered Cartilage in Unconfined Compression: Biomechanical Analysis. *Materials Today: Proceedings*, **2**(1), 355-364, 2015.

[6] T.H. Lin, H.Y. Jhang, F.C. Chu and C.A. Chung, Computational modeling of nutrient utilization in engineered cartilage. *Biotechnology progress*, **29**(2), 452-462, 2013.

[7] S. Zhou, Z. Cui, and J.P. Urban, Nutrient gradients in engineered cartilage: metabolic kinetics measurement and mass transfer modeling.*Biotechnology and bioengineering*, **101**(2), 408-421, 2008.

[8] Y. H Tsuang, Y. S. Lin, L. T. Chen, C. K. Cheng, and J.S. Sun, Effect of dynamic compression on in vitro chondrocyte metabolism. *The International journal of artificial organs*, **31**(5), 439-449, 2008.

[9] D. Liu, C.K. Chua, and K.F. Leong, A mathematical model for fluid shear-sensitive 3D tissue construct development. *Biomechanics and modeling in mechanobiology*, **12**(1), 19-31, 2013.

[10] R. Sacco, P. Causin, P. Zunino, and M.T. Raimondi, A multiphysics/multiscale 2D numerical simulation of scaffold-based cartilage regeneration under interstitial perfusion in a bioreactor. *Biomechanics and modeling in mechanobiology*, **10**(4), 577-589, 2011.

[11] R.L. Mauck, S.B. Nicoll, S.L. Seyhan, G.A. Ateshian and C.T. Hung, Synergistic action of growth factors and dynamic loading for articular cartilage tissue engineering. *Tissue engineering*, **9**(4), 597-611, 2003.

[12] K.W. Ng, R.L. Mauck, L.Y. Statman, E.Y. Lin, G.A. Ateshian, and C.T. Hung, Dynamic deformational loading results in selective application of mechanical stimulation in a layered, tissue-engineered cartilage construct.*Biorheology*, **43**(3, 4), 497-507, 2006.

[13] M.T. Raimondi, M. Moretti, M. Cioffi, C. Giordano, F. Boschetti, K. Laganà, and R. Pietrabissa, The effect of hydrodynamic shear on 3D engineered chondrocyte systems subject to direct perfusion. *Biorheology*, **43**(3, 4), 215.-222, 2006.

[14] T. Mesallati, C.T. Buckley, T. Nagel, and D.J. Kelly, Scaffold architecture determines chondrocyte response to externally applied dynamic compression. *Biomechanics and modeling in mechanobiology*, **12**(5), 889-899, 2013.

[15] C. Bandeiras, A. Completo, and A. Ramos, Influence of the scaffold geometry on the spatial and temporal evolution of the mechanical properties of tissue-engineered cartilage: insights from a mathematical model. *Biomechanics and modeling in mechanobiology*, **14**(5), 1057-1070, 2015.

[16] M.M. Nava, M.T. Raimondi, and R. Pietrabissa, A multiphysics 3D model of tissue growth under interstitial perfusion in a tissue-engineering bioreactor. *Biomechanics and modeling in mechanobiology*, **12**(6), 1169-1179, 2013.

[17] C.A. Chung, C.W. Chen, C.P. Chen, and C.S. Tseng, Enhancement of cell growth in tissue-engineering constructs under direct perfusion: Modeling and simulation. *Biotechnology and bioengineering*, **97**(6), 1603-1616, 2007.

[18] J.L. Alves, N. Yamamura, T. Oda, C. Teodosiu, Numerical simulation of musculoskeletal systems by V-Biomech, *Proceedings of 9th International Symposium CMBBE2010*, Valencia, Spain, 24-27 February, 43-48, 2010.

[19] M.T. Raimondi, E. Bonacina, G. Candiani, M. Laganà, E. Rolando, G. Talò, D. Pezzoli, R. D'Anchise, R. Pietrabissa and M. Moretti, Comparative chondrogenesis of human cells in a 3D integrated experimental–computational mechanobiology model.*Biomechanics and modeling in mechanobiology*, **10**(2), 259-268, 2011.

[20] S. Hossain, D.J. Bergstrom, and X.B. Chen, A mathematical model and computational framework for three-dimensional chondrocyte cell growth in a porous tissue scaffold placed inside a bi-directional flow perfusion bioreactor.*Biotechnology and bioengineering*, **112**(12), 2601-2610, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

HYBRID CELL CENTRED/VERTEX MODEL FOR LARGE TISSUE DEFORMATIONS

José J. Muñoz[1], Payman Mosaffa[1], Yanlan Mao[2], Rob Tetley[2], Nina Asadipour[1] and Antonio Rodríguez-Ferran[1]

[1]Laboratori de Càlcul Numèric (LaCàN), Universitat Politècnica de Catalunya, 08034 Barcelona, Spain
{j.munoz, payman.mosaffa, nina.asadipour, antonio.rodriguez-ferran}@upc.edu
[2]MRC Laboratori for Molecular Cell Biology, University College London
{y.mao,r.tetley}@ucl.ac.uk

Keywords: Cell-centred, vertex, tissue mechanics, Voronoi, network, wound healing, embryogenesis

Abstract. *Macroscopic deformations in embryonic soft tissues are due to the intra-cellular remodelling and cell intercalation. We here present a computational approach that can handle the two types of deformations, and also take into account the active cell response. The model resorts to cell-centred techniques, where particles represent cell nuclei, and to vertex models, where the vertices represent cell boundaries. This hybrid approach allows to consider separately intracellular and inter-cellular forces, and at the same time impose cell incompressibility.*

In the proposed model, the cell boundaries (defined by vertices) and cell nuclei (or cell-centres) networks are coupled through an interpolation scheme, which is eventually relaxed in order to smooth the cell boundaries. We show that this coupling between the two networks modifies the equilibrium equations and stabilises the vertex network. Incompressibility is implemented through a penalty method. The resulting model can be implemented in two- and three-dimensions, and is complemented with active rheological models.

We apply the model to simulate the stretching and relaxation of cell monolayers, and to simulate wound healing process in the wing disc of Drosophila fly embryo. We show that the numerical results agree with the experimental measurements.

1 INTRODUCTION

The development of computational models for studying the behaviour of biological tissue at cellular scale has progressively become a promising trend in modern biology. Due to the complexity of phenomena at cellular level, novel mechanical modelling is required to handle the observed tissue behaviour as well as being predictive to similar phenomena.

Due to the discrete nature of the tissue, and the relevance of the adhesive forces at the cell-cell junctions [2, 10] we resort to a cell-based approach which couples cell-centred [9] and vertex models [4, 8]. The two sets of points form two networks that we couple by using an appropriate interpolation. The combination of the two networks allows us modelling cellular rheology and adhesive forces in a joint manner, and inherit the advantages of cell-centres models (easy remodelling) and vertex models (explicit representation of boundary forces and area/volume preservation).

The developed model is employed to simulate embryogenesis of Drosophila wing disc [5] and a wound healing, which are experimentally analysed at the MRC-Laboratory for Molecular Cell Biology, at Universisty College London. In this process, it is observed that the mechanical signalling and the activity of the stress fibres plays a predominant role, which is as determinant as the genetic and biochemical signalling.

We first describe the kinematic description of the network and the coupling between them. Two approaches for describing the cell boundaries are described: Voronoi and barycentric tessellation. We resort to the latter approach and justify our approach in Section 2. In Section 3 we deduce the mechanical equilibrium equations for arbitrary cell rheology. The latter is defined in Section 4, where we describe a specific viscoelastic model based on an evolution equation of the resting length. Some illustrative results are presented in Section 5, where we compare the experimental and computational deformations of the embryonic tissue.

2 CELL NETWORK DEFINITION

2.1 General approach

We concentrate our attention on flat monolayers where the cell deformations along the thickness direction is assumed constant, and thus our description is restricted to the two-dimensional case. Further extensions of the model to curved monolayers may be found in [6].

The cell-cell connectivity of the tissue is defined by a set of N nodes $\boldsymbol{X} = \{\boldsymbol{x}_1 \dots \boldsymbol{x}_N\}$ that represent the cell nuclei (cell-centres) and a set of N_e bar elements connecting them and defined by a connectivity matrix $\boldsymbol{C} = \{\boldsymbol{c}_1 \dots \boldsymbol{c}_{N_e}\}$, with $\boldsymbol{c}_e = \{i\ j\}^T$ and i and j the nodes the element e connects. We will resort here to a Delaunay triangulation [3, 1] in order to define the cell-cell connectivity. Since Delaunay triangulations form convex hulls of the set of nodes, we filter those triangles with very high or low aspect ratio. More precisely, we remove those triangles whose inradius r and circumradius R satisfy

$$\frac{r}{R} < \mathrm{tol}_R$$

where tol_R is a tolerance here set to $\mathrm{tol}_R = 0.2$.

Based on Delaunay triangulation, we use a tessellation defined by a set of N_v connected vertices $\mathbf{V} = \{\boldsymbol{v}_1 \dots \boldsymbol{v}_{N_v}\}$, which represent the cell-cell boundaries. We resort to two types of tessellations:

Voronoi tessellation: vertices form the Voronoi diagram of the Delaunay triangulation [1]. Unbounded Voronoi areas located at the boundary of the tissue are removed.

Barycentric tessellation: vertices are formed by the barycentres of the triangles. Each cell boundary is formed by joining the triangles that surround it, and external nodes do not form a cell.

Figure 1a-b shows the Delaunay triangulation and two tessellations described above. The vertices $\boldsymbol{v}_I$ of the two tessellations can be computed by using standard Finite Element interpolations $N_i(\boldsymbol{\xi})$ of linear triangle elements as

$$\boldsymbol{v}_I = N_i(\boldsymbol{\xi}_I)\boldsymbol{x}_i \tag{1}$$

where summation is understood on the three nodes that form triangle I (see Figure 2). The interpolation parameter $\boldsymbol{\xi}_I$ takes different values depending on the choice of the tessellation, and corresponds to the coordinate of vertex $\boldsymbol{v}_I$ in the reference element. In Voronoi tessellations, $\boldsymbol{\xi}_I$ for each triangle can be computed from the actual position of $\boldsymbol{v}_I$ in the Voronoi diagram by solving a system of linear equations. In barycentric tessellations instead we have that $\boldsymbol{\xi}_I = \{1/3,\ 1/3\}$ for all vertices.

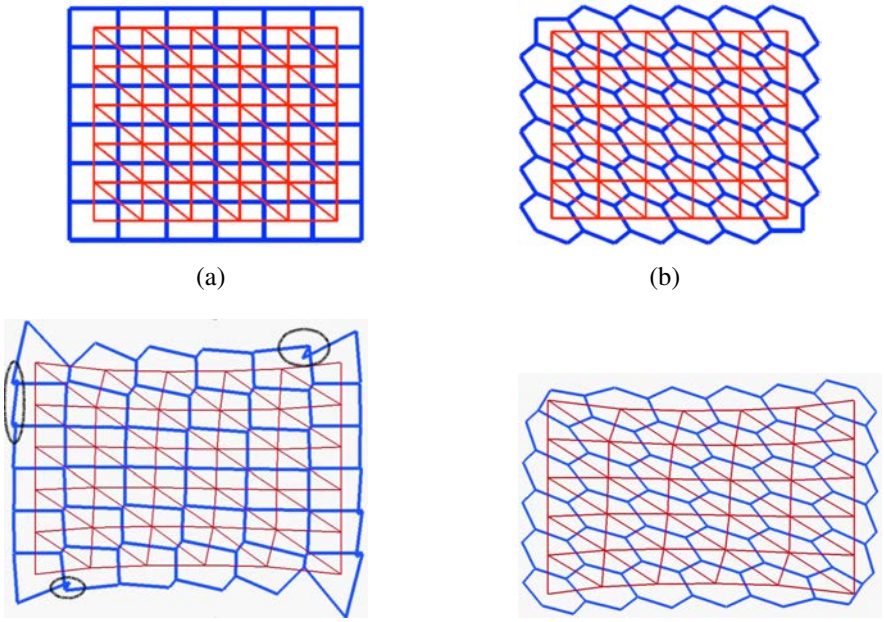


Figure 1: Triangulation and tessellation of a Cartesian 5×5 mesh of nodes. (a) Delaunay triangulation and Voronoi tessellation. (b) Delaunay triangulation and barycentric tessellation. (c) Same triangulation and tessellation as in (b), but with an applied displacement on the nodes. Crossovers of original rectangular areas are indicated with brushed ellipses. (d) Same triangulation and tessellation as in (c), but with an applied displacement on the nodes.

We note that in Delaunay triangulations, Voronoi tessellations guarantee that the cells are convex, but we do not necessarily have that

$$\boldsymbol{v}_I \in T(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3), \tag{2}$$

with $T(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ the triangle formed by nodes $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. Instead, vertices in barycentric tessellations satisfy the property in (2), although for very distorted meshes, cell centres may lay outside the cell boundary. More importantly though, if the value $\boldsymbol{\xi}_I$ is kept constant, while the nodal positions $\boldsymbol{x}_i$ are displaced, the Voronoi areas may easily yield non-convex polytopes or

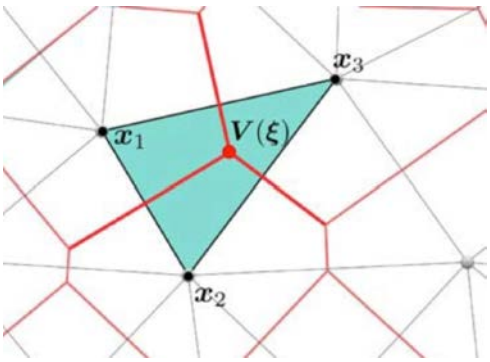


Figure 2: Schematic of the interpolation of the vertices on the Delaunay network.

with crossovers, giving rise to not valid tessellations, as Figure 2c shows. The displacement of the nodes with a constant interpolation parameter is considered in the next section. For this reason, only barycentric tessellations will be considered henceforth.

3 EQUILIBRIUM EQUATIONS

3.1 Elastic potentials

Mechanical equilibrium is here considered by assuming an elastic potential for each bar of the Delaunay triangulation and also for each segment of the tessellation. More specifically, and for simplicity, the following quadratic elastic potentials for the Delaunay (D) and vertices in the tessellation (V) are considered:

$$\begin{aligned} \Psi_D(\boldsymbol{x}) &= \frac{1}{2} k_D \sum_e \varepsilon_e^2, \\ \Psi_V(\boldsymbol{x}, \boldsymbol{\xi}) &= \frac{1}{2} k_V \sum_E \varepsilon_E^2, \end{aligned} \tag{3}$$

where $k_D > 0$ and $k_V > 0$ are the material stiffness constants. The elastic strains ε_e and ε_E are defined by

$$\begin{aligned} \varepsilon_e &= \frac{l^{ij} - L^{ij}}{L^{ij}}, \\ \varepsilon_E &= \frac{l^{IJ} - L^{IJ}}{L^{IJ}}. \end{aligned} \tag{4}$$

where bar e joints nodes i and j, and segment E joints vertices I and J. The lengths l^{ij} and l^{IJ} are the current (deformed) lengths, while L^{ij} and L^{IJ} are the resting or reference lengths, that will be further explained in Section 4. The current lengths are defined by

$$\begin{aligned} l^{ij} &= ||\boldsymbol{x}^i - \boldsymbol{x}^j||, \\ l_E^{IJ} &= ||\boldsymbol{v}^I - \boldsymbol{v}^J||. \end{aligned}$$

We note that although l^{IJ} depends on the positions $\boldsymbol{v}_I$, due to interpolation (1), the elastic potential Ψ_V depends on the nodal positions $\boldsymbol{x}_i$ and the parametric coordinates $\boldsymbol{\xi} = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_v}\}$, as indicated in (3). Also, due to the non-linear dependence of the lengths on nodal and vertex coordinates, the elastic potentials are non-linear, even if the reference length L is constant.

3.2 Area preservation

In order to impose that the areas of the tessellation are preserved, the following penalty term is considered:

$$\Psi_A(\boldsymbol{x},\boldsymbol{\xi}) = \frac{1}{2}k_A \sum_i (A_i - A_i^0)^2 \tag{5}$$

where $k_A > 0$ is a penalty factor, and A_i and A_i^0 are the current and initial areas, associated to each node i not located at the boundary. The areas A_i are in our case computed resorting to the divergence theorem and from the vertices positions surrounding each node i.

3.3 ξ-relaxation

In some occasions it becomes beneficial to relax the condition $\boldsymbol{\xi}_I = constant$. Indeed, as it will be shown in the next subsection, the interpolation in (1) implicitly prescribes the vertices' positions. This may have undesirable effects on the boundaries, and for this reason, the parametric coordinates of the vertices $\boldsymbol{\xi}_I$ are allowed to vary. This is numerically achieved by adding at each time-step t_{n+1} a penalty term with the form

$$\Psi_\xi^{n+1}(\boldsymbol{\xi}) = \frac{1}{2}k_\xi \sum_I ||\boldsymbol{\xi}_I^{n+1} - \boldsymbol{\xi}_I^n||^2. \tag{6}$$

with $k_\xi > 0$ a penalty parameter, and the superscript n indicating quantities at time t_n. We note that after assigning the value $k_\xi = \frac{\hat{k}_x i}{\Delta t}$, this penalty term is equivalent to adding a viscous term in the motion of the vertices with respect to the nodal positions $\boldsymbol{x}_i$.

3.4 Solution of equations

The equations of motion are solved at each time t_{n+1} by minimising the total potential

$$\Psi^{n+1}(\boldsymbol{x},\boldsymbol{\xi}) = \Psi_D^{n+1}(\boldsymbol{x}) + \Psi_V^{n+1}(\boldsymbol{x},\boldsymbol{\xi}) + \Psi_A^{n+1}(\boldsymbol{x},\boldsymbol{\xi}) + \Psi_\xi^{n+1}(\boldsymbol{\xi}) \tag{7}$$

with respect to the nodal positions $\boldsymbol{x}_i$ and parametric coordinates $\boldsymbol{\xi}_I$. This is equivalent to solving the following system of equations:

$$\left\{ \begin{array}{c} \boldsymbol{g}_x \\ \boldsymbol{g}_{\boldsymbol{\xi}} \end{array} \right\} := \left\{ \begin{array}{c} \nabla_x \Psi^{n+1} \\ \nabla_{\boldsymbol{\xi}} \Psi^{n+1} \end{array} \right\} = \boldsymbol{0} \tag{8}$$

We note that in the particular case that $k_D > 0$ and $k_V = k_A = k_\xi = 0$, the previous set of equations is equivalent to solving the nodal equilibrium equations:

$$\frac{\partial \Psi_D(\boldsymbol{x})}{\partial \boldsymbol{x}_i} = \boldsymbol{0} \Rightarrow \sum_{e \in i} \boldsymbol{t}_{ei} = \boldsymbol{0},\ i = 1, \ldots, N \tag{9}$$

with summation on all elements e connected to node i, and

$$\boldsymbol{t}_{ei} := k_D \varepsilon_e \frac{\partial \varepsilon_e}{\partial \boldsymbol{x}^i} \tag{10}$$

the traction force at node $\boldsymbol{x}^i$ along the bar element $\boldsymbol{c}^e = \{i\ j\}$. Other expressions of $\boldsymbol{t}_{ei}$ may be derived using more sophisticated elastic potentials $\Psi_V(\xi)$ in (3). In the work presented here, the

non-linear rheology will be introduced in the next section by using a variable resting length L^{ij}. In the case $k_V > 0$, and using the interpolation in (1), we have the following non-zero term,

$$\nabla_{\boldsymbol{x}} \Psi_V = \frac{\partial \Psi_V}{\partial \boldsymbol{v}_I} \frac{\partial \boldsymbol{v}_I}{\partial \boldsymbol{x}_i} = \frac{\partial \Psi_V}{\partial \boldsymbol{v}_I} \mathbf{P}_{Ii} \tag{11}$$

with

$$\mathbf{P}_{Ii} := \frac{\partial \boldsymbol{v}_I}{\partial \boldsymbol{x}_i} = N_i(\boldsymbol{\xi}_I)\mathbf{I}$$

When $\boldsymbol{\xi}_I = constant$, only the first block of equations in (8) needs to be solved. In this case, the system of equations in (9) turns into

$$\sum_{e \in i} \boldsymbol{t}_{ei} + \sum_{I \in i} \mathbf{P}_{Ii}^T \boldsymbol{t}_I = \mathbf{0},\ i = 1, \ldots, N \tag{12}$$

with,

$$\boldsymbol{t}_I = \frac{\partial \Psi_V}{\partial \boldsymbol{v}_I} = k_V \varepsilon_e \frac{\partial \varepsilon_e}{\partial \boldsymbol{v}_I} \tag{13}$$

the traction at vertex I and with the second summation in (12) on all the vertices surrounding node i. The equation above reveals two consequences of the coupling: the nodal equilibrium equations have contribution from the vertices, and the interpolation in (1) modifies the tractions at the vertices $\boldsymbol{t}_I$ through matrices $\mathbf{P}_{Ii}$. As a consequence, equilibrium at vertices is not fully preserved, even if $k_D = 0$. This fact has motivated the potential considered in the $\boldsymbol{\xi}$-relaxation.

When $\boldsymbol{\xi}_I$ is allowed to vary, the second block of equations in (8) reads,

$$\left(\frac{\partial \Psi_V}{\partial \boldsymbol{v}_J} + \frac{\partial \Psi_A}{\partial \boldsymbol{v}_J} \right) \mathbf{Q}_{JI} + k_\xi (\boldsymbol{\xi}_I^{n+1} - \boldsymbol{\xi}_I^n) = \mathbf{0},\ I = 1, \ldots, N_v$$

with

$$\mathbf{Q}_{JI} := \frac{\partial \boldsymbol{v}_J}{\partial \boldsymbol{\xi}_I} = \boldsymbol{x}_i \otimes \nabla N_i(\boldsymbol{\xi}_J)\delta_{IJ}$$

The non-linear equations are solved resorting to a Newton-Raphson scheme, which requires the computation of the Jacobian, which has the following general structure:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{xx} & \mathbf{K}_{x\xi} \\ \mathbf{K}_{\xi x} & \mathbf{K}_{\xi\xi} \end{bmatrix}$$

4 CELL RHEOLOGY AND CONTRACTILITY

So far, only elastic forces are considered, and the sole time-dependent term is the $\boldsymbol{\xi}$-relaxation term, which i equivalent to a viscous force acting on the relative motion of vertices with respect to the nodes. We introduce here an additional viscous-like rheology by using the following evolution of the resting law:

$$\frac{\dot{L}}{L} = \gamma \varepsilon_e \tag{14}$$

where γ is the *remodelling rate*, and represents the ability of the tissue to adapt to the elastic strain. As shown in [7], this evolution law of the resting length induces a similar response to the Standard Maxwell model. Furthermore, it also allows simulating cell contractility by using a modified version of the law in (14) as

$$\frac{\dot{L}}{L} = \gamma(\varepsilon_e - \varepsilon_c) \tag{15}$$

where ε_c represents a contractile strain that the cell asymptotically aims to achieve. This law can be either applied to the bars at the Delaunay network and at the segments of the cell boundaries. This is implemented by, according to the strain definitions in (4), using the following expression in (10) and (13),

$$\frac{\varepsilon_e}{\boldsymbol{x}_i} = \left(\frac{1}{L} - \frac{l}{L}^2 \frac{\partial L}{\partial l}\right) \frac{\partial l}{\partial \xi_i}$$
$$\frac{\varepsilon_e}{\boldsymbol{x}_i} = \mathbf{P}_{Ii}^T \left(\frac{1}{L} - \frac{l}{L}^2 \frac{\partial L}{\partial l}\right) \frac{\partial l}{\partial \boldsymbol{v}_I}$$

with

$$\frac{\partial l}{\partial \xi_i} = \frac{1}{l}(\xi_i - \xi_j)$$
$$\frac{\partial l}{\partial \boldsymbol{v}_I} = \frac{1}{l}(\boldsymbol{v}_I - \boldsymbol{v}_J)$$

and computing $\frac{\partial L}{\partial l}$ from a time-discretisation of the differential equation in (15).

5 RESULTS

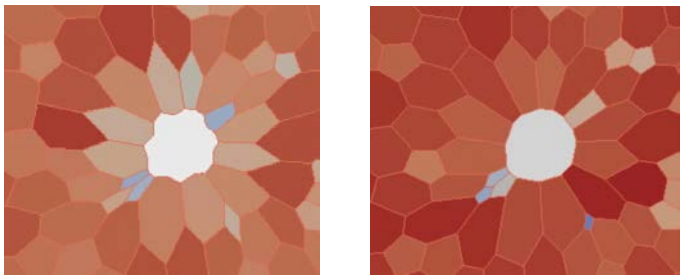
The MRC-Laboratory for Molecular Cell Biology at University College London is currently experimentally analysing the wound healing process of embryonic tissues of Drosophila fly. In this process, the mechanical signalling and the activity of the stress fibres play a predominant role that is as important as the genetic and biochemical signalling.

The locations of the cells have been obtained from the phase contrast images, and the ablation of the tissue has been simulated by setting $k_D = 0$ on those bars connected to ablated cells, and $k_V = 0$ on those segments located at boundaries between dead cells.

We have resorted to the hybrid model in order to simulate the healing process. The material properties k_D and k_V have been fitted in order to match the evolution of the rate of the wound closure and the expansion after ablation. No areas preservation has been included, since this has not been found to be necessary.

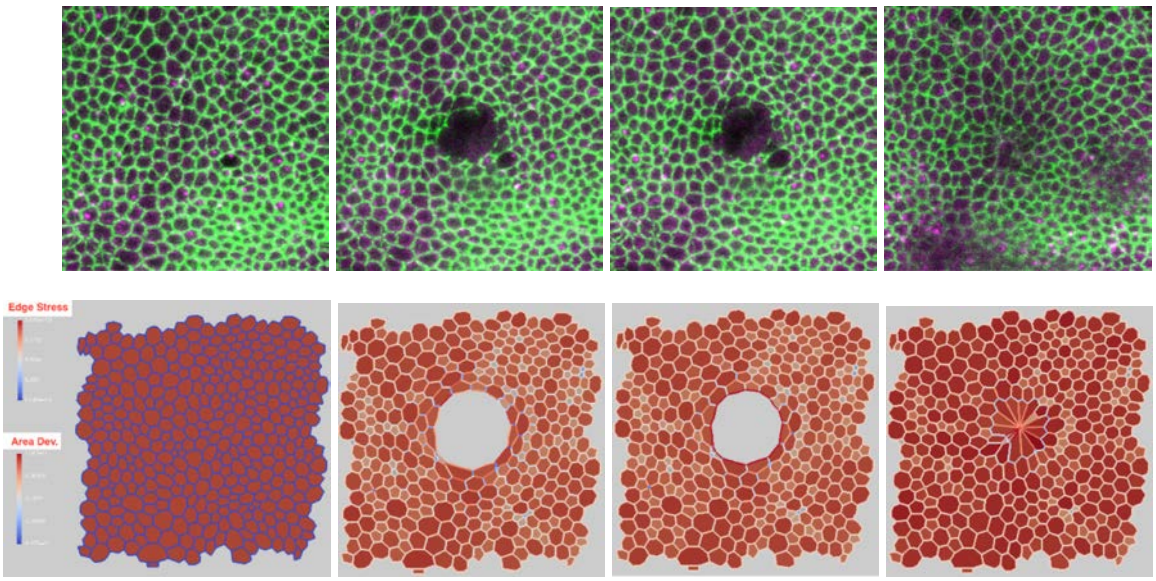
The equations using $\boldsymbol{\xi}$-relaxation have become necessary in order to avoid a zig-zag effect on the wound edge. Figure 3 compares the different wound profiles obtained without and with variation on parameter $\boldsymbol{\xi}$. As it can be observed, the interpolation constraint has detrimental effects on the profile of the wound, which is smoothed after incrementally relaxing the positions of the vertices.

The wound closure is simulated by imposing a linearly incremental contractility ε_c at the wound edge. The sequence of wounded areas experimentally measured are compared to the numerical simulations in Figure 4. While the evolution and expansion of the wound and the trend of the closure fit the microscopic images, the model needs to be extended with remodelling capabilities in order to better simulate the final stages of the reparing process.



(a) (b)

Figure 3: Effect of ξ-relaxation. (a) Wound edge without ξ-relaxation. (b) Wound edge applying ξ-relaxation.



(a) (b) (c) (d)

Figure 4: Response of tissue to applied damage by laser ablation of cells. Top: experimental images. Bottom: results of computational model. (a) Tissue before ablation, (b) Tissue recoil right after laser ablation applied. (c) Concentration of Myosin at the wound's edges causing wounds edge contractility. (d) Tissue after wound closure.

6 CONCLUSIONS

Wound closure of monolayers on elastic process has been well inspected [2]. However, the *in vivo* healing process is far more uncertain. The mechanical signalling and the tissue polarisation are currently solid hypotheses. These though have not been quantitatively measured or simulated.

By bridging the experimental measurements with rigorous mathematical and mechanical modelling we aim to couple the biochemical signalling and the forces exerted within the tissue. This understanding is necessary to predict tissue behaviour in other developmental processes and more generally in regenerative medicine.

The hybrid model and the numerical results presented here shows that the *in vivo* wound healing process can be successfully simulated. The model though requires to be complemented with remodelling capabilities. In order to inspect different mechanisms along the tissue thickness, it is also desirable to extend the model for the three-dimensional case. These features are currently investigated.

REFERENCES

[1] CB Barber, DP Dobkin, and H.T. HT Huhdanpaa. The Quickhull algorithm for convex hulls. *ACM Trans. Math.. Soft.*, 22(4):469–483, 1996. http://www.qhull.org.

[2] A Brugués, E Anon, V Conte, JH Veldhuis, M Gupta, J Collombelli, J J Muñoz, GW Brodland, B Ladoux, and X Trepat. Forces driving epithelial wound healing. *Nature Phys.*, 10:683–690, 2014.

[3] B. Delaunay. Sur la sphère vide. *Bull. Acad. Sc. URSS, cl. sc. math. nat.*, 6:793–800, 1934.

[4] H. Honda, M. Tanemura, and T. Nagai. A three- dimensional vertex dynamics cell model of space-filling polyhedra simulating cell behavior in a cell aggregate. *J. Theor. Biol.*, 226:439–453, 2004.

[5] Y. Mao, A.L. Tournier, A. Hoppe, L. Kester, B.J. Thompson, and N. Tapon. Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. *EMBO J.*, 32(21):2790, 2803.

[6] P. Mosaffa, N. Asadipour, D Millán, A. Rodríguez-Ferran, and J.J. Muñoz. Cell-centred model for the simulation of curved cellular monolayers . *Comp. Part. Mech.*, 2(4):359–370, 2015.

[7] J.J. Muñoz and S. Albo. Physiology-based model of cell viscoelasticity. *Phys. Rev. E*, 88(1):012708, 2013.

[8] S Okuda, Y Inoue, M Eiraku, Y Sasai, and T Adachi. Modeling cell proliferation for simulating three-dimensional tissue morphogenesis based on a reversible network reconnection framework. *Biomech. Model. Mechanobiol.*, 12:987–996, 2013.

[9] P Pathmanathan, J Cooper, A Fletcher, G Mirams, P Murray, J Osborne, J Pitt-Francis, A Walter, and S J Chapman. A computational study of discrete mechanical tissue models. *Phys. Biol.*, 6:036001, 2009.

[10] M Rauzi, P F Lenne, and T Lecuit. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature*, 468:1110–1114, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SENSITIVITY OR TEMPERATURE FIELD IN THE SYSTEM PROTECTIVE CLOTHING – FOREARM WITH RESPECT TO PERTURBATIONS OF EXTERNAL HEATING CONDITIONS

Bohdan Mochnacki[1], **Mariusz Ciesielski**[2]

[1] University of Occupational Safety Management in Katowice
Bankowa 8, 40-007 Katowice, Poland
e-mail: bmochnacki@wszop.edu.pl

[2] Institute of Computer and Information Sciences, Czestochowa University of Technology
Dabrowskiego 69, 42-200 Czestochowa, Poland
mariusz.ciesielski@icis.pcz.pl

Keywords: Bio-heat transfer, Protective clothing, Sensitivity analysis, Control Volume Method, Numerical simulations.

Abstract. *In the paper the problem concerning the numerical modeling of thermal processes in domain of biological tissue secured by a layer of protective clothing being in the thermal contact with the environment is discussed. The cross-section of the forearm (2D problem) is treated as the non-homogeneous domain in which the sub-domains of skin tissue, fat, muscle, bone and blood vessels are distinguished. Between skin tissue and protective clothing the air gap is taken into account. This sub-domain is treated as a solid body which thermal conductivity is defined in the special way. The process of external heating is determined by the Neumann boundary condition and the sensitivity analysis with respect to the perturbations of the boundary heat flux is discussed. Both the basic boundary-initial problem and the sensitivity one are solved by means of the control volume method using the Voronoi polygons.*

1 INTRODUCTION

The problem of skin tissue heating including the layer of protective clothing is described by the system of partial differential equations (energy equations), the boundary condition given on the external surface of the system, the boundary conditions between skin tissue and protective clothing, the boundary conditions on the surfaces limiting the successive sub-domains of forearm and the initial conditions. In the version presented in this paper the air gap between fabric and skin tissue is treated as a solid body (see: next chapter) and on its boundaries the continuity of temperature field and heat fluxes are assumed.

The transient temperature field in the tissue sub-domains is determined by the Fourier-type equation called the Pennes equation [1-7]. This equation contains two additional components (the source functions), this means the perfusion heat source and the metabolic heat source. The Pennes equation belongs to the group of macroscopic tissue models. In the recent years the others model based on the Cattaneo-Vernotte equation (e.g. [8]) or the dual phase lag equation (e.g. [9]) appeared, but the Pennes approach is, so far, most commonly used. The domain considered (a section of forearm – 2D problem) is the non-homogeneous one and constitutes a composition of skin tissue, fat, muscle, bone and blood vessels (arteries and veins). The successive sub-domains differ in the values of thermal parameters, in this paper the data taken from [10] are applied. Temperature fields in the domains of fabric and trapped air are described by the well known Fourier equation (diffusion equation). The thermophysical parameters of textiles can be found in [11].

At the stage of numerical computations the Control Volume Method (CVM) is applied, in other words, the domain considered is divided into a certain number of small cells and the governing equations in the integral form are used individually to each one of them. This procedure guarantees, a priori, the conservation of physical quantities like mass, momentum and energy, is extremely flexible and conceptually simple. In this paper, the 2D control volumes correspond to the Voronoi polygons (also called the Thiessen or Dirichlet cells in two dimensions) have been used. Such a version of CVM was in details discussed by Ciesielski and Mochnacki in [12].

The sensitivity analysis presented here concerns the changes of transient temperature field in domain considered due to the perturbations of the boundary heat flux determined by the Neumann boundary condition given on the external surface of the system. The sensitivity model is created by the differentiation of energy equations and boundary-initial conditions with respect to the parameter considered (a direct approach, e.g. [13-16]).

2 MATHEMATICAL DESCRIPTION OF THE PROCESS

The domain considered (cross section of the forearm middle part [17]) is shown in Figure 1.

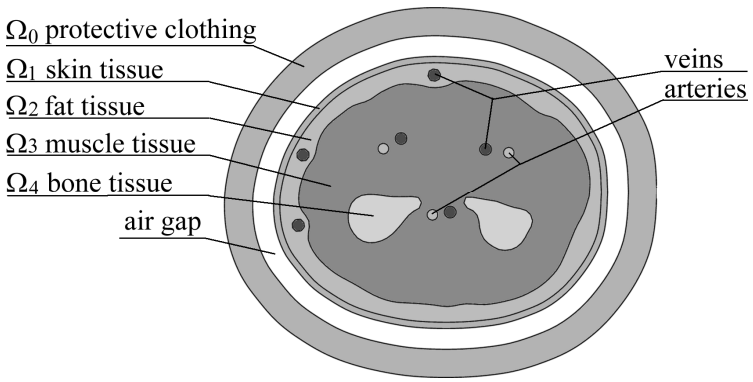


Figure 1: Simplified 2D geometrical model of the forearm with protective clothing cross-section.

Heat transfer processes proceeding in the tissue volume are described by the system of the Pennes equations of the form

$$c_e(T)\frac{\partial T_e(x,t)}{\partial t} = \nabla\left[\lambda_e(T)\nabla T_e(x,t)\right] + Q_{per\,e}(x,t) + Q_{met\,e}(x,t), \qquad e = 1,...,4 \tag{1}$$

where e = 1, 2, 3, 4 refers to the tissue sub-domains (skin, fat, muscle and bone, respectively), c_e is the volumetric specific heat, λ_e is the thermal conductivity, Q_{per} and Q_{met} [W/m^3] are the capacities of volumetric internal heat sources connected with the blood perfusion and metabolism, T, $x = \{x_1, x_2\}$, t denote the temperature, spatial co-ordinates and time. The perfusion heat source is given by the formula

$$Q_{per\,e}(x,t) = c_b G_{be}(T)\left[T_b - T_e(x,t)\right], \qquad T_b = \left(T_{b\,artery} + T_{b\,vein}\right)/2 \tag{2}$$

where G_{be} is the blood perfusion [m^3$_{blood}$/(s m^3$_{tissue}$)], c_b is the blood volumetric specific heat and $T_{b\,artery}$ and $T_{b\,vein}$ are the arterial and vein blood temperatures. Metabolic heat source Q_{met} (x, t) can be treated as a constant value.

Equation describing the transient temperature field in the domain of fabric is of the form

$$c_0(T)\frac{\partial T_0(x,t)}{\partial t} = \nabla\left[\lambda_0(T)\nabla T_0(x,t)\right] \tag{3}$$

The similar equation with the parameters corresponding to the properties of air describes the temperature field in the domain of trapped air between fabric and forearm.

So, the layer of trapped air is treated as a solid body (a convection in this sub-domain can be neglected). Through the air gap heat flows by the radiation and conduction, in particular

$$q_r(x,t) = \alpha_r\left[T_F(x,t) - T_T(x,t)\right], \qquad q_c(x,t) = \frac{\lambda_a}{\delta}\left[T_F(x,t) - T_T(x,t)\right] \tag{4}$$

where α_r is the radial heat transfer coefficient, λ_a is the air thermal conductivity, δ is the air gap thickness, T_F, T_T are the averaged temporary temperatures of fabric and skin surfaces. As is well known

$$\alpha_r = 10^{-4}\varepsilon_{F-T}C_c(T_F + T_T + 546)\left[\left(\frac{T_F + 273}{100}\right)^2 + \left(\frac{T_T + 273}{100}\right)^2\right] \tag{5}$$

and

$$\frac{1}{\varepsilon_{F-T}} = \frac{1}{\varepsilon_F} + \frac{1}{\varepsilon_T} - 1 \tag{6}$$

is the substitute emissivity, while $\varepsilon_F, \varepsilon_T$ are the emissivities of fabric and skin tissue surfaces and C_c = 5.67 W/m^{2}K^2.

The total heat flux exchanged between the surfaces is equal to

$$q(x,t) = \alpha_r\left[T_F(x,t) - T_T(x,t)\right] + \frac{\lambda_a}{\delta}\left[T_F(x,t) - T_T(x,t)\right] = \left(\alpha_r + \frac{\lambda_a}{\delta}\right)\left[T_F(x,t) - T_T(x,t)\right] \tag{7}$$

Let us introduce the substitute air thermal conductivity

$$q(x,t) = \frac{\lambda_{sub}}{\delta}\left[T_F(x,t) - T_T(x,t)\right] = \left(\alpha_r + \frac{\lambda_a}{\delta}\right)\left[T_F(x,t) - T_T(x,t)\right] \tag{8}$$

and finally

$$\lambda_{sub} = \delta(\alpha_r + \frac{\lambda_a}{\delta}) \tag{9}$$

It should be pointed out, that the volumetric specific heat of trapped air corresponds to the temperature-dependent values which can be found in literature (e.g. [18]).

On the contact surface between the tissue sub-domains, the continuity of temperature and heat fluxes are assumed

$$x \in \Gamma_{e,e+1} : \begin{cases} -\lambda_e \dfrac{\partial T_e(x,t)}{\partial n} = -\lambda_{e+1} \dfrac{\partial T_{e+1}(x,t)}{\partial n} \\ T_e(x,t) = T_{e+1}(x,t) \end{cases}, \qquad e = 1,2,3 \tag{10}$$

where $\partial / \partial n$ is a temperature derivative in normal direction. The same condition is accepted on the surfaces limiting the air gap.

On the outer surface of the fabric, the combined Robin and Neumann boundary conditions are taken into account

$$x \in \Gamma_{out} : \quad -\lambda_0 \frac{\partial T_0(x,t)}{\partial n} = -\alpha_{out}\left(T_{amb} - T_0(x,t)\right) + \begin{cases} -aq_b, & \text{if } t \in (0, t_{heating}] \\ 0, & \text{if } t \in (t_{heating}, t_{final}] \end{cases} \tag{11}$$

where q_b is the boundary heat flux, a is the absorptivity of the outer surface of the protective clothing, α_{out} is the heat transfer coefficient and T_{amb} is the ambient temperature. On the surfaces between the blood vessels and soft tissue sub-domains the Robin condition is taken into account, at the same time the heat transfer coefficients and the arterial and vein blood temperatures are assumed to be known.

The initial conditions are also given

$$t = 0 : \quad T_e(x,t) = T_{steady}(x), \qquad e = 0,1,...,4 \tag{12}$$

where T_{steady} is the temperature distribution corresponding to the steady state conditions in the tissue-fabric domain for the given initial ambient temperature and the initial external heat transfer coefficient.

3 SENSITIVITY MODEL

The sensitivity model presented in this paper concerns the changes of transient temperature field in domain considered due to the perturbations of the boundary heat flux. The equations creating the sensitivity model have been obtained by the differentiation of the basic equations and conditions with respect to q_b.

Let us denote the sensitivity function $\partial T(x,t) / \partial q_b$ by $U(x, t)$. Then the sensitivity equations for the tissue sub-domains are of the form

$$c_e \frac{\partial U_e(x,t)}{\partial t} = \lambda_e \nabla^2 U_e(x,t) - c_b G_{be} U_e(x,t), \qquad e = 1,...,4 \tag{13}$$

while for the fabric

$$c_0 \frac{\partial U_0(x,t)}{\partial t} = \lambda_0 \nabla^2 U_0(x,t) \tag{14}$$

One can see that the above equations concern the constant values of fabric and tissue thermophysical parameters (the authors do not dispose the convincing information connected with the temperature-dependent relations).

For the trapped air the sensitivity equation is the following

$$c_a \frac{\partial U_a(x,t)}{\partial t} + \frac{\mathrm{d}c_a}{\mathrm{d}T} U_a(x,t) \frac{\partial T_a(x,t)}{\partial t} = \nabla\left[\lambda_a \nabla U_a(x,t)\right] + \frac{\mathrm{d}\lambda_a}{\mathrm{d}T} \nabla\left[U_a(x,t) \nabla T_a(x,t)\right] \quad (15)$$

On the external surface the boundary condition has a form

$$x \in \Gamma_{out}: \quad -\lambda_0 \frac{\partial U_0(x,t)}{\partial n} = \alpha_{out} U_0(x,t) + \begin{cases} -a, & \text{if } t \in (0, t_{heating}] \\ 0, & \text{if } t \in (t_{heating}, t_{final}] \end{cases} \quad (16)$$

'Rebuilding' of the others boundary and initial conditions is rather simple and it will not be discussed here.

4 RESULTS OF COMPUTATIONS

In Figure 2 an example of the control volume mesh covering the sub-domains considered is shown. Here, the domain being a composition of the forearm, protective clothing and air gap is divided into 3429 control volumes.

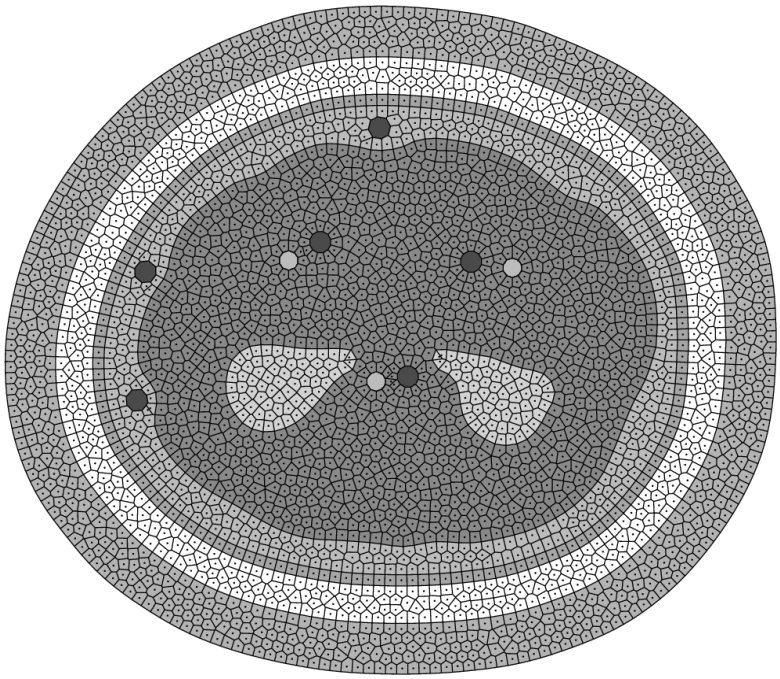


Figure 2: Control volume mesh of the considered sub-domains for the cross-section of the forearm and protective clothing.

The initial temperature distribution of the fabric-forearm system (see: Figure 3) corresponds to the steady state conditions. They are found for the air temperature equals T_{amb} = 20 °C, and heat transfer coefficient α_{out} = 5 W/m^{2}K. Additionally the following blood temperatures are assumed: $T_{b\,artery}$ = 36 °C, $T_{b\,vein}$ = 35 °C, while $\alpha_{artery} = \alpha_{vein}$ = 5000 W/m^{2}K.

In the numerical simulation at the moment t = 0 the process of heating starts (q_b = 2100 W/m^2, a = 0.3, while T_{amb} = 30 °C and α_{out} = 5 W/m^{2}K) and this process is continued

until $t_{heating}$ = 3 min. Next, the cooling process takes place (q_b = 0 W/m^2, T_{amb} = 30 ^{o}C and α_{out} = 5 W/m^{2}K).

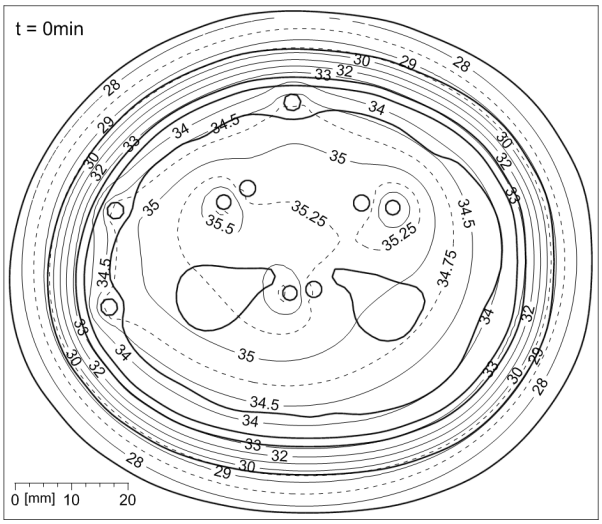


Figure 3: The initial temperature distribution.

The basic solution in the form of heating curves representing the average temperature in the layers of skin, the inner and outer layers of protective clothing is shown in Figure 4. The next Figure presents the temporary solutions for times 3 min. and 6 min.

In Figures 6 and 7 the numerical solution of sensitivity problem is shown. In particular, Figure 6 illustrates the changes of sensitivity function $U_e(x,t) = \partial T_e(x,t) / \partial q_b$ (also representing the average values in the skin layer, inner and outer layers of protective clothing). The temporary distributions of sensitivity with respect to perturbation of the boundary heat flux are shown in Figure 7.

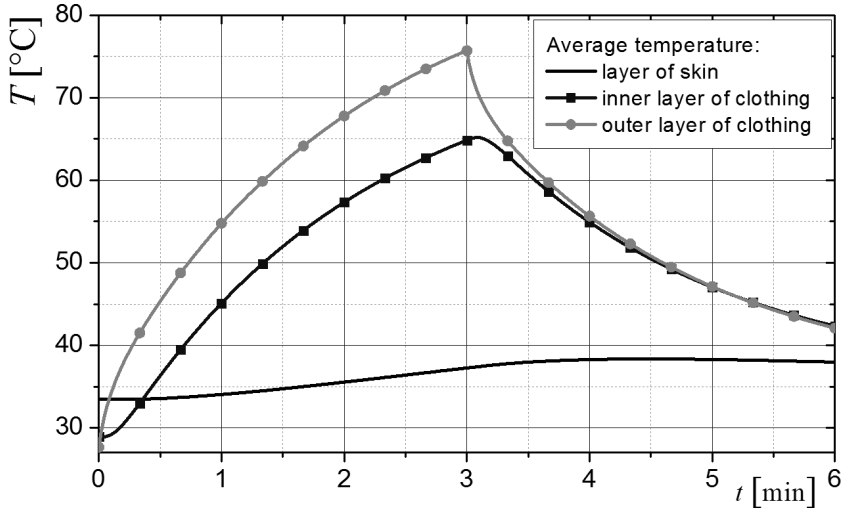


Figure 4: Heating curves representing the average temperature in the skin layer, inner and outer layers of protective clothing.

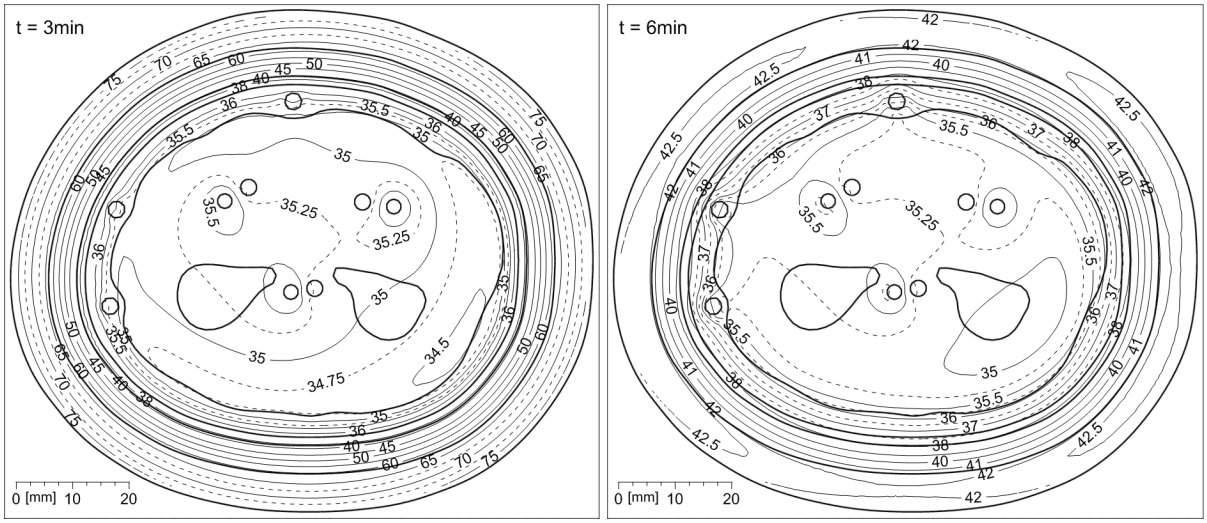


Figure 5: Temporary solutions for 3 and 6 min.

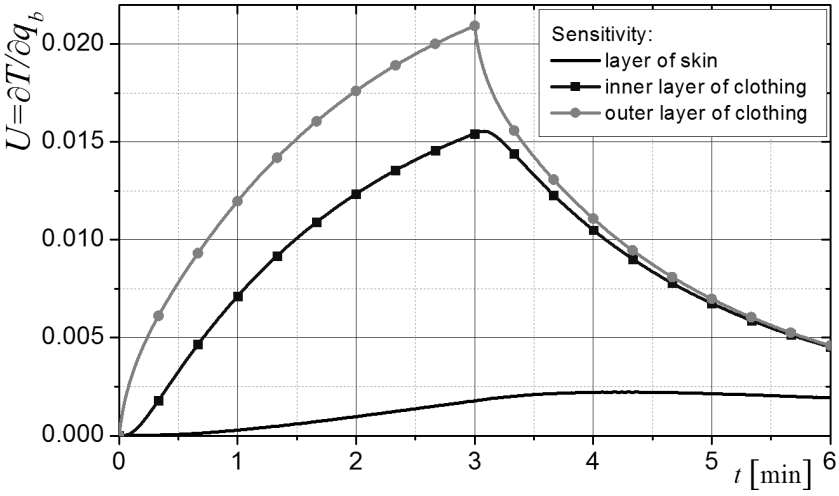


Figure 6: Sensitivity with respect to the boundary heat flux.

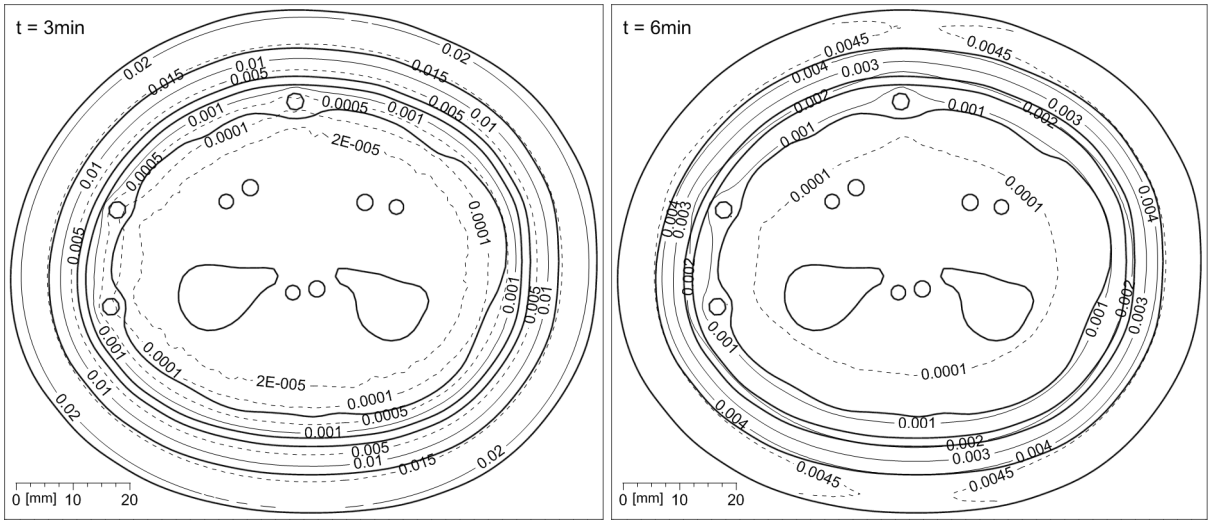


Figure 7: Distribution of sensitivity with respect to the boundary heat flux for times 3 and 6 min.

5 CONCLUSIONS

In this paper the numerical model of thermal processes occurring in the tissue-fabric domain exposed to an external heating by the strong boundary heat flux have been discussed. The computations have been performed on the basis of the author's variant of CVM using the Voronoi polygons. Such a method can be easily adapted to the needs of bio-heat transfer problems and it is an effective approach to the numerical modeling of thermal processes proceeding in the domain of biological tissue. Moreover, a discretization of 2D domain of the complex heterogeneous shape using the Voronoi polygons has many advantages, among others, the shapes of sub-domains can be exactly reconstructed. The CVM assures the correct modeling of the energy balances and the implementation of the different types of boundary conditions is simple.

The analysis of the results presented in this paper, gives a number of the significant information. The increase of temperature in the fabric domain proceeds essentially faster in comparison with the tissue domain. In the case of accidental contact between hot fabric and human body the tissue burns can take place. One can note that at the final moment of the heating process, the temperature of the protective clothing is significantly higher than the temperature of the skin layer and at the stage of the cooling of outer fabric surface, the heat is still transferred to the forearm domain. In order to achieve the thermal comfort of a person during the heating process (the maximum temperature of the skin tissue layer should not exceed the value more than 38 °C) the adequate period of the heating time in the presented simulation has been chosen. Such an analysis of results allows one to the prediction of a situation which is not preferable by a person. Further analysis allows one to determine the maximum stay time at the given heat flux for a person under secure conditions.

In this work the sensitivity analysis with respect to the perturbations of the boundary heat flux is considered and one can formulate the following conclusions:

- the maximum value reaches at the final stage of heating and the most essential changes are in the outer layer of fabric,
- after period of the heating process, the sensitivity rapidly decreases,
- during the considered simulation time in the tissue domain the sensitivity is practically close to zero.

ACKNOWLEDGEMENT

The paper has been sponsored by WSZOP Katowice (Research Project No. 3/2013).

REFERENCES

[1] H.H. Pennes, Analysis of tissue and arterial blood temperatures in the resting human forearm, *Journal of Applied Physiology*, **1**, 93-122, 1948.

[2] H. Arkin, L.X. Xu, K.R. Holmes, Recent development in modeling heat transfer in blood perfused tissues, *IEEE Trans Bio-Med. Eng.* **41**, 97-107, 1994.

[3] E. Majchrzak, Application of different variants of the BEM in numerical modeling of bio-heat transfer processes, *MCB: Molecular & Cellular Biomechanics* **10, 3**, 201-232, 2013.

[4] E. Majchrzak, B. Mochnacki, M. Dziewonski, M. Jasinski, Numerical modeling of hyperthermia and hypothermia processes, *Advanced Materials Research* **268-270**, 257-262, 2011.

[5] E. Majchrzak, B. Mochnacki, M. Jasinski, Numerical modeling of bioheat transfer in multi-layer skin tissue domain subjected to a flash fire, *Computational Fluid and Solid Mechanics* **1-2**, 1766-1770, 2003.

[6] M. Duda, B. Mochnacki, 3D model of thermal interactions between human forearm and environment, *Journal of Applied Mathematics and Computational Mechanics*, **14, 3**, 12-23, 2015.

[7] B. Mochnacki, A. Piasecka-Belkhayat, Numerical modeling of skin tissue heating using the interval finite difference method, *MCB: Molecular & Cellular Biomechanics*, **10, 3**, 233-244, 2013.

[8] W. Roetzel , N. Putra, S.K. Das, Experiment and analysis for non-Fourier conduction in materials with non-homogeneous inner structure, *Int. J. Therm. Sci*. **42**, 541–552, 2003.

[9] E. Majchrzak, L. Turchan, The general boundary element method for 3D dual-phase lag model of bioheat transfer, *Engineering Analysis with Boundary Elements*, **50**, 76-82, 2015.

[10] D. Fiala, K.J. Lomas, M. Stohrer, A computer model of human thermoregulation for a wide range of environmental conditions: the passive system, *Journal of Applied Physiology* **87, 5**, 1957-1972, 1999.

[11] P. Chitrphiromsri, *Modeling of thermal performance firefighter protecting clothing during the intense heat exposure*, Mechanical Engineering, Doctoral Theses, Raleigh, 2004.

[12] M. Ciesielski, B. Mochnacki, Application of the Control Volume Method using the Voronoi polygons for numerical modeling of bio-heat transfer processes, *Journal of Theoretical and Applied Mechanics*, **52, 4**, 927-935, 2014.

[13] B. Mochnacki, E. Majchrzak, Sensitivity of the skin tissue on the activity of external heat sources, *CMES-Computer Modeling in Engineering &Sciences*, **4, 3-4**, 431-438, 2003.

[14] E. Majchrzak, B. Mochnacki, Sensitivity analysis and inverse problems in bio-heat transfer modelling, *Computer Assisted Mechanics and Engineering Sciences*, **13**, 85-108, 2006.

[15] K. Dems, B. Rousselet, Sensitivity analysis for transient heat conduction in a solid body – Part I, *Structural Optimization* **17**, 36-45, 1999.

[16] M. Kleiber, *Parameter Sensitivity in Non-linear Mechanics*, J. Willey & Sons, London, 1997.

[17] M. Schuenke, E. Schulte et al., *General Anatomy and Musculosceleton Systerm (Atlas of Anatomy)*, Thieme, Stuttgart – New York, 2010.

[18] J. Szargut, *Thermodynamics*, PWN, Warsaw. 2000 (in Polish).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

INFLUENCE OF MUSCLE COMPRESSION ON DYNAMIC MUSCLE PERFORMANCE

Tobias Siebert[1], Norman Stutzig[1], Olaf Till[2] and Christian Rode[2]

[1] University of Stuttgart, Sport and Motion Science
Allmandring 28, 70569 Stuttgart, Germany
e-mail: {tobias.siebert, norman.stutzig}@inspo.uni-stuttgart.de

[2] Friedrich-Schiller-University Jena, Motion Science
Seidelstr. 20, 07749 Jena, Germany
christian.rode@uni-jena.de

Keywords: transversal muscle loading, viscoelastic material properties, impact load, isometric contraction, Rattus norvegicus.

Abstract. *During daily activities such as sitting, carrying loads or wearing orthoses, skeletal muscles are compressed. Surrounding muscles and bones can also induce force leading to compression of the muscle. The aim of the present study was to examine the influence of varying unidirectional transversal muscle loading on contraction dynamics. We performed isometric experiments on isolated rat M. gastrocnemius medialis* (n = 10) *without and with five different transversal loads (0.64, 1.13, 1.62, 2.11, 2.60 N). The muscles were loaded by a custom-made plunger in transversal direction. The muscle force was measured at the distal tendon and the vertical movement of the plunger was captured with a high-speed camera during supramaximal muscle stimulation. Increasing transversal muscle force resulted in an almost linear decrease in maximum isometric muscle force (F_{im}) from 5.0 ± 1.5 % F_{im} to 13.1 ± 2.1 % F_{im}. Compared with an unloaded isometric contraction, the rate of force development (RFD) decreased from 20.7 ± 4.1 % RFD to 35.2 ± 5.8 % RFD. The lifting height of the plunger decreased from 1.7 ± 0.2 mm at the lowest transversal load to 0.6 ± 0.2 mm at the maximum load applied. Unidirectional transversal compression depresses longitudinal force development in skeletal muscles and, thus, should be considered in musculo-skeletal models simulating the interaction of muscles with their mechanical environment, e.g. in impact biomechanics. The interactions between the muscle and one unidirectional transversal load could be explained reasonably well with a simple model including a hill-type muscle model, a geometric model transferring transversal forces into a longitudinal direction and a viscoelastic model representing the characteristics of passive muscle tissue. Data presented in this study may be important to develop and validate muscle models which enable simulation of muscle contractions under compression and reveal the mechanisms behind.*

INTRODUCTION

It is well known that active muscle force depends on different parameters such as muscle length [1,2] or contraction velocity [3]. Traditionally these dependencies were determined on isolated muscle preparations. However, *in vivo* skeletal muscles are embedded in an environment of other muscles, connective tissue and bones. These surrounding structures are supposed to exert external forces to the muscle belly, thereby influencing active muscle force. External tensile forces can be transmitted via extramuscular connective tissue to the muscle [4]. These forces act mainly in longitudinal direction (in the line between muscle origin and insertion: line of action). In addition, forces may act in transversal direction (perpendicular to the line of action) thereby compressing the muscle [5]. Transversal muscle loading may be induced during sports exercises, e.g. by lifting belts, bench shirts and compressing swimsuits or occurs during daily activities, such as carrying loads, wearing orthoses, or sitting.

In a recent study [5] the influence of increasing transversal muscle loading on contraction dynamics was examined using rat *M. gastrocnemius medialis* (GM). The muscle was loaded unidirectional by a plunger which was able to move in transversal direction (Fig. 1). During muscle contraction the muscle deforms, and work was performed to lift the load. Compared with the unloaded contraction, transversal muscle loading with 0.64 N resulted in a decrease in maximal isometric muscle force (F_{im}) and rate of force development (RFD) by 5% and 25%, respectively. Due to given contact area of the plunger (0.5 cm^2) the applied pressure was 1.3 N/cm^2, which corresponds to about the mean pressure at the body-seat interface during human sitting [6,7]. Comparatively higher muscle compressions are, however, expected for e.g. car drivers during accidents or athletes involved in contact sports (e.g. boxing and wrestling). Thus, the aim of the present study was to examine the influence of varying unidirectional transversal muscle loading on contraction dynamics of isolated rat *M. gastrocnemius medialis*. Parts of this study are published in [8].

METHODS

Experimental procedure was similar to [5]. Experiments were approved according to Section 8 of the German animal protection law (Tierschutzgesetz, BGBl I 1972, 1277). Experiments were performed on rat (*Rattus norvegicus*, Wistar) GM (n = 10, mass GM = 894 ± 88 mg). Muscle preparation, anaesthesia and setup have been described in detail earlier [9]. Briefly, the rats were anaesthetized with sodium pentobarbital (100mg/kg body mass). Body temperature was maintained at 33-36°C using a heating pad (Föhr Medical Instruments). The muscle was freed from its surrounding tissues and fixed in a horizontal position. The distal tendon was attached to a muscle lever system (Aurora scientific 305B-LR). The sciatic nerve was stimulated (Aurora Scientific 701C) with 100 µs square wave pulses of 3–8 mA (threefold the twitch threshold) at100 Hz (maximal tetanical muscle stimulation).

The transversal load was applied in the middle of the muscle belly by a custom made plunger with 0.5 cm^2 contact area (Fig. 1). To withstand the induced transversal forces, the muscle was supported from below by a horizontal plate. The plunger was able to move freely in vertical direction and its movement was measured with a high-speed camera (Vosskühler, 462 frames/sec). To minimize friction between the plunger and the muscle surface, the attachment area was coated with paraffin oil. Variable suffixes ‘lo’ and ‘unlo’ distinguish loaded and unloaded experiments, respectively.

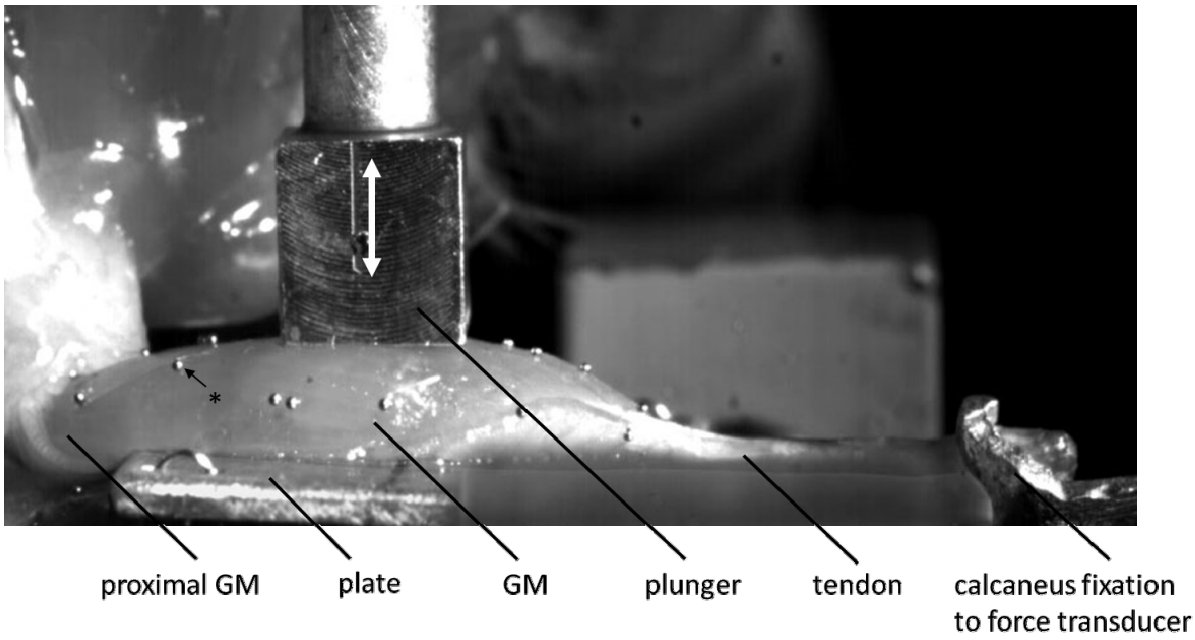


Figure 1: Experimental setup. Transversal load is induced on rat *M. gastrocnemius medialis* (GM) by a plunger which is able to move freely in vertical direction. The small markers on the muscle surface (*) were used for analyses in another study.

An alternating sequence of experiments without a transversal load and with a transversal load was conducted. The muscle was stimulated for 300 ms at optimum muscle length. The resting phase between the experiments was 3 min to minimize muscle fatigue. Five loaded experiments were performed with increasing transversal loads (0.64, 1.13, 1.62, 2.11, and 2.60 N) corresponding to increasing pressures (1.3, 2.3, 3.3, 4.3, and 5.3 N/cm^2). The plunger covered about 20 % of the muscle's surface area visible from the top. After each loaded experiment the muscle was able to reproduce F_{im} in the unloaded reference contraction.

We determined F_{im} as well as *RFD* from each experiment. *RFD* was calculated as force difference between 10–90% F_{im} divided by the corresponding difference in time. ΔF_{im} and ΔRFD indicate percentage changes of F_{im} and *RFD*, respectively, from unloaded to loaded experiments. In the loaded experiments, the lifting height $\Delta h = h - h_0$ was measured, where h_0 is the initial height of the load in the passive condition. The maximum value of Δh over contraction time is the maximum lifting height h_{max}.

For statistical analyses we used a linear model for the dependencies of ΔF_{im}, ΔRFD, and Δh, respectively, on transversal force. The slopes of the dependencies were calculated for each muscle by linear regression, using the Matlab® (The Mathworks, Inc., Nattick, MA, USA) function `regress.m'. To obtain confidence intervals at a confidence level of 0.99 for the slopes, linear regressions were performed over the combined experimental values of all ten muscles.

RESULTS

Unidirectional transversal muscle loading with different loads impacts longitudinal muscle force, rate of force development and lifting height of the load (Fig. 2).

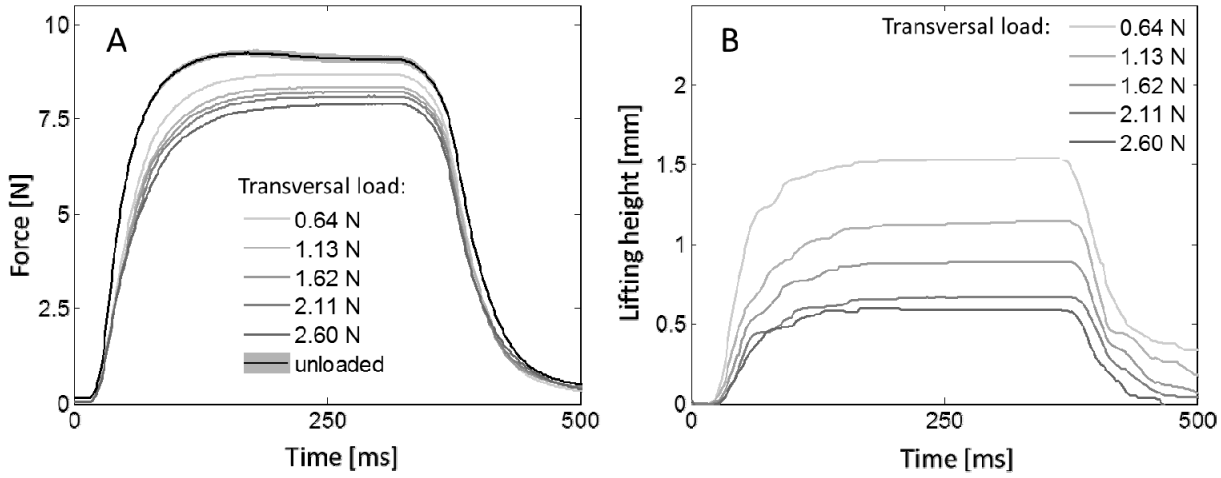


Figure 2: Force-time (A) and lifting height - time traces depending on different transversal loads exemplarily shown for GM9. Mean (black line) and standard deviation (grey) are given for the force-time traces of the unloaded reference contractions. Stimulation started at $t = 0$ and ended at $t = 300$ ms.

Increasing the transversal load from 0.64 N to 2.60 N resulted in consistent and almost linear changes of ΔF_{im} from 5.0 ± 1.5 % F_{im} to 13.1 ± 2.1 % F_{im} ($R^2 > 0.95$; Fig. 3A), of ΔRFD from 20.7 ± 4.1 % RFD to 35.2 ± 5.8 % RFD ($R^2 > 0.84$; Fig. 4), and of Δh from 1.7 ± 0.2 mm to 0.6 ± 0.2 mm ($R^2 > 0.97$; Fig. 3B), respectively. Linear regression analyses indicated a strong influence of load on ΔF_{im}, ΔRFD, and Δh (Fig. 3, Table 1). The upper as well as the lower bounds of the .99 confidence intervals of the slopes were all negative, indicating a strong effect of load on the considered variables (Table 1). The calculated confidence interval for the slope of ΔF_{im} implicates a decrease of longitudinal muscle force by at least 5.9% over the range of examined load values.

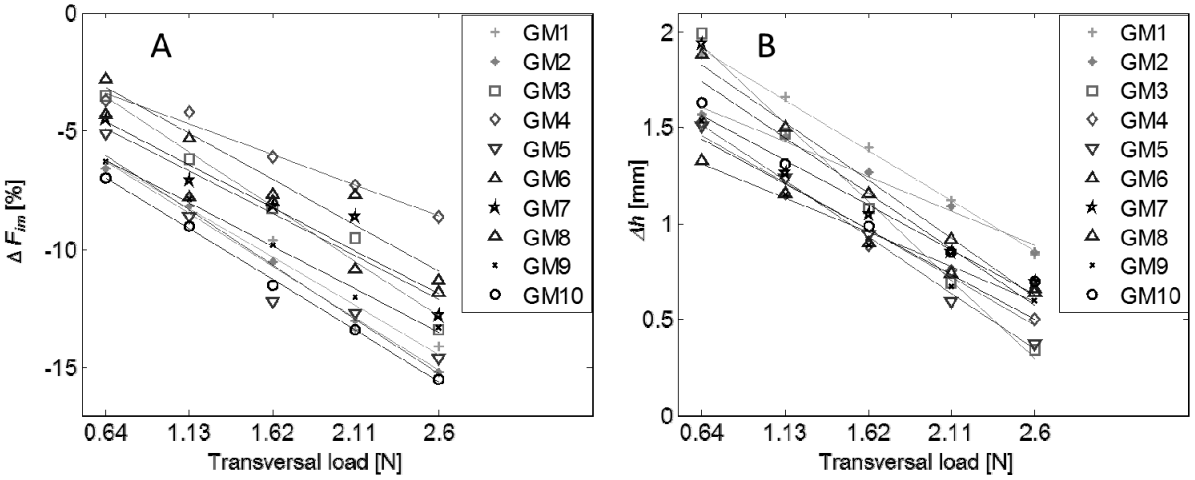


Figure 3: Influence of transversal loading with different loads (0.64, 1.13, 1.62, 2.11, 2.60 N) on isometric muscle force (A) and lifting height (B). Data and individual regression lines are shown for all ten muscles.

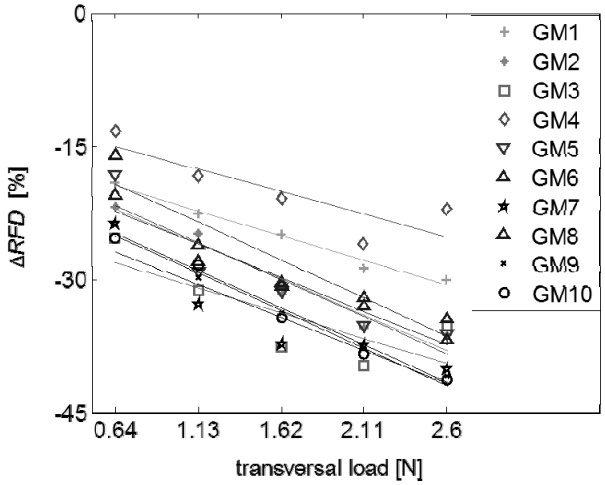


Figure 4: Influence of transversal loading with different loads (0.64, 1.13, 1.62, 2.11, 2.60 N) on rate of force development (*ΔRFD*). Data and individual regression lines are shown for all ten muscles.

independent variable	dependent variable	range of calculated slopes	.99 confidence interval of slope	
			lower bound	upper bound
load [N]	ΔF_{im} [%]	-4.714 to -2.633	-5.026	-3.000
	ΔRFD [%]	-8.735 to -5.143	-10.089	-4.764
	Δh [mm]	-0.833 to -0.367	-0.638	-0.437

Table 1: Slopes of linear regressions and the corresponding 99 % confidence intervals for the dependencies of change in isometric longitudinal muscle force ΔF_{im}, change of rate of force development ΔRFD, and change of lifting height Δh on transversal load. Confidence intervals were calculated over the combined experimental values of all muscles.

DISKUSSION

Varying unidirectional transversal force influences longitudinal contraction dynamics (Fig. 2). Increasing transversal load resulted in an almost linear decrease in maximum isometric force and rate of force development. In our experiments, muscles produce work to strain the series elastic component, to lift the load, and to deform the muscle itself during an end-held isometric contraction. After performing this work, the muscle has to balance the transversal load in the static case. These interactions between the muscle and its mechanical environment could be explained reasonably well with a simple phenomenological model [5] including a hill-type muscle model, a viscoelastic model representing the characteristics of passive muscle tissue and a geometric model transferring transversal forces into a longitudinal direction. The gearing ratio of the geometric model can be explained by a simple ellipsoid muscle model [10].

Various factors may contribute to the observed reduction in muscle force induced by transversal muscle loading. For example, considering a spatial arrangement of muscle fibers transversal muscle loading may induce changes in muscle thickness and fiber pennation angle influencing muscle force generation [11]. Moreover, muscle fibers are connected to the

extracellular matrix [12]. Local deformations induced by the plunger may induce local length changes along muscle fiber and thus decrease muscle force.

As most muscles are packed in 'muscle packages' [13] transversal forces induced by surrounding tissues may impact on large parts of the muscle surface. This may be especially true for inner muscles like rat soleus and plantaris which are almost completely surrounded by other calf muscles. If surrounding muscles contract, they will deform [14,15] and thus may transfer transversal forces over the complete surface to these inner muscles. Thus, the given plunger contact area used in this study may represent rather a local loading compared with the physiological loading situation in muscle packages. Thus, it seems of great importance to examine the intermuscular pressure (between neighboring muscles) in muscle packages during contractions. Based on these insights, basic questions about evolution of muscle architecture can be addressed and it may be speculated whether muscle architectures are designed to minimize mutual transversal loading in between neighboring muscles. Maybe muscle architectures are adjusted such that muscles exert only small intermuscular pressures during contraction generating. This might be one explanation for complex muscle architectures, exhibiting e.g. distributions of varying fiber lengths and pennation angles or inner tendon sheets [16,17,18,19]. If, however, minimization of intermuscular pressure is no design criteria, transversal forces induced by contracting muscles may have a function. For example, they could contribute to stabilization of muscle packages in impact situations (e.g. foot contact during cyclic locomotion). Moreover, transversally compressed muscles can act under dynamically favourable conditions, e.g. transversal loading can result in increased work production of the contractile component [5]. Furthermore, work performed transversally on adjacent muscles may be saved and released subsequently in addition to energy savings in series [20,21] and parallel elastic structures [22] which might enhance efficiency of cyclical locomotion. However, these questions have to be addressed in future studies.

Summing up, experimental observation of the impact of transversal muscle loading on contraction dynamics may help to better understand muscle tissue properties and to describe more precisely compression of the human body e.g. in rehabilitation studies. Moreover, applying transversal loads to muscles opens a window to analyze three-dimensional muscle force generation. The data presented in this study can be used to advance and validate muscle models which enable simulation of muscle contractions under compression and possibly reveal the mechanisms behind.

REFERENCES

[1] A.M. Gordon, A.F. Huxley, F.J. Julian, The variation in isometric tension with sarcomere length in vertebrate muscle fibres. *J Physiol* 184, 170-92, 1966.

[2] C. Rode, T. Siebert, A. Tomalka, R. Blickhan, *Proc R Soc B*, Myosin filament sliding through the Z-disc relates striated muscle fibre structure to function. http://dx.doi.org/10.1098/rspb.2015.3030, 2016.

[3] A.V. Hill, The Heat of Shortening and the Dynamic Constants of Muscle. *Proceedings of the Royal Society of London. Series B*, *Biological Sciences* 126, 136-195, 1938.

[4] H. Maas, G.C. Baan, P.A. Huijing, Intermuscular interaction via myofascial force transmission: effects of tibialis anterior and extensor hallucis longus length on force transmission from rat extensor digitorum longus muscle. *J Biomech* 34, 927-40, 2001.

[5] T. Siebert, O. Till, R. Blickhan, Work partitioning of transversally loaded muscle: experimentation and simulation. *Comput Methods Biomech Biomed Engin* 17, 217-29, 2014.

[6] A. Gefen, N. Gefen, E. Linder-Ganz, S.S. Margulies, In vivo muscle stiffening under bone compression promotes deep pressure sores. *J Biomech Eng* 127, 512-24, 2005.

[7] E. Linder-Ganz, N. Shabshin, Y. Itzchak, A. Gefen, Assessment of mechanical conditions in sub-dermal tissues during sitting: a combined experimental-MRI and finite element approach. *J Biomech* 40, 1443-54, 2007.

[8] T. Siebert, O. Till, N. Stutzig, M. Gunther, R. Blickhan, Muscle force depends on the amount of transversal muscle loading. *J Biomech* 47, 1822-1828, 2014.

[9] O. Till, T. Siebert, C. Rode, R. Blickhan, Characterization of isovelocity extension of activated muscle: a Hill-type model for eccentric contractions and a method for parameter determination. *J Theor Biol* 255, 176-87, 2008.

[10] T. Siebert, M. Gunther, R. Blickhan, A 3D-geometric model for the deformation of a transversally loaded muscle. *J Theor Biol* 298, 116-121, 2012.

[11] J.M. Wakeling, M. Jackman, A.I. Namburete, The effect of external compression on the mechanics of muscle contraction. *J Appl Biomech* 29, 360-4, 2013.

[12] C.A. Yucesoy, Epimuscular myofascial force transmission implies novel principles for muscular mechanics. Exerc Sport Sci Rev 38, 128-34, 2010.

[13] T. Siebert, K. Leichsenring, C. Rode, C. Wick, N. Stutzig, H. Schubert, R. Blickhan, M. Böl, Three-Dimensional Muscle Architecture and Comprehensive Dynamic Properties of Rabbit Gastrocnemius, Plantaris and Soleus: Input for Simulation Studies. *PLoS ONE 10*, e0130985, 2015.

[14] M. Böl, K. Leichsenring, C. Weichert, M. Sturmat, P. Schenk, R. Blickhan, T. Siebert, Three-dimensional surface geometries of the rabbit soleus muscle during contraction: input for biomechanical modelling and its validation. *Biomech Model Mechanobiol* 12, 1205-1220, 2013.

[15] M. Böl, K. Leichsenring, M. Ernst, C. Wick, R. Blickhan, T. Siebert, Novel microstructural findings in M. plantaris and their impact during active and passive loading at the macro level. *J Mech Behav Biomed Mater* 51, 25-39, 2015.

[16] S. Gorb, M.S. Fischer, 3D-analysis of the arrangement and length distribution of fascicles in the triceps muscle of the Galea musteloides (Rodentia, Cavimorpha). *Zoomorphology* 120, 91-97, 2000

[17] P. Schenk, T. Siebert, P. Hiepe, D. Gullmar, J.R. Reichenbach, C. Wick, R. Blickhan, M. Böl, Determination of three-dimensional muscle architectures: validation of the DTI-based fiber tractography method by manual digitization. *J Anat* 223, 61-8, 2013.

[18] H. Stark, R. Frober, N. Schilling, Intramuscular architecture of the autochthonous back muscles in humans. *J Anat* 222, 214-22, 2013.

[19] P. Hiepe, K.H. Herrmann, D. Gullmar, C. Ros, T. Siebert, R. Blickhan, K. Hahn, J.R. Reichenbach, Fast low-angle shot diffusion tensor imaging with stimulated echo encoding in the muscle of rabbit shank. *NMR Biomed* 27, 146-57; 2014.

[20] A. Biewener, R. Baudinette, In vivo muscle force and elastic energy storage during steady-speed hopping of tammar wallabies (Macropus eugenii) *J Exp Biol* 198, 1829-41, 1995.

[21] F. Mörl, T. Siebert, D. Haufle, Contraction dynamics and function of the muscle-tendon complex depend on the muscle fibre-tendon length ratio: a simulation study. *Biomech Model Mechanobiol*, doi: 10.1007/s10237-015-0688-7, 2015.

[22] C. Rode, T. Siebert, W. Herzog, R. Blickhan, The effects of parallel and series elastic components on the active cat soleus force-length relationship. *Journal of Mechanics in Medicine and Biology* 9, 105-122, 2009.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A 3D ELECTROMECHANICAL FEM-BASED MODEL FOR CARDIAC TISSUE

Minh Tuan Duong[1,2], **Alexander Jung**[1], **Ralf Frotscher**[1] **and Manfred Staat**[1]

[1] Aachen University of Applied Sciences
Institute of Bioengineering
Heinrich-Mußmann-Str. 1, 52428 Jülich, Germany
e-mail: {a.jung, frotscher, m.staat}@fh-aachen.de

[2] Hanoi University of Science and Technology
DaiCoViet No.1, Hanoi, Vietnam
tuan.duongminh@hust.edu.vn

Keywords: Cardiac Tissue, CellDrum, Cell Models, Drug Modeling, Electromechanical Coupling

Abstract. *The CellDrum provides an experimental setup to investigate the electromechanics of a cardiac tissue construct and particularly the effect of drugs. Experimental data can be used to parametrize and validate computational electromechanical models. Until now, the experiments have been performed with a thin tissue layer of a mixture including ventricular, atrial and sinoatrial cells. Mechanically, it is modeled as a materially and geometrically nonlinear shell. For future experiments with a thick tissue layer of mixtures and single cardiac cell types the mechanical model is extended to a nonlinear 3D continuum which can be used as a step towards whole heart modeling. Comparisons with the 2D shell model are presented with and without the consideration of a chosen drug and a simulation with a model for thick cardiac tissue construct is carried out to give a first impression on how the experiments might look like.*

1 INTRODUCTION

A human heart is divided into four chambers, two atria and two ventricles, connected by four valves. The right atrium receives oxygen-poor blood from the body and pumps it to the right ventricle from where it is pumped to the lungs. The left atrium receives oxygen-rich blood from the lungs and pumps it to the left ventricle from where it is pumped back to the body. Both the filling and ejection processes are controlled by electrophysiological processes. At the cellular level the main components in the heart are cardiomyocytes, cardiac fibroblasts, endothelial cells and vascular smooth muscle cells [1]. Cardiomyocytes are subdivided into two types, namely the pacemaker-conduction and the contractile cardiomyocytes. Each cardiac cycle is initiated by sinoatrial cells located in the upper wall of the right atrium. Sinoatrial cells are pacemaker cardiomyocytes that can generate spontaneous action potentials, i.e. a rapid change in the membrane potential. Over a conduction system the action potential propagates throughout the whole heart so that the contractile cardiomyocytes in the atria and ventricles are electrically excited. Cardiomyocytes are composed of myofilaments enabling the cells to contract. Even pacemaker-conduction cardiomyocytes are able to contract but by far not as strong as the contractile cardiomyocytes. In a process called excitation-contraction coupling the action potential results in a release of Ca^{2+} from intracellular stores which activate the sarcomeres so that tension can be created and the cells can contract.

Computational cardiac models have been developed with the motivation to investigate the physiology and pathologies of the heart and ultimately to optimize therapies. In 1952, Hodgkin and Huxley published the first electrophysiological model of an excitable cell [2]. For a squid giant axon they linked the kinetics of ion channel conformation change with ion currents through the cell membrane and the change of the membrane potential. Ten years later, this groundbreaking work was extended by Noble to model the electrophysiology of cardiac myocytes [3]. Since then, the understanding of cellular mechanisms and the experimental potential has increased. This has led to the development and improvement of models for different human cardiomyocytes [4-7]. Since the models are already at a mature stage, they can serve as a basis to investigate the effect of various drugs on the cardiac electrophysiology [7, 8]. When these models are applied to tissues the action potential propagation from one cell to its neighboring cells needs to be taken into account. For this, the monodomain and bidomain reaction-diffusion equations are used. Bidomain equations account for the different electrical conductivities of the intracellular and extracellular spaces whereas only one conductivity is used in the monodomain equations. Differences between monodomain and bidomain results were reported to be extremely small [9] so that the monodomain equations with a much lower computational cost is widely preferred. Solving them with the Finite Element Method (FEM) gives the membrane averaged over a set of myocytes in each element.

While many questions can be addressed with electrophysiological models, others e.g. regarding the dilated cardiomyopathy necessitates a complete representation of the electromechanics. Improved computer performance has made it possible to develop FEM models where the electrophysiology is coupled with the mechanics. Electromechanical coupling is based on the relationship between the electrical activation of the tissue, the respective active tension generated by the sarcomeres and the resulting deformation [10]. Here, the electrophysiology is modeled on the cellular level [11-15] or is simplified using eikonal equations [16] or the Fitz-Hugh-Nagumo model [17,18]. There exist a wide range of electromechanical models from the cell [10,12] over the tissue [14] to the organ level including left ventricular models [11,15], models of both ventricles [12-15,17], and whole heart models [18]. They feature a varying degree of complexity with respect to the anatomical representation, associated mesh fineness, boundary conditions, material models, underlying electrophysiological models and coupling

models depending on the research objective. Different from the electrophysiological models electromechanical models are still in a maturing phase. Magnetic resonance imaging makes it now possible to create patient-specific highly resolved anatomical models [13], ventricle muscle fiber directions can be determined and assigned with a rule-based algorithm [19] and the orthotropic electric [20] and mechanical properties [21, 22] of the ventricular myocardium are well studied. However, there are still many tasks towards a highly sophisticated whole heart computer model. Examples include the model parametrization, verification, and validation procedures [23].

For thin cardiac tissue constructs these tasks have been addressed with a device called *CellDrum* [24, 25]**.** The *CellDrum* is a circular, 4 μm thin silicone membrane with a diameter of 16 mm. Human-induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CM) are seeded and cultivated on top of the silicon membrane which is coated with fibronectin. The tissue monolayer has a thickness of up to 19 μm and consists of the following mixture: 60% myocardial ventricular, 35% atrial, and 5% nodal hiPSC-CM (Cor.4U, Axiogenesis AG, Germany). The cardiomyocytes live within a culture medium and beat autonomously as in the real heart. Clamped in a fixed ring the composite material can be inflated by a syringe pump and the resulting cellular contraction-dependent deflection is measured using a laser sensor. Thus, parameters for passive and coupling models can be determined [26]. Drugs have already been applied to the culture medium leading to phenomenologically observable changes in the beating frequency, deflection, contraction duration, activation time, relaxation time and resting time [25]. Frotscher et al. [14, 26, 27] developed an electromechanical model of the *CellDrum* which they parametrized, verified and validated using respective experiments. On the level of the electrophysiological cell models also the drug action was implemented. Until now, the effect of four drugs could be successfully simulated. Currently, the thin tissue construct is modeled as a materially and geometrically nonlinear shell whereas electrically it is viewed as a nonlinear three- dimensional continuum.

The next step is to perform experiments with thick cardiac tissue. For this, the mechanical model is modified towards a 3D continuum as well. Comparisons between the 2D shell formulation and 3D continuum formulation are presented. Furthermore, a new geometry is introduced to roughly predict the results of future experiments with thick cardiac tissue. Simulation results with and without the consideration of a chosen drug are shown.

2 ELECTROMECHANICAL MODEL

2.1 Electrophysiological model

Hodgkin-Huxley based cell models for sinoatrial [5], atrial [6] and ventricular [4] cardiomyocytes were employed, each having the general form

$$\frac{\partial V_m}{\partial t} = \frac{1}{C_m} I_{tot} \tag{1}$$

$$I_{tot} = I_{stim} - \sum_{i=1}^{n} I_i(g_j) \quad (j = 1, \dots, m) \tag{2}$$

$$\frac{\partial g_j}{\partial t} = \alpha_j^+(V_m)\left(1 - g_j\right) + \alpha_j^-(V_m) g_j, \tag{3}$$

where V_m is the cell membrane potential, C_m is the electrical capacitance of the cell membrane, I_{tot} is the total current for each cell model, I_{stim} is an electrical stimulus which is zero in the sinoatrial model, I_i are ionic currents related to the ion i and n is the number of ionic currents in the respective cell model. Gate variables g_j, usually one or two, control the opening and closure of each ion channels. Here, α_j^+ and α_j^- are experimentally determined opening and closure rates which depend on a threshold value of the membrane potential for each ion channel. Channel currents take the form

$$I_i = G_i \prod_{j=1}^{m} g_j (V_m - E_i), \tag{4}$$

where G_i is the maximum conductance of channel i and E_i is the reversal potential for the type of ion which flows through this channel.
Furthermore, the employed cell models provide ordinary differential equations for free intracellular sodium, potassium and calcium concentrations as well as for internal calcium stores.

Cellular interaction on the micro-scale level was took into account with a simple homogenization approach. Once the total current density of the sinoatrial model falls below a threshold $I_{tot} < -0.5\ \mu A \mu F^{-1}$, the atrial and ventricular model receive a stimulus

$$I_{stim} = -\frac{50 \mu A m m^{-3}}{C_m} \cdot \frac{V_{cell}}{A_m} \tag{5}$$

for 2 ms depending further on their ratio of cell volume V_{cell} to cell surface area A_m.
Concerning the respective volume fraction and cell volume to cell surface ratio of every cell model (s: sinoatrial, a: atrial, v: ventricular) the highly nonlinear source current per unit length reads

$$I_{src} = \left(\theta_s \frac{A_m^s}{V_{cell}^s} I_{tot}^s + \theta_a \frac{A_m^a}{V_{cell}^a} I_{tot}^a + \theta_v \frac{A_m^v}{V_{cell}^v} I_{tot}^v \right) (1 - \theta_{ECM}). \tag{6}$$

Here, also the volume fraction of the extracellular matrix θ_{ECM} including fibroblasts are taken into account. With this approach the microscopic cellular action potential are homogenized in order to get a macroscopic action potential that propagate throughout the tissue. The monodomain describing the action potential propagation reads

$$C \frac{\partial V}{\partial t} = \nabla(\mathbf{G}(\mathbf{B}) \nabla V) + I_{src}, \tag{7}$$

with the action potential V, capacitance C, isotropic conductance $\boldsymbol{G}$ and the left Cauchy-Green tensor $\mathbf{B}$.

2.2 Electromechanical coupling

Electromechanical coupling is described by the following two ordinary differential equations for each cell model:

$$\frac{\partial Ca_b}{\partial t} = f_1(Ca_i, Ca_b, T_0) \tag{8}$$

$$\frac{\partial z}{\partial t} = f_2(z, Ca_b, \mathbf{B}). \tag{9}$$

The calcium concentration Ca_b that is bound to troponin C in the myofilaments depends on the freely available calcium concentration Ca_i and on the cellular stress T_0. The strain-dependent state of activation is expressed by the activation variable $z \in (0,1)$. Finally, the cellular stress for each cell model is computed according to Niederer and Smith [10] by

$$T_0 = T_{ref}\big(1 + \beta_0(\lambda - 1)\big)K\frac{z}{z_{max}}, \tag{10}$$

with T_{ref} being the cellular reference stress, β_0 is a scalar determining the length dependence of the tension, λ is the one-dimensional cellular stretch in fiber direction, K reflects the binding velocity and z_{max} is the maximally available activation level. Due to a missing cell type specific parametrization an equal parametrization was used for all cell types. For the whole tissue construct the cellular reference stress was determined to be $T_{ref} = 0.5808$ kPa [26] which replaced the original model parameter [10].

2.3 Drug action

Drug action can be implemented in (4) using the conductance-block formulation where the maximal conductance of the channel i is reduced by the factor

$$g_i = g_{i,control}\left[1 + \left(\frac{D}{IC_{50_j}}\right)^n\right]^{-1}. \tag{11}$$

Here, $g_{i,control}$ is the drug-free maximal conductance, D is the drug concentration the IC_{50} value represents the half-inhibitory value, i.e. the concentration of a given drug that will cause the current flowing through the channel i to be reduced by 50%. It has been shown that drug effects at steady-state concentration can be well represented by this model [28]. In this study, Verapamil is used as an example associated with blocking effects on the hERG-, L-type calcium- and sodium-channels of the ventricular cardiomyocytes. Respective IC_{50} values and effective free therapeutic plasma concentrations (EFTPC) used for D can be found in [8].

2.4 Mechanical model

Derivation of the mechanical properties of cardiac muscle tissue can be performed by employing either the active strain or active stress approach. The active strain approach utilizes a multiplicative split of the deformation gradient into passive $\mathbf{F}_\mathrm{p}$ and active deformation $\mathbf{F}_a$

$$\mathbf{F} = \mathbf{F}_p\mathbf{F}_a, \tag{12}$$

whereas in the active stress approach the total stress in the tissue is additively decomposed into the passive and active stress. Both approaches are considered to lead to comparable solutions [29].

Since the experimental parametrization was based on the level of stresses, the latter approach was used. Based on the work of Hunter et al. [30] the total Cauchy stress reads

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_p + \boldsymbol{\sigma}_a = 2J^{-1}\mathbf{B}\frac{\partial\psi}{\partial\mathbf{B}} - p\mathbf{I} + T(t,\mathbf{B})\mathbf{a}\otimes\mathbf{a}, \tag{13}$$

with the strain energy function ψ, the determinant of the deformation gradient J, the hydrostatic pressure p, the identity tensor $\mathbf{I}$, the active stress T being a scalar, the time t and the vector $\mathbf{a}$ oriented in the direction of the cardiac muscle fiber. Global isotropy of the muscle fibers is given in the thin cardiac tissue construct on the *CellDrum* [26], thus

$$\mathbf{a}\otimes\mathbf{a} = \mathbf{I}. \tag{14}$$

is assumed. Finally, the 2nd Piola-Kirchhoff stress $\mathbf{S}$ is given as

$$\mathbf{S} = \mathbf{S}_p + \mathbf{S}_a = 2\frac{\partial\psi}{\partial\mathbf{C}} - p\mathbf{C}^{-1} + T(t,\mathbf{C})\mathbf{C}^{-1}, \tag{15}$$

in terms of the right Cauchy-Green tensor $\mathbf{C}$. According to the experimentally measured pressure-deflection curves the quasi-incompressible hyperelastic neo-Hookean constitutive law

$$\psi = C_{10}(I_1 - 3) \tag{16}$$

was used where I_1 is the first invariant and $C_{10} = 0.0838284$ [26].

2.5 Finite Element Models

Finite element models were created for 19 µm and 100 µm thick cardiac tissue constructs. Due to symmetry only one quarter of the *CellDrum* was modeled. For a 19 µm thick cardiac tissue construct two models were developed. The first model is composed of a 2D mechanical model and a 3D electrical model. Shell formulation was used for the mechanical model which consists of 120 seven-node triangular elements. The electrical model consists of 7664 ten-node quadratic tetrahedral elements. The second model is completely formulated in 3D and both the mechanical and the electrical part are computed using the same mesh which is described above.

For a 100 µm thick cardiac tissue construct another 3D model was created. Here, the mesh consists of 7639 ten-node quadratic tetrahedral elements.

Firstly, the inflation process is simulated by applying a certain pressure to the clamped mechanical mesh. X Pa is necessary to displace the central point of the 2D 19 µm model to 1.2 mm. This has been experimentally validated [14]. Displacements are then projected onto the 3D electrical mesh. Subsequently, the autonomous cell contractions is simulated using an algorithm to solve the electromechanical problem which is described in [14]. When using the same mesh for the mechanical and the electrical part, the displacements do no longer need to be projected. The open source code *Code_Aster* was used to solve the nonlinear mechanical and electrical problem, whereas the cellular ordinary differential equations were solved internally using a fourth-order singly-diagonally implicit Runge-Kutta method.

3 RESULTS

3.1 Comparison between 2D and 3D mechanical model

Figure 1 shows the central node deflection of two 19 µm thick cardiac tissue construct models as a function of time. Models composed of a 2D and a 3D mechanics model are compared. As it was to be expected, the difference between both models is negligibly small. The same is true when the drug Verapamil is applied to the respective ventricular cell models [4]. Simulations with a low (25 nM) and a high effective free therapeutic plasma concentration (81 nM) were carried out. The resulting membrane potentials of the respective ventricular cell models correspond to [8]: the duration of the action potential decreases with increasing drug concentration. As distinguished from pure electrophysiology models *CellDrum* models are capable of predicting the drug effect on the deflection. Effects on the deflection are shown in Figure 2 for the second model: the deflection decreases with increasing drug concentration.

Figure 1: Central node deflection of a 19 µm thick cardiac tissue constructs. A 2D- (solid line) and a 3D-mechanical model (dashed line) are compared.

Figure 2: Effect of Verapamil on the central node deflection of a 19 µm thick cardiac tissue construct. The second 3D model was used. Results of the control (0 nM, solid line), low EFTPC (25 nM, dashed line), and high EFTPC (81 nM, dotted line) are shown.

3.2 Comparison between a 19 µm and a 100 µm thick cardiac tissue construct model

Figure 4 shows the time function of the central node deflection of two 3D cardiac tissue construct models. Models with different tissue construct thicknesses, 19 µm and 100 µm, are compared. In order to have the same start position, x Pa more pressure is needed to inflate the 100 µm model up to the same height as the 19 µm model. The deflection of the 100 µm model is lower compared to the 19 µm model. Furthermore, the effect of Verapamil on the deflection was investigated in both models. The application of 81 nM Verapamil results in a peak deflection difference of x1 (x2) mm in the 19 (100) µm models.

Figure 3: Central node deflection of two differently thick tissue constructs. 19 µm thick (solid line) and 100 µm thick (dashed line) cardiac tissue construct models are compared. Comparisons were made with no addition of a drug.

4 OUTLOOK

In the present model the electric and passive material properties were modeled to be isotropic according to global isotropy of the cardiac tissue on the *CellDrum*. However, myocardial tissue of the left ventricle is orthotropic in both electrical [20] and passive mechanical properties [21] and thus, it is unsuitable for heart models. Currently, an orthotropic model is developed which is described in the following.

Myocardial tissue is organized in layers with collagen fibers between the sheets and are referred to a right-handed orthonormal coordinate system with the fiber axis $\mathbf{f}_0$, the sheet axis $\mathbf{s}_0$, and sheet normal axis $\mathbf{n}_0$ (Figure 4).

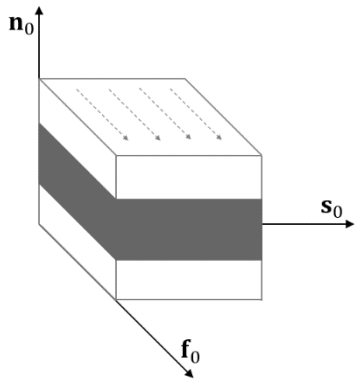


Figure 4: Right-handed orthonormal coordinate system with fiber axis $\mathbf{f}_0$, sheet axis $\mathbf{s}_0$ and sheet-normal axis $\mathbf{n}_0$ in the layered organization of myocardial tissue. The fiber direction is indicated with dashed arrows.

Orthotropic electrical behavior is considered in the monodomain equation (7). Here, an anisotropic conductance tensor $\mathbf{G}$ is used in order to incorporate differences in the action potential propagation in all axes as has been done in [13]. Orthotropic passive material behavior is implemented using the strain energy function published Holzapfel and Ogden [21]

$$\psi = \frac{a}{2b}\exp[b(I_1 - 3)] + \sum_{i=f,s}\frac{a_i}{2b_i}\{exp[b_i(I_{4i} - 1)^2] - 1\} + \frac{a_{fs}}{2b_{fs}}\left[exp\left(b_{fs}I_{8fs}^2\right) - 1\right], \quad (17)$$

where a, b, a_f, a_s, b_f, b_s, a_{fs} and b_{fs} are eight positive material constants which can be derived from [22]. The given strain energy function consists of an isotropic term in I_1, a transversely isotropic term in I_{4f} and I_{4s} and an orthotropic term in I_{8fs}. The respective invariants read

$$I_1 = tr\mathbf{C}, \quad I_{4f} = \mathbf{f}_0 \cdot (\mathbf{C}\mathbf{f}_0), \quad I_{4s} = \mathbf{s}_0 \cdot (\mathbf{C}\mathbf{s}_0), \quad I_{8fs} = \mathbf{f}_0 \cdot (\mathbf{C}\mathbf{s}_0). \quad (18)$$

Using the *CellDrum* geometry the fiber direction will be set to be orientated radially from the center to the edge. Analogue experiments might be possible in the future using mainly myocardial ventricular cells which are cultivated in a way that they grow in the desired direction. This allows to parameterize the active stress model (10) by measuring the cell type specific reference stress having the mentioned passive stress model as basis. All in all, improved models can help to better investigate the effect of drugs *in-silico*.

REFERENCES

[1] A.R. Pinto, A. Ilinykh, M.J. Ivey, J.T. Kuwabara, M.L. D´Antoni, R. Debuque, A. Chandran, L. Wang, K. Arora, N.A. Rosenthal, M.D. Tallquist, Revisiting cardiac cellular composition. *Circulation Research*, **118**, 400-409, 2016.

[2] A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, **117**, 500-544, 1952.

[3] D. Noble, A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pace-maker potentials. *Journal of Physiology*, **160**, 317-352, 1962.

[4] K.H.W.J. ten Tusscher, A.V. Panfilov, Alternans and spiral breakup in a human ventricular tissue model. *American Journal of Physiology – Heart and Circulatory Physiology*, **29**, H1088-H1100, 2006.

[5] N.J. Chandler, I.D. Greener, J.O. Tellez, S. Inada, H. Musa, P. Molenaar, D. DiFrancesco, M. Baruscotti, R. Longhi, R.H. Anderson, R. Billeter, V. Sharma, D.C. Sigg, M.R. Boyett, H. Dobrzynski, Molecular architecture of the human sinus node: insights into the function of the cardiac pacemaker. *Circulation*, **119**, 1562-1575, 2009.

[6] M. Paci, J. Hyttinen, K. Aalto-Setälä, S. Severi, Computational models of ventricular- and atrial-like human induced pluripotent stem cell derived cardiomyocytes. *Annals of Biomedical Engineering*, **41**, 2334-2348, 2013.

[7] O.V. Aslanidi, M. Al-Owais, A.P. Benson, M. Colman, C.J. Garratt, J.P. Greenwood, A.V. Holden, S. Kharche, E. Kinnell, E. Pervolaraki, S. Plein, J. Stott, H. Zhang, Virtual tissue engineering of the human atrium: modeling pharmacological actions on atrial arrhythmogenesis. *European Journal of Pharmaceutical Sciences*, **46**, 209-221, 2012.

[8] G.R. Mirams, Y. Cui, A. Sher, M. Fink, J. Cooper, B.M. Heath, N.C. McMahon, D.J. Gavaghan, D. Noble, Simulation of multiple ion channel block provides improved early prediction of compounds´ clinical torsadogenic risk. *Cardiovascular Research*, **91**, 53-61, 2011.

[9] M. Potse, B. Dubé, J. Richer, A. Vinet, R.M. Gulrajani, A comparison of monodomain and bidomain reaction-diffusion models for action potential propagation in the human heart, *IEEE Transactions on Biomedical Engineering*, **53**, 2425-2435, 2006.

[10] S.A. Niederer, N.P. Smith, An improved numerical method for strong coupling of excitation and contraction models in the heart, *Progress in Biophysics and Molecular Biology*, **96**, 90-111, 2008.

[11] T.S.E. Eriksson, A.J. Prassl, G. Plank, G.A. Holzapfel, Modeling the dispersion in electromechanically coupled myocardium. *International Journal of Numerical Methods in Biomedical Engineering*, **29**, 1267-1284, 2013.

[12] R. Ruiz-Baier, A. Gizzi, S. Rossi, C. Cherubini, A. Laadhari, S. Filippi, A. Quarteroni, Mathematical modelling of active contraction in isolated cardiac myocytes, *Mathematical Medicine and Biology*, **31**, 259-283, 2014.

[13] C.M. Augustin, A. Neic, M. Liebmann, A.J. Prassl, S.A. Niederer, G. Haase, G. Plank, Anatomically accurate high resolution modeling of human whole heart electromechanics: a strongly scalable algebraic multigrid solver method for nonlinear deformation, *Journal of Computational Physics*, **305**, 622-646, 2016.

[14] R. Frotscher, D. Muanghong, G. Dursun, M. Goßmann, A. Temiz-Artmann, M. Staat, Sample-specific adaption of an improved electromechanical model of in vitro cardiac tissue, *Journal of Biomechanics*, 2016 (in press).

[15] A. Quarteroni, T. Lassila, S. Rossi, R. Ruiz-Baier, Integrated heart – Coupling multiscale and Multiphysics models for the simulation of the cardiac function, *Computer Methods in Applied Mechanics and Engineering*, 2016 (submitted), https://www.mate.polimi.it/biblioteca/add/qmox/05-2016.pdf.

[16] M. Sermesant, R. Chabiniok, P. Chinchapatnam, T. Mansi, F. Billet, P. Moireau, J.M. Peyrat, K. Wong, J. Relan, K. Rhode, M. Ginks, P. Lambiase, H. Delingette, M. Sorine, C.A. Rinaldi, D. Chapelle, R. Razavi, N. Ayache. Patient-specific electromechanical models of the heart for the prediction of pacing acute effects in CRT: a preliminary clinical validation. *Medical Image Analysis*, **16**, 201-215, 2012.

[17] E. Berberoglu, H.O. Solmaz, S. Göktepe, Computational modeling of coupled cardiac electromechanics incorporating cardiac dysfunctions. *European Journal of Mechanics – A/ Solids*, **48**, 60-73, 2014.

[18] B. Baillargeon, N. Rebelo, D.D. Fox, R.L. Taylor, E. Kuhl, The Living Heart Project: A robust and integrative simulator for human heart function. *European Journal of Mechanics –A/ Solids*, **48**, 38-47, 2014.

[19] J.D. Bayer, R.C. Blake, G. Plank, N.A. Trayanova, A novel rule-based algorithm for assigning myocardial fiber orientation to computational heart models, *Annals of Biomedical Engineering*, **40**, 2243-2254, 2012.

[20] B.J. Caldwell, M.L. Trew, G.B. Sands, D.A. Hooks, I.J. LeGrice, B.H. Smaill, Three distinct directions of intramural activation reveal nonuniform side-to-side electrical coupling of ventricular myocytes. *Circulation: Arrhytmia and Electrophysiology*, **2**, 433-44, 2009.

[21] G.A. Holzapfel, R.W. Ogden, Constitutive modeling of passive myocardium: a structurally based framework for material characterization. *Philosophical Transactions of the Royal Society A – Mathematical, Physical and Engineering Sciences*, **367**, 3445-3475, 2009.

[22] G. Sommer, A.J. Schriefl, M. Andrä, M. Sacherer, C. Viertler, H. Wolinski, G.A. Holzapfel, Biomechanical properties and microstructure of human ventricular myocardium, *Acta Biomaterialia*, **24**, 172-192, 2015.

[23] P. Pathmanathan, R.A. Gray, Ensuring reliability of safety-critical clinical applications of computational cardiac models. *Frontiers in Physiology*, **4**, 358, 2013.

[24] P. Linder, J. Trzewik, M. Rüffer, G. Artmann, I. Digel, R. Kurz, A. Rothermel, A. Robitzki, A. Temiz-Artmann, Contractile tension and beating rates of self-exciting monolayers and 3D-tissue constructs of neonatal rat cardiomyocytes, *Medical & Biological Engineering & Computing*, **48**, 59-65, 2010.

[25] M. Goßmann, R. Frotscher, P. Linder, S. Neumann, R. Bayer, M. Epple, M. Staat, A. Temiz-Artmann, G. Artmann, Mechano-pharmacological characterization of cardiomyocytes derived from human induced stem cells. *Cellular Physiology and Biology*, 2016 (in press)

[26] R. Frotscher, M. Goßmann, H.J. Raatschen, A. Temiz-Artmann, M. Staat, *Simulation of cardiac cell-seeded membranes using the edge-based smoothed FEM.* In: H. Altenbach,

G.I. Mikhasev (eds), *Shell and membrane theories in mechanics and biology: From macro- to nanoscale structures*, Springer, 2015.

[27] R. Frotscher, J.P. Koch, M. Staat, Computational investigation of drug action on human-induced stem cell derived cardiomyocytes. *Journal of Biomechanical Engineering*, **137**, 071002-071002-7, 2015.

[28] T. Brennan, M. Fink, B. Rodriguez, Multiscale modeling of drug-induced effects on cardiac electrophysiological activity. *European Journal of Pharmaceutical Sciences*, **36**, 62-77, 2009.

[29] A. Gizzi, C. Cherubini, S. Filippi, A. Pandolfi, Theoretical and numerical modeling of nonlinear electromechanics with applications to biological active media. *Communications in Computational Physics*, **17**, 93-126, 2015.

[30] P. J. Hunter, A.D. McCulloch, H.E.D.J. ter Keurs, Modelling the mechanical properties of cardiac muscle. *Progress in Biophysics and Molecular Biology*, **69**, 289–331, 1998.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

BIOMECHANICAL PROPERTIES OF CORNEAL EXTRACELLULAR MATRIX USING EXPERIMENTAL AND COMPUTATIONAL METHODS

Hamed Hatami-Marbini[1]

[1] Department of Mechanical & Industrial Engineering,

University of Illinois at Chicago, Chicago, Illinois 60607

e-mail: hatami@uic.edu

Keywords: cornea, soft tissue, computational method, experiments, mechanical behavior, hydration.

Abstract. *The cornea is a transparent soft tissue at the front of the eye and is mainly responsible for refracting the incoming light and protecting the internal contents of the eye. Although the cornea is composed of different components, it is well-known that its biomechanical properties are due to the structure and function of its extracellular matrix, called stroma. The stroma is composed of regularly arranged collagen fibers embedded in a hydrated proteoglycans matrix. From a mechanics viewpoint, the stroma can be considered as a composite structure in which the collagen fibrils and the hydrated matrix take tensile and compressive forces, respectively.*

In this work, experimental and computational studies conducted to fully characterize the mechanical properties of the cornea in compression and tension. The unconfined compression experiments were conducted on corneal disks while uniaxial tensile tests were performed on corneal rectangular strips. A transversely isotropic constitutive model was used to numerically analyze the experimental measurements and determine the material constants of the cornea. It is shown that the material properties of the cornea in tension and compression are hydration dependent. Therefore, careful attention must be taken in corneal hydration during material characterization experiments.

1 INTRODUCTION

The cornea is a hydrated soft tissue which transmits the incident light while acting as a resilient barrier to the mechanical forces. From the anterior to posterior, the cornea is composed of epithelial cells, Bowman's zone, stroma, Descemet's membrane, and endothelial cells. The biomechanical properties of the cornea are mainly derived from the stromal layer. The mechanical behavior of corneal stroma can be determined using uniaxial strip extentiometry, inflation tests, indentation tests, and compressive experiments.

The uniaxial tensile tests are done using rectangular corneal strips subjected to tensile loading [1-8]. Because of the well-known limitations of this testing technique [9], it has often been used for comparative studies [3, 6, 10]. Compared to the uniaxial testing technique, inflation tests are able to better represent natural conditions of the cornea as they do not use excised samples [11, 12]. However, the use of pressurization fluid during inflation experiments results in corneal swelling [8, 11]. This unwanted swelling may affect the precision of experimental measurements. The compressive experiments have also been used to study the mechanical behavior of the cornea in terms of its swelling pressure–thickness relation [13-17]. The findings of the above experimental studies show a wide range of variation in the reported material parameters for the corneal tissue. This variation has been often explained in terms of inherent differences in sample properties and experimental conditions.

The objective of the present manuscript is to show that corneal mechanical properties are hydration dependent and the lack of control on tissue hydration explain the existing discrepancies in the literature. For this purpose, we investigated the effects of hydration on tensile and compressive properties of the cornea [18-22]. We performed unconfined compression experiments and used a transversely isotropic biphasic model to determine corneal material parameters as a function of hydration.

2 MATERIALS AND METHODS

2.1 Sample Preparation

Porcine and bovine eyes were obtained from a slaughterhouse. At the laboratory, corneoscleral rims were excised from normal looking corneas. The epithelium and endothelium were removed from the corneal surface. A biopsy punch and double blade device were used to prepare corneal disks and corneal strips. The stromal buttons were weighed with 0.1 mg accuracy and their thickness was measured using an ultrasound pachymeter. The stromal buttons were used in unconfined experiments. The corneal strips were divided into the following groups based on their average thicknesses: 0.3, 0.35, 0.4, 0.5, 0.7, 0.9, and 1.1 mm. In order to alter the thickness (hydration) of the strips, all specimens were air-dried in a desiccator and were then immersed in an ionic solution until the desired thickness of one of the groups was obtained. In order to obtain the hydration of the specimens, the linear hydration-thickness relation, i.e. $H_w = 5.3t - 0.67$, was used [23]. In hydration-thickness relation, H_w is hydration in mg water / mg dry tissue and t denotes the thickness in mm.

2.2 Uniaxial Tensile Experiments

The samples were tested using a custom-built micro tensile device composed of a linear stepper motor, a submersible load cell, and serrated grips. The strips were mounted between serrated grips and glue was used to prevent any possible slippage. Mineral oil was used to fill the chamber in order to prevent any changes in the hydration (thickness) of the specimens. In order to remove possible variations in experimental measurements caused by stress-strain his-

tory, a preconditioning step was done at the beginning of all experiments [6, 7]. The preconditioning was composed of five loading/unloading cycles and three successive relaxation tests. A tare load of 100 mN was then applied to straighten the samples and determine the initial length of specimens. This initial length was used to obtain the engineering strain. The experiments were conducted using an engineering strain of 5% and a displacement rate of 2 mm/min. The stress was calculated from dividing the recorded force by the initial cross-sectional area.

2.3 Unconfined Compression Experiments

A DHR-2 Rheometer was used to perform unconfined compression experiments. The chamber of the rheometer was first filled with NaCl solution and the corneal buttons were placed on the center of the impermeable platen surface. Initially A preload equivalent to a stress of 1.5 KPa was applied in order to create a uniform starting point in all experiments. The equilibrium thickness of corneal disks at this load was taken as the initial thickness for calculating engineering strain. A series of ramp-hold compressive experiments was then performed. The reaction stress was calculated by dividing the recorded force by the initial area of the samples. A constant displacement rate of 0.15 μm/s was used in all experiments.

2.4 Data Analysis

The stress-strain curves obtained from uniaxial experiments were represented using the exponential expression $\sigma = A(e^{B\varepsilon} - 1) + \sigma_0$ and power-law function $\sigma = \alpha\varepsilon^{\beta} + \sigma_0$ [1, 24]. Here, ε is the engineering strain, σ is the stress, σ_0 is the initial tare stress and A, *B*, α, and β are fit parameters.

The unconfined compression experimental measurements at each ramp-and-hold were curve-fitted to a transversely isotropic biphasic model. The compressive stress of a stromal disk with radius *a* subjected to ramp-hold compressive strain is given by [25],

$$\sigma(t) = \begin{cases} E_z \dot{\varepsilon}_0 t + E_r \dfrac{\dot{\varepsilon}_0 a^2}{\kappa_r \mathrm{H}} B_3 \left(\dfrac{1}{8} - \displaystyle\sum_{n=1}^{\infty} \dfrac{e^{-x_n^2 \kappa_r \mathrm{H} a^{-2} t}}{x_n^2 \left(B_2^2 x_n^2 - B_1 \left(1 + \nu_{r\theta}\right)^{-1} \right)} \right) & t \le t_0 \\ E_z \dot{\varepsilon}_0 t_0 - E_r \dfrac{\dot{\varepsilon}_0 a^2}{\kappa_r \mathrm{H}} B_3 \displaystyle\sum_{n=1}^{\infty} \dfrac{e^{-x_n^2 \kappa_r \mathrm{H} a^{-2} t} - e^{-x_n^2 \kappa_r \mathrm{H} a^{-2} (t - t_0)}}{x_n^2 \left(B_2^2 x_n^2 - B_1 \left(1 + \nu_{r\theta}\right)^{-1} \right)} & t \ge t_0 \end{cases} \tag{1}$$

where

$$B_1 = 1 - \nu_{r\theta} - 2\nu_{zr}^2 E_r \mathrm{E}_z^{-1}\ , B_2 = \left(1 - \nu_{zr}^2 E_r \mathrm{E}_z^{-1}\right)\left(1 + \nu_{r\theta}\right)^{-1}, \tag{2}$$

$$B_3 = \left(1 - 2\nu_{zr}^2\right) B_2 B_1^{-1}\ , \mathrm{H} = E_r \left(1 - \nu_{zr}^2 E_r \mathrm{E}_z^{-1}\right)\left((1 + \nu_{r\theta}) B_1\right)^{-1},$$

t_0 is the ramp time, $\dot{\varepsilon}_0$ is the ramp compressive strain rate, E_r is the in-plane Young's modulus, E_z is the out-of-plane Young's modulus, $\nu_{r\theta}$ is the in-plane Poison's ratio, ν_{zr} is out-of-plane Poison's ratio, κ_r is the permeability coefficient and x_n's are the solution of following equation

$$J_1(x) - \frac{1 - \nu_{zr}^2 E_r \mathrm{E}_z^{-1}}{1 - \nu_{r\theta} - 2\nu_{zr}^2 E_r \mathrm{E}_z^{-1}} x J_0(x) = 0, \tag{3}$$

and J_0 and J_1 are the zeroth and first order Bessel functions of the first kind, respectively. The experimental reaction stress was fit to equation **Error! Reference source not found.** in order to find the material properties.

In all cases, the coefficients of determinations R^2 were calculated to determine the goodness of the fits. Also, One way ANOVA was used to quantify the effect of corneal thickness on biomechanical properties; statistical significance of $p<0.05$ was used in all statistical analyses. Data is reported as mean ± standard deviation.

3 RESULTS

Figure 1a shows the stress-strain behavior of corneal samples with seven different thicknesses. The corneal strips with lower hydration showed a stiffer response. Both exponential and power-law expression successfully curve-fitted the stress-strain curves of samples with thickness t>0.35mm. Dashed lines in Figure 1a show the power-law fits. The average tangent modulus for different thickness groups is obtained from the stress-strain curves, Figure 1b. This plot show that there is a significant difference between tangent modulus of different groups (P<0.05).

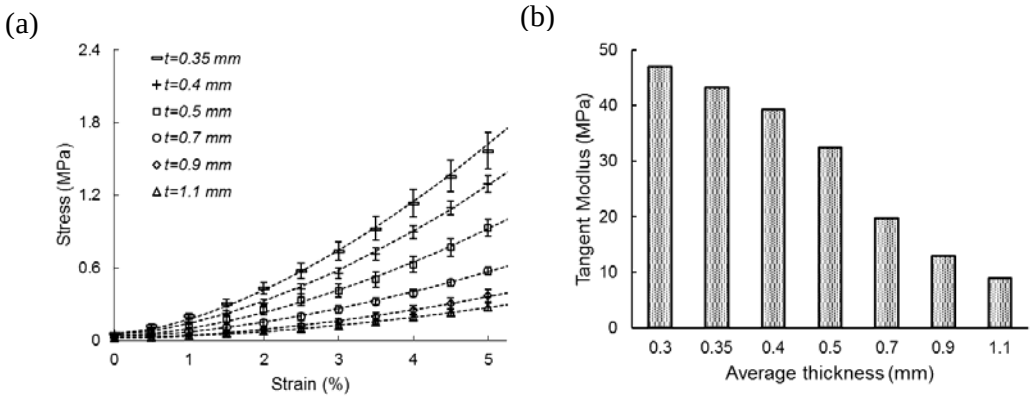


Figure 1: (a) The thickness effects on tensile behavior of corneal strips. The dashed lines show the fits obtained using the power-law relation. (b) The effect of thickness on tangent modulus at 5% strain.

Figure 2 shows the reaction stress variation as a function of time in a typical stepwise unconfined compression experiment. The reaction stress increased non-linearly during the ramp loading and decreased to an equilibrium stress during the relaxation stage. The slope of the equilibrium stress-strain curve was used to calculate the equilibrium moduli of corneal samples as a function of average thickness, Figures 2b. ANOVA analysis showed that variation of these moduli was not statistically significant when the thickness is between *1.07 mm* and *1.27 mm*. The in-plane Young's modulus and permeability coefficient were determined by fitting the transversely isotropic biphasic model to the experimental measurements. An average coefficient of determination $r^2=0.979$ was obtained for the fits. Figures 2c-d show the corneal in-plane Young's modulus and permeability coefficient as a function of average thickness.

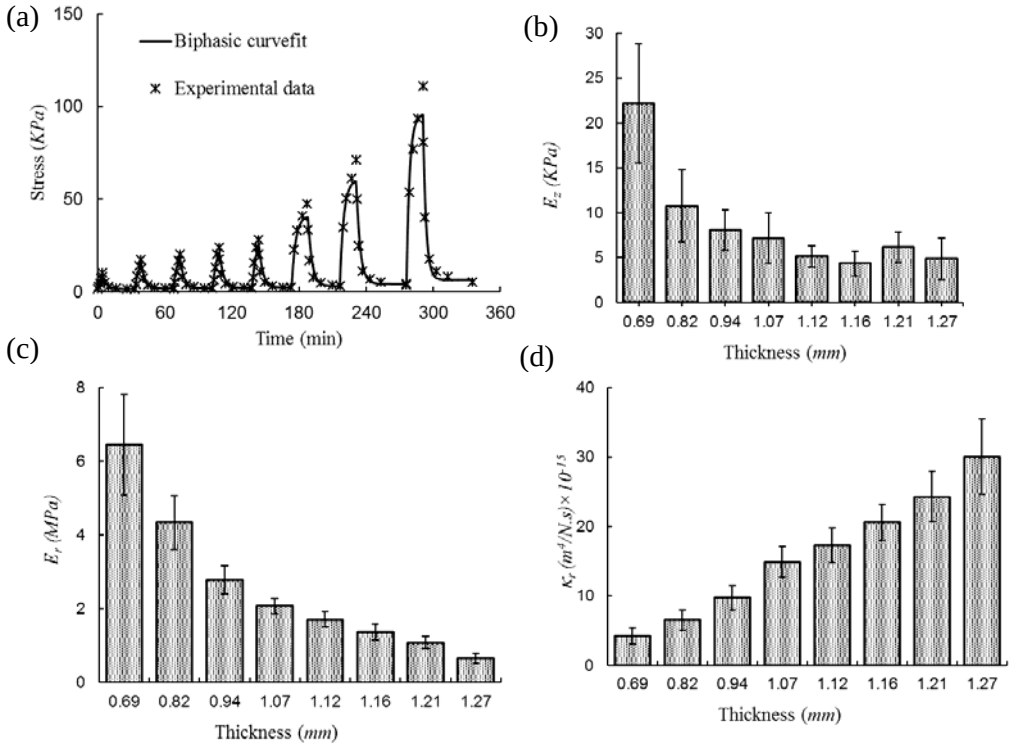


Figure 2: (a) Typical experimental measurements from stepwise unconfined compression tests. The solid line shows the curve-fit of the experimental measurements with a transversely isotropic biphasic model. (b) The equilibrium out-of-plane Young's modulus (E_z), (c) the in-plane Young's modulus, and (d) the permeability coefficient of the cornea as a function of average thickness.

4 DISCUSSION

The present study investigated the relation between corneal thickness (hydration) and material properties. For this purpose, the effects of corneal hydration were investigated in extensometry and unconfined compression experiments. This extensometry has been widely used to determine the stress-strain behavior of the cornea; which resulted in a wide range of variations for material parameters [8, 9, 24]. The present study showed that lack of control on thickness (hydration) has significantly contributed to this variation. In order to further investigate the hydration dependent mechanical behavior of the cornea, corneal buttons were tested under unconfined compression. A transversely isotropic model was used to curve-fit the experimental data and obtain the hydration dependent in-plane Young's modulus, out-of-plane compressive modulus, and permeability coefficient of the cornea. It is seen that mechanical properties strongly depend on the corneal thickness (hydration), Figure 2.

The hydration dependent properties of the cornea can be explained in terms of its microstructure. The extracellular matrix of the cornea like many other soft tissues is primarily composed of collagen fibrils and proteoglycans. In particular, regularly distributed collagen fibrils are surrounded by a proteoglycan matrix. The presence of proteoglycan creates a strong tendency for the tissue to swell when immersed in a water-based solution. With increasing thickness (caused by hydration changes), the bonds between collagen fibrils and proteoglycan side chains possibly break and a softer mechanical response is obtained [26-28]. Similarly, with increasing the thickness, the total fixed charge density (and subsequently the osmotic pressure)

decreases. Therefore, it can be proposed that hydration-dependent properties of the cornea are due to the molecular-level interactions between collagen fibrils and proteoglycans.

5 CONCLUSIONS

The present work investigated the hydration dependent material properties of the cornea. Two different experimental techniques, i.e. uniaxial tensile tests and unconfined compression experiments, were used to characterize the corneal mechanical behavior. Furthermore, numerical models were used to analyze the experimental measurements. It was concluded that swollen corneal samples have softer mechanical properties in comparison with dehydrated ones. Therefore, careful attention must be taken in interpreting experimental studies that do not fully control the thickness.

REFERENCES

[1] D.A. Hoeltzel, et al., Strip Extensiometry for Comparison of the Mechanical Response of Bovine, Rabbit, and Human Corneas. *Journal of Biomechanical Engineering*. **114**(2), 202-215, 1992.

[2] J. Kampmeier, et al., Thermal and Biomechanical Parameters of Porcine Cornea. *Cornea*. **19**(3), 355-363, 2000.

[3] Y. Zeng, et al., A comparison of biomechanical properties between human and porcine cornea. *Journal of Biomechanics*. **34**(4), 533-537, 2001.

[4] G. Wollensak, E. Spoerl, T. Seiler, Stress-strain measurements of human and porcine corneas after riboflavin–ultraviolet-A-induced cross-linking. *Journal of Cataract and Refractive Surgery*. **29**(9), 1780-1785, 2003.

[5] A. Elsheikh, D. Alhasso, Mechanical anisotropy of porcine cornea and correlation with stromal microstructure. *Experimental Eye Research*. **88**(6), 1084-1091, 2009.

[6] B.L. Boyce, et al., Stress-controlled viscoelastic tensile response of bovine cornea. *Journal of Biomechanics*. **40**(11), 2367-2376, 2007.

[7] A. Elsheikh, W. Kassem, S.W. Jones, Strain-rate sensitivity of porcine and ovine corneas. *Acta of Bioengineering and Biomechanics*. **13**(2), 25-36, 2011.

[8] F. Boschetti, et al., Mechanical Characterization of Porcine Corneas. *Journal of Biomechanical Engineering*. **134**(3), 031003-9, 2012.

[9] A. Elsheikh, K. Anderson, Comparative study of corneal strip extensometry and inflation tests. *Journal of The Royal Society Interface*. **2**(3), 177-185, 2005.

[10] A. Elsheikh, D. Alhasso, P. Rama, Biomechanical properties of human and porcine corneas. *Experimental Eye Research*. **86**(5), 783-790, 2008.

[11] B.L. Boyce, et al., Full-field deformation of bovine cornea under constrained inflation conditions. *Biomaterials*. **29**(28), 3896-3904, 2008.

[12] A. Elsheikh, D. Wang, D. Pye, Determination of the modulus of elasticity of the human cornea. *Journal of Refractive Surgery*. **23**(8), 808-18, 2007.

[13] T. Olsen, S. Sperling, The swelling pressure of the human corneal stroma as determined by a new method. *Experimental Eye Research*. **44**(4), 481-490, 1987.

[14] B.O. Hedbys, C.H. Dohlman, A new method for the determination of the swelling pressure of the corneal stroma in vitro. *Experimental Eye Research*. **2**(2), 122-129, 1963.

[15] M.H. Friedman, et al., Contribution of the Donnan osmotic pressure to the swelling pressure of corneal stroma. *American Journal of Physiology*. **222**(6), 1565-1570, 1972.
[16] T. Wiley, I. Fatt, Effects of solutes on the swelling pressure of the corneal stroma. *Investigative Ophthalmology*. **14**(9), 684-688, 1975.
[17] S. Hodson, et al., Observation on the human cornea in vitro. *Exp Eye Res*. **32**(3), 353-60, 1981.
[18] H. Hatami-Marbini, E. Etebu, Rate Dependent Biomechanical Properties of Corneal Stroma in Unconfined Compression *Biorheology*. **50**(3-4), 133-47, 2013.
[19] H. Hatami-Marbini, E. Etebu, An experimental and theoretical analysis of unconfined compression of corneal stroma. *Journal of Biomechanics*. **46**(10), 1752-1758, 2013.
[20] H. Hatami-Marbini, E. Etebu, Hydration dependent biomechanical properties of the corneal stroma. *Exp Eye Res*. **116**, 47-54, 2013.
[21] H. Hatami-Marbini, Hydration Dependent Viscoelastic Tensile Behavior of Cornea. *Ann Biomed Eng*. **42**(8), 1740-8, 2014.
[22] H. Hatami-Marbini, R. Maulik, A Biphasic Transversely Isotropic Poroviscoelastic Model for the Unconfined Compression of Hydrated Soft Tissue. *Journal of biomechanical engineering*. **138**(3), 031003-1-6, 2016.
[23] B.O. Hedbys, S. Mishima, The thickness-hydration relationship of the cornea. *Experimental Eye Research*. **5**(3), 221-228, 1966.
[24] S.L. Woo, et al., Nonlinear material properties of intact cornea and sclera. *Exp Eye Res*. **14**(1), 29-39, 1972.
[25] B. Cohen, W.M. Lai, V.C. Mow, A transversely isotropic biphasic model for unconfined compression of growth plate and chondroepiphysis. *Journal of biomechanical engineering*. **120**(4), 491-6, 1998.
[26] J.E. Scott, Proteoglycan: collagen interactions and corneal ultrastructure. *Biochem. Soc. Trans*. **19**(4), 877-881, 1991.
[27] X. Cheng, H. Hatami-Marbini, P.M. Pinsky, Modeling Collagen-Proteoglycan Structural Interactions in the Human Cornea, in *Computer Models in Biomechanics*, G.A. Holzapfel and E. Kuhl, Editors., Springer.11-24, 2013.
[28] H. Hatami-Marbini, P.M. Pinsky. On Mechanics of Connective Tissue: Assessing the Electrostatic Contribution to Corneal Stroma Elasticity. in *Material Reseaech Society Proceedings*. Boston, 2009

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELING OF PATIENT-SPECIFIC CASES OF ATHEROSCLEROSIS IN CAROTID ARTERIES

Timur Gamilov[1,2], Roman Pryamonosov[1,2] and Sergey Simakov[1,2]

[1]Moscow Institute of Physics and Technology
9, Institutskiy per.,Dolgoprudny, Russia
e-mail: gamilov@crec.mipt ru

[2] Institute of Numerical Mathematics RAS
8, ul. Gubkina, Moscow, Russia

Keywords: Circle of Willis, Patient-specific, Carotid artery stenosis, Vessel segmentation, Microcirculation, 1d haemodynamics

Abstract. *1D model is used to simulate blood flow in major vessels of the upper body and head. The 1D part is stated in terms of viscous incompressible fluid flow in the network of elastic tubes. Two different types of junctions are considered: junctions between major vessels and junctions between arteries and veins. Vessel network reconstruction algorithm consists of vessel segmentation, thinning-based obtaining of set of centerlines, and graph reconstruction. Input data is 3D DICOM datasets, obtained with contrast enhanced Computed Tomography (CT) Angiography. Constructed model is used to study the influence of carotid artery stenosis on the direction of blood flow in the circle of Willis.*

1 Introduction

Investigation of pulsatile blood flow in cerebral normal and diseased vessels provides important information for clinical diagnosis. One-dimensional modelling is a well-developed approach in this area [1, 2]. It allows to simulate large number of arteries and veins and take into account regulatory mechanisms. Computational costs of such approach allow to run numerous numerical experiments in a short time on a laptop. 1D models can be used to study the impact of stenosis on blood flow patterns in cerebral region for different groups of patients. They can also help to choose the most optimal strategy of treatment.

The structure of circle of Willis can be very individual. Simulations based on pre-made averaged vessel structures could produce huge errors. This increases the importance of patient-specific approach in simulations of cerebral blood circulation.

In this work the 1D blood flow model [3] is used to investigate the impact of a stenosis in internal carotid artery on blood flow in the circle of Willis. Microcirculation regions between arteries and veins are simulated with the help of Poiseuille's pressure drop condition. An important feature of this work is reconstruction of patient-specific vessel structure based on CT images.

2 Methods

2.1 Blood flow model

1D haemodynamics model used in this work is the model of viscous incompressible fluid in a network of elastic tubes. Network of arteries is obtained from patient's CT images. Veins are considered to have the same structure as arteries, but different parameters. In this section brief description of the model is presented, for details we refer to [3, 4]. Blood flow in each vessel is described by hyperbolic set of mass and momentum balances

$$\partial A_k/\partial t + \partial(A_k u_k)/\partial x = 0, \tag{1}$$

$$\partial u_k/\partial t + \partial\left(u_k^2/2 + p_k/\rho\right)/\partial x = f_{fr}(A_k, u_k), \tag{2}$$

where k is the index of the vessel; t is the time; x is the distance along the vessel counted from the vessel junction point; ρ is the blood density (constant); $A_k(t,x)$ is the vessel cross-section area; p_k is the blood pressure; $u_k(t,x)$ is the linear velocity averaged over the cross-section; f_{tr} is the friction force. Relationship between pressure and cross-section is given by wall-state equation:

$$p_k(A_k) - p_{*k} = \rho_w c_k^2 f(A_k), \tag{3}$$

where ρ_w is the vessel wall density (constant); $f(A)$ is function

$$f(A_k) = \begin{cases} \exp\left(A_k/A_{0k} - 1\right) - 1, & A_k/A_{0k} > 1 \\ \ln A_k/A_{0k}, & A_k/A_{0k} \leqslant 1, \end{cases} \tag{4}$$

p_{*k} is the pressure in tissues surrounding the vessel; A_{0k} is the unstressed cross-sectional area; c_k defines elastic properties of the wall and can be considered as velocity of small disturbances propagation [5].

At the terminal point of the venous system the pressure $p_H = 8mmHg$ is set as the boundary condition. At the entry point of the aorta the blood flow is assigned

$$u(t,0)\,A(t,0) = Q_H(t). \tag{5}$$

Here function $Q_H(t)$ corresponds to the heart rate value of 1 Hz and stroke volume of 65 ml [6].

Bifurcation points are divided into two groups: 1) common junctions (between arteries and arteries or veins and veins); 2) microcirculation junctions (between arteries and veins). At common junctions continuity of pressure is postulated

$$p_i\left(A_i\left(t,\tilde{x}_i\right)\right) - p_j\left(A_j\left(t,\tilde{x}_j\right)\right) = 0, \tag{6}$$

where i, j are indexes of the vessels. At junction point between M arteries and veins the Poiseuille's pressure drop condition is used

$$p_k\left(A_k\left(t,\tilde{x}_k\right)\right) - p_{node}\left(t\right) = \varepsilon_k R_k A_k\left(t,\tilde{x}_k\right) u_k\left(t,\tilde{x}_k\right), k = k_1, k_2, \ldots, k_M, \tag{7}$$

where R_k represents the hydraulic resistance of *k*-th vessel in the junction; $\varepsilon = 1, \tilde{x}_k = L_k$ for incoming vessels, $\varepsilon = -1, \tilde{x}_k = 0$ for outgoing vessels; p_{node} is the pressure in the junction. Parameters R_k are adjusted to simulate pressure drop between arteries and veins. To close the system, we add the mass conservation condition and compatibility conditions of hyperbolic set (1),(2) (see [4]).

2.2 Reconstruction of patient specific vessel structure

Vessel network reconstruction algorithm consists of vessel segmentation, thinning-based extraction of centerlines, and graph reconstruction [11]. Input data is 3D DICOM datasets, obtained with contrast enhanced Computed Tomography Angiography. Resolution of each 2D transverse slice is 512x512 voxels. Only quasi-isotopic voxel grids with deviation from cubic grids less than 10% were used; other grids can be resampled into isotopic ones.

Anatomy of cerebral vessels in human body is represented by carotid arteries rooting at aorta, vertebral arteries separating from subclavian arteries (that also rooting at aorta), and small arteries in brain. Carotid and vertebral arteries merge in the circle of Willis that allows blood bypass stenoses in neck vessels.

Since all arteries studied originate from aorta, it is natural to segment aorta first (Fig. 2.2 (d)). Our method utilize Hough Circleness [8] and IDT algorithm [9] to produce the aorta mask [11]. Firstly, on the bottommost slice of dataset Hough Circleness algorithm finds largest bright disk corresponding to aorta cut. Secondly, using lowest intensity inside the aorta cut as the threshold we obtain mask M. This mask usually consists of aorta, cerebral vessels, and other bright tissues. Thirdly, IDT algorithm cuts mask M at bottlenecks and outputs mask M_A including aorta and some parts of cerebral and coronary vessels. Finally, morphological operations are used to remove these vessel parts from M_A.

Second segmentation step is to decrease intensities of bones (Fig. 2.2 (c)). We use multi-scale MMBE algorithm [10] which requires two datasets: enhanced with contrast agent and not enhanced. This is the main drawback of our method because every patient has to experience procedure twice and obtains double dose of radiation. However, this step is essential because vertebral arteries are very hard to separate from neck bones on CT images.

Once the bones are darkened, Frangi Vesselness filter [7] is used to segment arteries (Fig. 2.2 (e)). Rather than assuming that the segmentation is the biggest connectivity component of voxels with high vesselness values, we search for all such voxel connectivity components next to the aorta mask. This approach considers that stenoses may divide carotid and vertebral arteries in several parts.

Finally, segmentation errors near the aorta border must be removed (Fig. 2.2 (f)). Despite Frangi Vesselness effectiveness, it can produce segmentation "leaks" near large bright structures

like aorta. Let us denote the mask of cerebral vessels by M_v. Leak removal is represented by iterative procedure working with voxel layers L_i. Each layer L_i is a set of voxels in mask M_v separated from the aorta border by i voxels. On every iteration algorithm removes voxels from layer L_i that have no adjacent voxels in layer L_{i+1}. Algorithm iterating for parameter values $i = i_{max} \dots 1$. Start parameter value i_{max} must be bigger than thickness of "leak" and experiments has shown that i_{max} can be set to 15 voxels.

Entire segmentation algorithm is shown in Fig. 2.2. Bone removal and aorta segmentation are independent steps and may be performed in any order. If the aorta is segmented first, the initial mask M will contain neck bones and skull as shown in Fig. 2.2 (d2).

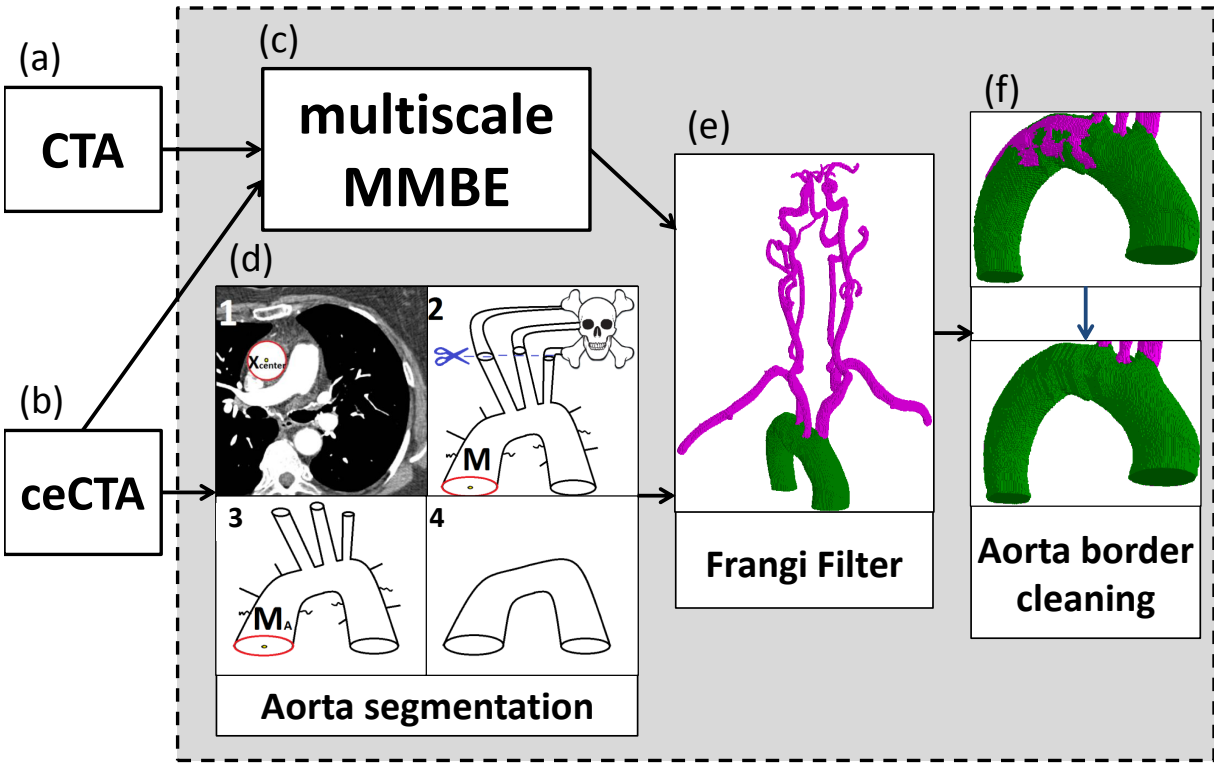


Figure 1: Segmentation algorithm takes (a) CT angiography image and (b) contrast enhanced CT image; (c) bones are removed by multiscale Matched Mask Bone Elimination; (d) essential steps of aorta segmentation: (d1) Hough Circleness, (d2) IDT algorithm cuts mask M at bottlenecks, removing bones and other bright structures and producing (d3) mask M_A, (d4) aorta arch branches and coronary vessels are removed from M_A via morphological operations; (e) cerebral arteries are segmented by Frangi Vesselness Filter; (f) cleaning of aorta border. Mask of aorta is shown by green, cerebral arteries are purple.

After the segmentation, vessel centerlines are extracted with a version of Thinning method with False Twigs Elimination algorithm [11]. These centerlines are used to produce a graph of arterial network where each node corresponds to a bifurcation or an end of vessel and to each edge the length and mean radius of corresponding vessel tube are assigned.

2.3 Computational domain

The extracted network of arteries is presented in Fig. 2. The network of veins is assumed to have similar structure and be connected to the arteries at terminal points. Parameters c_k and R_k were adjusted so that the model reproduces physiological values of blood flows as well as blood velocities measured in control points. The velocities were measured with the Doppler ultrasound method.

The circle of Willis for this particular patient is not closed and consists of vessels 36-35-28-73-85-10-7. The stenosis with degree 90 % is located in the left internal carotid artery. It was simulated by separating the stenosed vessel into three parts: the stenosed part, the proximal part and the distal part. Parameters of the proximal and distal parts correspond to the parameters of the initial non-stenosed vessel. The cross-section of the stenosed part was decreased.

3 Results

Table 1 shows comparison between simulated blood flow velocities and measured velocities for the patient with 90 % stenosis. Measurements after stenosis treatment were not available, so they are substituted by physiological range according to medical literature [6].

Artery (vessel on Fig. 2)	Velocity with stenosis, cm/s						Velocity without stenosis, cm/s	
	Right			Left				
	Sim.	Meas.	Error	Sim.	Meas.	Error	Sim.	Physiol. range
Common carotid (26,3)	50	55	9%	51	54	5.5%	60	50-104
Internal carotid (27,86)	72	67	7%	240	220	10%	48	32-100

Table 1: Simulated blood flow velocities (sim.), measured velocities before stenosis treatment (meas.), physiological range of velocities without stenosis (physiol. range)

Two series of calculation were performed. In the first series all parameters c_k were decreased by 50%, simulating very elastic vessels. In the second series all parameters c_k were increased by 50%, simulating very stiff vessels. Blood flow through vessel 85 (see Fig. 2) was studied for different degrees of stenosis. Results are shown on Fig. 3. Low values of blood flow can be explained by very weak interaction between left and right parts of the Willis circle of the particular patient. Curves in Fig. 3 are almost identical and cross zero line near 54% value.

4 Conclusion

The numerical results show that the proposed vessel network reconstruction algorithm provides a good basis for patient specific simulations. It can be used for the prediction of results of stenosis treatment and for calculation blood flow patterns in hardly accessible regions. The main problems of our approach are necessity to perform two CT scans (with and without contrast) and some difficulties in detection of small vessels in the circle of Willis.

According to Fig. 3, the direction of blood flow in the circle of Willis depends heavily on the degree of stenosis. Certain stenoses can stop blood flow in certain vessels, increasing risk of formation of blood clots. As a result, a relatively weak stenosis (54 %) can be more dangerous than a stronger one.

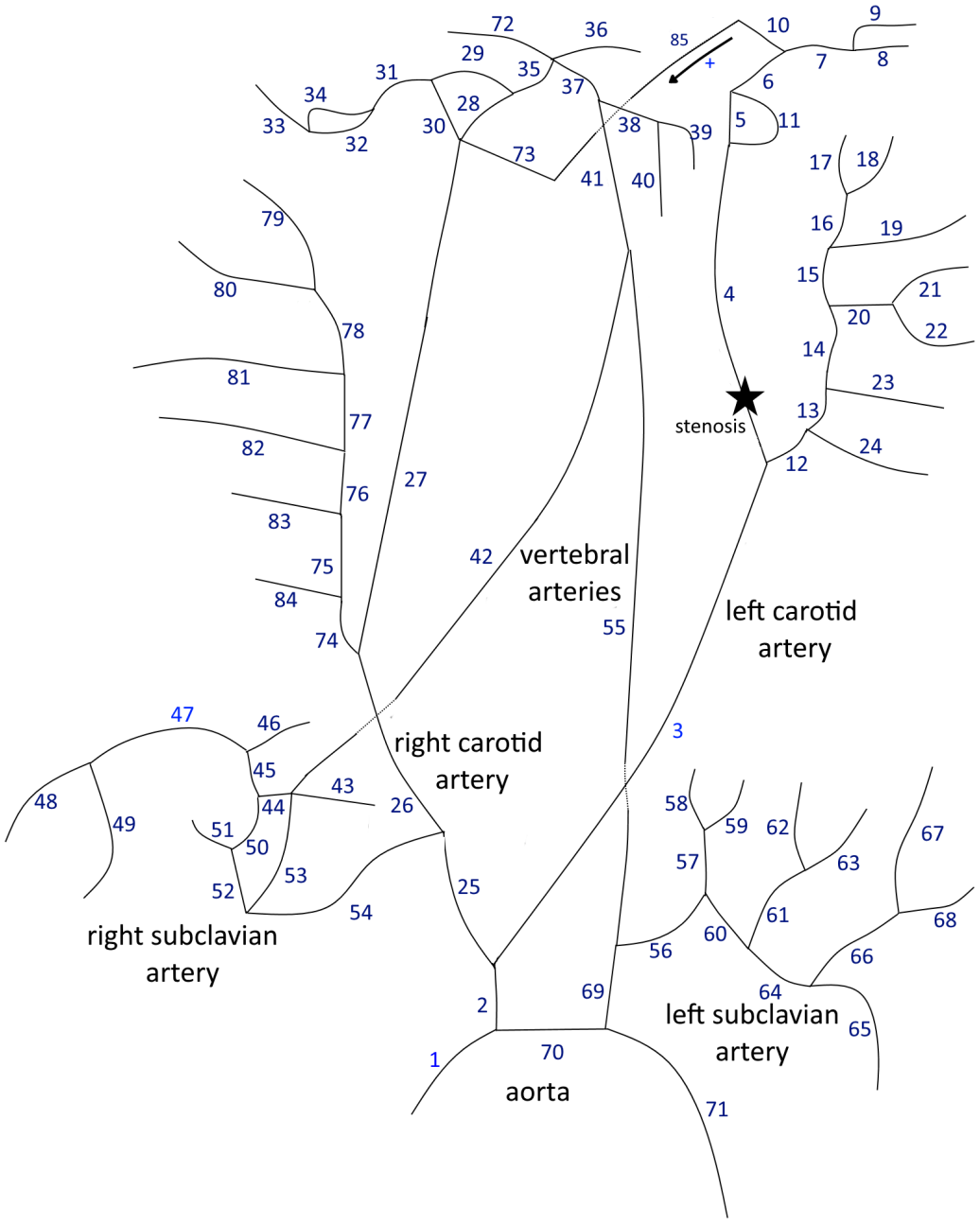


Figure 2: The structure of reconstructed arterial part based on anonymous patient-specific data set. The star designates stenosis. The arrow shows positive blood flow direction in vessel 85. Parameters of the vessels are presented in Tab. 4

Another interesting result is the very small difference between curves for elastic and rigid vessels in Fig. 3. This shows that risks caused by changes in blood flow patterns due to the stenosis are the same for different groups of people: old people, athletes, smokers, etc.

Cerebral blood circulation model presented in this work does not take into account many important regulation mechanisms, such as baroreflex, regulation of CO_2, etc. Regulatory mechanisms are essential in sustaining the blood pressure in cerebral arteries at necessary level. This drawback is partially compensated by pressure continuity condition in bifurcations, but still limits possible applications of the model.

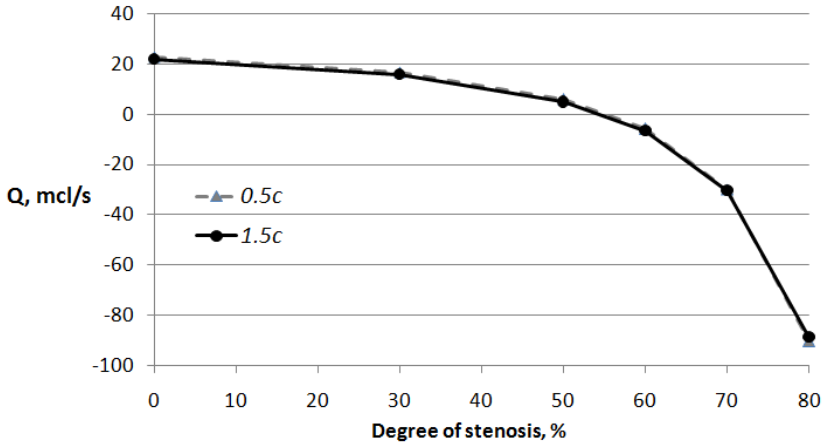


Figure 3: Blood flow in vessel 85 for different degrees of stenosis in left internal carotid artery (see Fig. 2) for elastic vessels ($0.5c$) and rigid vessels ($1.5c$).

Acknowledgement. The research was supported by Russian Science Foundation (RSF) grant 14-31-00024. The authors acknowledges the staff of I.M.Sechenov First Moscow State Medical University and especially Philip Kopylov for giving us patient-specific data.

REFERENCES

[1] Lucas O. Mller and Eleuterio F. Toro, A global multiscale mathematical model for the human circulation with emphasis on the venous system, *IJNMBE*, **30(7)**, 681–725, 2014.

[2] J. Alastruey, S. M. Moore, K. H. Parker, T. David, J. Peir, S. J. Sherwin, Reduced modelling of blood flow in the cerebral circulation: Coupling 1-D, 0-D and cerebral autoregulation models, *International Journal for Numerical Methods in Fluids*, **56(8)**, 1061–7, 2008.

[3] T.M. Gamilov, P.Y. Kopylov, R.A. Pryamonosov, S.S. Simakov, Virtual fractional flow reserve assessment in patient-specific coronary networks by 1D hemodynamic model, *RJNAMM*, **30(5)**, 269–276, 2015.

[4] S.S. Simakov, T.M. Gamilov TM, Y.N. Soe, Computational study of blood flow in lower extremities under intense physical load , *RJNAMM*, **28(5)**, 485–504, 2013.

[5] I. B. Wilkinson, J. R. Cockcroft, D. J. Webb. Pulse wave analysis and arterial stiffness *J. Cardiovasc. Pharmacol.*, **32, Suppl.3**, S33–7, 1998.

[6] W.F. Ganong, *Review of Medical Physiology*, Stamford, CT, Appleton and Lange, 1999.

[7] A.F. Frangi et al., Multiscale Vessel Enhancement Filtering. *Medical Image Computing and Computer-Assisted InterventationMICCAI98*, **1496**, 130–137, 1998.

[8] T.S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, R. Whitaker, Engineering and Algorithm Design for an Image Processing API: A Technical Report on ITK - The Insight Toolkit, In Proc. of *Medicine Meets Virtual Reality*, J. Westwood, ed., IOS Press Amsterdam 586-592, 2002.

[9] L. Grady, Fast, Quality, Segmentation of Large Volumes – Isoperimetric Distance Trees. *Computer Vision – ECCV 2006*, **3953**, 3711-3723, 2007.

[10] H.A. Gratama van Andel, Removal of Bone in CT Angiography by Multiscale Matched Mask Bone Elimination. *Medical Physics*, **34**, 3711-3723, 2007.

[11] A. Danilov, Yu. Ivanov, R. Pryamonosov, Yu. Vassilevski, Methods of Graph Network Reconstruction in Personalized Medicine. *International Journal for Numerical Methods in Biomedical Engineering*, 2015. doi: 10.1002/cnm.2754

k	l, cm	d, cm	c, cm/s	R, $\frac{kba \cdot s}{cm^3}$	k	l, cm	d, cm	c, cm/s	R, $\frac{kba \cdot s}{cm^3}$
1	3.05	2.99	600		44	2.29	0.34	700	
2	1.21	2.42	700		45	0.46	0.92	700	
3	14.9	0.74	700		46	2.72	0.29	700	50
4	15.34	0.4	700		47	8.40	0.72	700	
5	0.24	0.25	700		48	1.75	0.34	700	50
6	0.77	0.28	700		49	5.27	0.60	700	50
7	1.43	0.23	700		50	0.81	1.00	700	
8	0.56	0.17	700	25	51	7.11	0.25	700	50
9	1.64	0.16	700	25	52	0.42	1.10	700	
10	1.72	0.18	700		53	2.79	0.49	700	
11	2.76	0.76	700		54	3.02	0.96	700	
12	0.51	0.39	700		55	26.9	0.28	700	
13	1.30	0.37	700		56	2.15	0.83	700	
14	2.14	0.38	700		57	0.78	0.49	700	
15	4.55	0.35	700		58	0.55	0.31	700	25
16	0.27	0.29	700		59	1.04	0.24	700	25
17	8.82	0.18	700	100	60	0.41	0.7	700	
18	0.59	0.16	700	100	61	2.05	0.42	700	
19	6.98	0.24	700	100	62	2.25	0.19	700	25
20	5.48	0.27	700		63	4.48	0.23	700	25
21	4.37	0.19	700	100	64	0.17	0.69	700	
22	1.59	0.16	700	100	65	2.15	0.34	700	25
23	8.84	0.19	700	100	66	7.57	0.66	700	
24	1.36	0.19	700	100	67	1.62	0.29	700	25
25	4.66	1.25	700		68	6.58	0.59	700	25
26	9.56	0.75	910		69	3.36	1.07	700	
27	19.8	0.42	910		70	3.05	2.27	600	
28	0.22	0.21	700		71	7.94	2.39	600	4
29	0.29	0.15	700		72	3.04	0.18	700	10
30	0.24	0.26	700		73	1.27	0.19	700	
31	0.46	0.23	700		74	0.91	0.53	700	
32	1.25	0.20	700		75	2.30	0.47	700	
33	1.70	0.16	700	10	76	0.40	0.45	700	
34	2.11	0.19	700		77	1.38	0.34	700	
35	2.50	0.18	700		78	4.79	0.35	700	
36	0.63	0.18	700	10	79	3.48	0.22	700	37.5
37	3.12	0.23	700		80	7.54	0.23	700	37.5
38	3.58	0.24	700		81	7.65	0.21	700	37.5
39	2.55	0.19	700	25	82	11.4	0.18	700	37.5
40	4.34	0.18	700	25	83	7.06	0.23	700	37.5
41	2.77	0.27	700		84	1.59	0.25	700	37.5
42	22.9	0.39	700		85	0.8	0.22	700	
43	1.3	0.21	700	50	86	1.0	0.40	700	

Table 2: Parameters of the artery network shown in Fig. 2: k is the index of the vessel, l is the length, d is the diameter, c is the stiffness ((3)), R is the resistance (7). Veins are considered to have the same structure with c_k lowered by 20% and dubled d.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PERSONALIZED COMPUTATION OF FRACTIONAL FLOW RESERVE IN CASE OF TWO CONSECUTIVE STENOSES

Yuri Vassilevski[1], Timur Gamilov[2] and Philip Kopylov[3]

[1]Institute of Numerical Mathematics RAS
Gubkina 8, Moscow, Russia, and
Moscow Institute of Physics and Technology
Instituskii Lane 9, Dolgoprudny, Russia
e-mail: yuri.vassilevski@gmail.com
[2] Moscow Institute of Physics and Technology
Instituskii Lane 9, Dolgoprudny, Russia, and
Institute of Numerical Mathematics RAS
Gubkina 8, Moscow, Russia e-mail: gamilov@crec.mipt.ru
[3] I.M. Sechenov First Moscow State Medical University
Trubetskaya ul., 8/2, Moscow, and
Institute of Numerical Mathematics RAS
Gubkina 8, Moscow, Russia

Keywords: Coronary vessels, Fractional Flow Reserve, Stenosis, Bifurcation, Wall-state equation

Abstract. *This paper addresses the problem of personalized computation of fractional flow reserve (FFR). FFR is considered to be the golden standard for making decision on surgical treatment of coronary vessels with multiple stenoses. Computer simulations could simplify the process by non-invasive estimation of FFR based on one-dimensional blood flow model. In this work a problem of two consecutive stenoses is studied. It is shown that in some cases only one of the stenoses has to be to removed, while the other one is insignificant. The structure of vessels is based on patient CT scans. An emphasis is made on the boundary conditions in bifurcation points of the vessels.*

1 Introduction

Heart diseases are the leading causes of sudden deaths in developed countries. One of the main reasons of heart failures are coronary stenoses. They can cause myocard ischemia which frequently results in disability or death. The main treatment of severe coronary stenosis is invasive endovascular intervention, i.e. stenting. Some stenoses can be treated with drugs without surgical intervention. Decision on the type of treatment is based on the estimate of haemodynamic importance of the lesion.

The modern golden standard for making decision on the type of treatment is the fractional flow reserve (FFR) [1, 2, 3]. FFR is defined as the ratio of average blood pressure distal to stenosis to average blood pressure in aorta under conditions of vasodilator administration [1]. Stenosis with FFR below 0.7-0.8 is considered to be severe and should be treated with endovascular surgical intervention. Stenosis with FFR above 0.8 can be treated with drug therapy. The FFR based assessment of stenosis helps to reduce the number of expensive operations as well as the number of incidences which caused disability or death [4].

Modern methods of the numerical FFR estimation involve 3D blood flow modelling in the local region of the studied vessel [1, 2]. It requires substantial computational resources but gives a detailed picture of blood flow. Another approach is based on 1D blood flow modelling of coronary region [5, 6]. One-dimensional approach allows to simulate substantial part of coronary region with several stenoses in different locations.

Cases with several stenoses complicate the FFR estimation [8, 9]. In some cases all lesions should be treated surgically while in others certain lesions can be ignored. 1D models allow to run a series of simulations for each surgical strategy and estimate its haemodynamical impact in a short time. In this work we demonstrate how 1D simulations can be used to study the interaction between two consecutive stenoses. The structure of computational domain was obtained from patient's CT-scans [10].

2 Methods

2.1 Blood flow model

The blood flow model used to estimate FFR considers unsteady viscous incompressible fluid flow through a 1D network of elastic tubes [6, 11]. The patient-specific network of coronary arteries was constructed on the basis of CT images [10]. In this section a brief description of the model is presented, for details we refer to [6, 11]. The flow in every vessel is described by mass and momentum balances

$$\partial A_k/\partial t + \partial(A_k u_k)/\partial x = 0, \tag{1}$$

$$\partial u_k/\partial t + \partial\left(u_k^2/2 + p_k/\rho\right)/\partial x = f_{fr}(A_k, u_k)\,, \tag{2}$$

where k is the index of the vessel; t is the time; x is the distance along the vessel counted from the vessel junction point; ρ is the blood density (constant); $A_k(t,x)$ is the vessel cross-section area; p_k is the blood pressure; $u_k(t,x)$ is the linear velocity averaged over the cross-section; f_{tr} is the friction force. The relationship between pressure and cross-section is defined by the wall-state equation:

$$p_k(A_k) - p_{*k} = \rho_w c_k^2 f(A_k)\,, \tag{3}$$

where ρ_w is the vessel wall density (constant); p_{*k} is the pressure in tissues surrounding the vessel; c_k defines elastic properties of the wall and can be considered as the speed of small

disturbances propagation [12]; $f(A_k)$ is defined by

$$f(A_k) = \begin{cases} \exp\left(A_k/A_{0k} - 1\right) - 1, & A_k/A_{0k} > 1 \\ \ln A_k/A_{0k}, & A_k/A_{0k} \leqslant 1, \end{cases} \tag{4}$$

A_{0k} is the unstressed vessel cross-sectional area.

At the entry point of the aorta the blood flow is assigned

$$u(t,0)\, A(t,0) = Q_H\left(t\right). \tag{5}$$

Here function $Q_H(t)$ corresponds to the heart rate value of 1 Hz and stroke volume of 65 ml [13].

At the bifurcation points the Poiseuille's pressure drop condition is used

$$p_k\left(A_k\left(t,\tilde{x}_k\right)\right) - p^l_{node}\left(t\right) = \varepsilon_k R^l_k A_k\left(t,\tilde{x}_k\right) u_k\left(t,\tilde{x}_k\right), k = k_1, k_2, \ldots, k_M, \tag{6}$$

where R_k represents the hydraulic resistance of *k*-th vessel. Networks of veins and arteries are connected by similar pressure drop conditions.

At the terminal point of the venous system the pressure $p_H = 8mmHg$ is set as the boundary condition. The network of veins is considered to have the same structure as the network of arteries with doubled diameters and 20%-lowered c_k. To close the system, we add the mass conservation condition and second-order compatibility conditions of hyperbolic set (1),(2) (see [11]).

Myocardial compression during systolic phase is the essential part of coronary haemodynamics. The majority of perfusion occurs during the heart diastole [13]. Wall-state equation (3) is modified by setting $p_* = P^{cor}_{ext}\left(t\right)$. The shape of the function $P^{cor}_{ext}\left(t\right)$ is presented in [6] and it is similar to the pressure in the ventricle. The amplitude of P^{cor}_{ext} is 120 $mmHg$ and 30 $mmHg$ for terminal vessels of left and right coronary artery, respectively. Increased resistance of micro-circulation region during systole is simulated by multiplying R_k in (6) for all coronary vessels by the factor of 3 [14].

FFR is calculated as the ratio of average pressure in coronary artery distal to stenosis ($\overline{P}_{dist}$) to average aortic pressure ($\overline{P}_{aortic}$, aorta is vessel 1 in Fig. 1) during vasodilator administration:

$$FFR = \frac{\overline{P}_{dist}}{\overline{P}_{aortic}}. \tag{7}$$

Vasodilator administration is simulated by doubling S_0 in the studied vessel and decreasing resistance R by the factor of 5. This method provided good agreement with experimental results according to [6].

2.2 Compatibility conditions

Hyperbolic equations (1), (2) can be solved by different numerical methods. In this work we use the second-order grid-characteristic method [15]. The method requires the second-order approximation of compatibility conditions. A brief derivation of the second-order compatibility conditions is presented here. For more information we refer to [11].

Compatibility conditions of the hyperbolic set (1), (2) can be derived from characteristic form of (1) and (2):

$$\boldsymbol{\omega}_{ki} \cdot \left(\partial \mathbf{V}_k/\partial t + \partial \mathbf{F}_k/\partial x\right) = \boldsymbol{\omega}_{ki} \cdot \left(\partial \mathbf{V}_k/\partial t + \lambda_{ki}\partial \mathbf{V}_k/\partial x\right) = \boldsymbol{\omega}_{ki} \cdot \mathbf{g}_k, \quad i = 1,2 \tag{8}$$

where λ_{ki} are the eigenvalues of the Jacobi matrix $\mathbf{A}_k = \partial \mathbf{F}_k/\partial \mathbf{V}_k$; $\boldsymbol{\omega}_{ki}$ are the eigenvectors of $\mathbf{A}_k$; $\mathbf{V}_k = (A_k, u_k)^T$; $\mathbf{F}_k = (A_k u_k, u_k^2/2 + p_k/\rho)^T$; $\mathbf{g}_k = (0, f_{fr})^T$.

After the finite differences discretization we can find linear dependence between the linear velocity $u_k(t_{n+1}, \tilde{x}_k)$ and the cross section area $A_k(t_{n+1}, \tilde{x}_k)$ at the time step t_{n+1} at the end or at the beginning of every vessel composing a junction node

$$u_k(t_{n+1}, \tilde{x}_k) = \alpha_k A_k(t_{n+1}, \tilde{x}_k) + \beta_k. \tag{9}$$

For each vessel we use an uniform 1D mesh composed of J edges:

$$M = \{(x_j, t_n) : x_j = hj, j = \overline{0, J}; t_n = \sum_{p=1}^{n} \tau_p\}$$

where h is the mesh size; τ_p is the p-th time step; L is the length of the vessel. For the beginning of the vessel outgoing from the junction node we use the second-order approximation of spatial derivative at time step t_{n+1}. Approximation of equation (8) ($i = 1$) is written at the point $(0, t_{n+1})$ as follows:

$$\left(\frac{\partial \mathbf{V}}{\partial x}\right)_{0,t_n} \approx \frac{-3\mathbf{V}_0^{n+1} + 4\mathbf{V}_1^{n+1} - \mathbf{V}_2^{n+1}}{2h},$$

$$\left(\frac{\partial \mathbf{V}}{\partial t}\right)_{0,t_n} \approx \frac{\mathbf{V}_0^{n+1} - \mathbf{V}_0^{n}}{\tau^{n+1}},$$

$$(\boldsymbol{\omega}_i)_{0,t_{n+1}} \approx (\boldsymbol{\omega}_i)_0^n, \quad (\lambda_i)_{0,t_{n+1}} \approx (\lambda_i)_0^n.$$

Denoting $w_0^n = \left(\sqrt{\frac{1}{\rho A}\left(\frac{\partial p}{\partial A}\right)}\right)_0^n$, $\mathbf{W}_0^n = \{w_0^n, (-1)^i\}$, $\sigma_0^n = \frac{\tau_{n+1}}{h}(\lambda_i)_0^n$ we obtain:

$$\mathbf{W}_0^n \cdot \left(\frac{\mathbf{V}_0^{n+1}) - \mathbf{V}_0^n}{\tau^{n+1}} + (\lambda_i)_0^n \frac{-3\mathbf{V}_0^{n+1} + 4\mathbf{V}_1^{n+1} - \mathbf{V}_2^{n+1}}{2h}\right) = \mathbf{W}_0^n \cdot \mathbf{g}_0^{n+1}.$$

After substitution $\sigma_0^n = \frac{\tau_{n+1}}{h}(\lambda_i)_0^n$ and taking the dot-product

$$\begin{aligned} &w_0^n[S_0^{n+1} - S_0^n + \sigma_0^n(-\frac{3}{2}S_0^{n+1} + 2S_1^{n+1} - \frac{1}{2}S_2^{n+1})] - \\ &-[u_0^{n+1} - u_0^n + \sigma_0^n(-\frac{3}{2}u_0^{n+1} + 2u_1^{n+1} - \frac{1}{2}u_2^{n+1})] = \tau^{n+1}(w_0^n\phi_0^{n+1} - \psi_0^{n+1}) \end{aligned}$$

we find α and β from (9) at x_0

$$\begin{aligned} &\alpha = w_0^n, \quad \beta = [w_0^n(\sigma_0^n(2S_1^{n+1} - \frac{1}{2}S_2^{n+1}) - S_0^n) - \\ &- (\sigma_0^n(2u_1^{n+1} - \frac{1}{2}u_2^{n+1}) - u_0^n) - \tau^{n+1}(w_0^n\phi_0^{n+1} - \psi_0^{n+1})]/(1 - \frac{3}{2}\sigma_0^n). \end{aligned} \tag{10}$$

In the similar way we derive α and β from (9) at x_J:

$$\begin{aligned} &\alpha = -w_J^n, \quad \beta = [-w_J^n(\sigma_J^n(\frac{1}{2}S_{J-2}^{n+1} - 2S_{J-1}^{n+1}) - S_J^n) - \\ &- (\sigma_J^n(\frac{1}{2}u_{J-2}^{n+1} - 2u_{J-1}^{n+1}) - u_J^n) + \tau^{n+1}(w_J^n\phi_J^{n+1} + \psi_J^{n+1})]/(1 + \frac{3}{2}\sigma_J^n). \end{aligned} \tag{11}$$

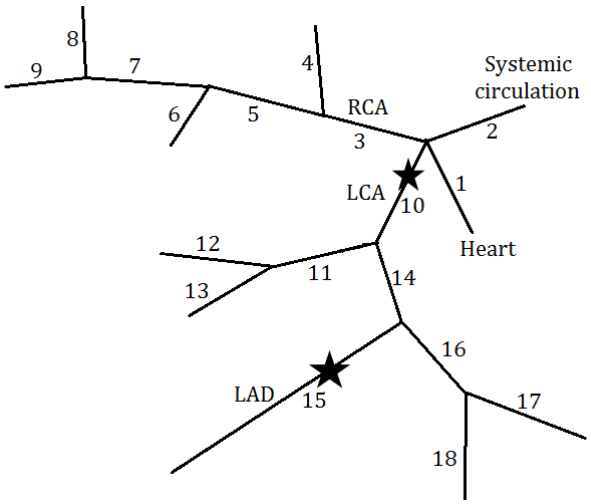


Figure 1: The structure of reconstructed arterial part of anonymous patient-specific data set. Stars designate stenoses. Parameters of the vessels are presented in Tab. 1. LCA is the left main coronary artery, LAD is the left anterior descending artery, RCA is the right main coronary artery

k	l_k, cm	d_k, mm	$c_k, \frac{cm}{s}$	$R_k, \frac{ba \cdot s}{cm^3}$	k	l_k, cm	d_k, mm	$c_k, \frac{cm}{s}$	$R_k, \frac{ba \cdot s}{cm^3}$
1	5.28	21.7	1050	20	10	0.59	3.6	950	720
2	60.0	25.1	840	20	11	6.1	3.0	950	720
3	2.72	3.1	1200	7200	12	2.05	1.17	950	720
4	1.44	1.31	1200	7200	13	1.75	1.21	950	720
5	1.40	2.73	1200	7200	14	1.39	3.8	950	720
6	6.75	1.52	1200	7200	15	12.1	2.05	950	720
7	5.01	2.50	1200	7200	16	5.4	1.91	950	720
8	1.27	1.19	1200	7200	17	0.38	1.01	950	720
9	5.65	0.157	1200	7200	18	2.62	1.19	950	720

Table 1: Parameters of the arterial tree: k is the index of the vessel according to Figure 1, l_k is the length, d_k is the diameter, c_k is the stiffness (3), R_k is the resistance (6). Veins are considered to have the same structure with c_k lowered by 20 %, d_k doubled.

2.3 Patient specific 1D coronary network

The 1D vascular network was generated on the basis of patient-specific data. Geometry of coronary vessels was extracted from CT scans with the help of image processing and segmentation algorithms described in [7]. The network is presented in Fig. 1.

In this work two consecutive stenoses are evaluated. The stenosis in LCA is one third of the vessel's length, the stenosis in LAD is a half of the vessel's length. Note that the patient had other stenoses which were removed from the model in order to investigate the interaction of two consecutive stenoses. Each stenosis was modelled by separating the hosting vessel into three parts: the stenosed part, the proximal part and the distal part [6]. The parameters of the proximal and distal parts correspond to the parameters of the hosting non-stenosed vessel. The parameters of the stenosed part were modified: $A_{0_{stenosed}} = (1-\gamma)^2 A_0$, $R_{stenosed} = \frac{L_{sten}}{L(1-\gamma)^4} R$, where γ is the stenosis fraction (degree of stenosis), R is the resistance of the hosting vessel, L is the length of the hosting vessel, L_{sten} is the length of the stenosed part of the vessel.

3 Results

Three series of simulations were performed for three degrees of stenosis in LCA: no stenosis 0%; slight stenosis 25%; severe stenosis 90%. In each series FFR in LAD was calculated, results are shown in Fig. 2. FFR in LCA was also calculated and turned to be independent of the degree of stenosis in LAD. In addition, we computed FFR in LCA for the following degrees of LCA stenosis: 0% gives FFR=0.96, 25% results in FFR=0.88, 90% results in FFR=0.5.

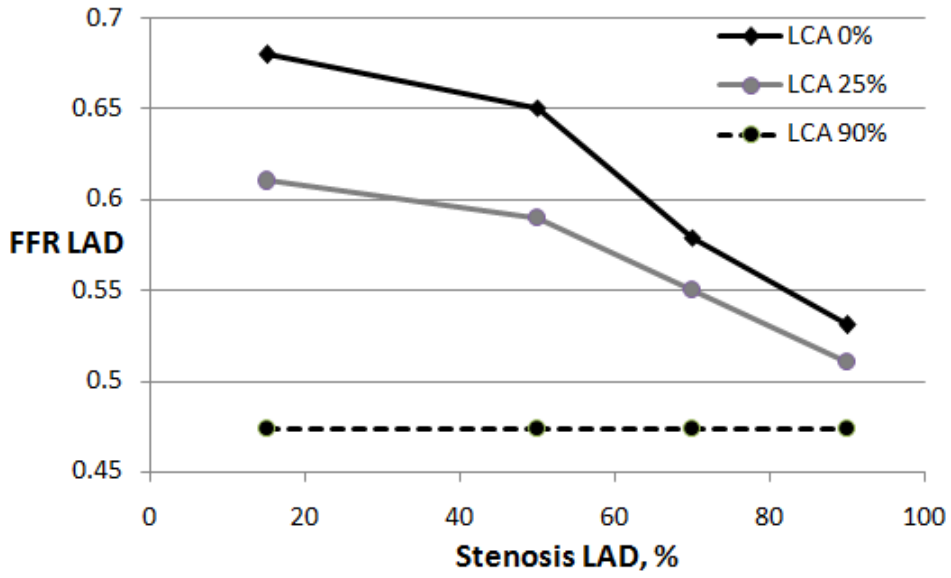


Figure 2: FFR in LAD for various stenosis degrees in LCA.

According to Fig. 2, FFR sensitivity to the degree of the distal stenosis depends on the degree of proximal stenosis. In the case of severe stenosis in LCA, the estimated FFR in LAD is independent of LAD degree of stenosis.

4 Conclusions

The results allow us to make a few conclusions. Firstly, longer lesions (LAD on Fig. 1) have significant impact on haemodynamics even when the degree of the stenosis is low. Secondly, the proximal stenosis has significant impact on FFR of the distal stenosis. The severe proximal stenosis nullifies blood flow to coronary arteries. As a result, lumen of distal arteries has little effect on haemodynamics: it can be very small and still be wide enough for a small amount of blood to flow through. This can be seen in Fig. 1 when the stenosis in LCA is 90%: FFR in LAD is constant since the distal stenosis does not affect the blood flow distribution. Therefore, FFR measurements for distal stenoses in presence of severe proximal lesion can be deceiving and do not allow to assess their impact on haemodynamics.

It should be noted that our method of virtual FFR estimation does not provide FFR=1.0 in case of 0% stenosis. The reason is boundary conditions (6) at the junction nodes that imply pressure drops in bifurcations. In order to achieve FFR=1.0 for healthy vessels, these conditions should be modified.

Cases with multiple stenoses are very hard to analyze without simulations. The presented method can be used for analysis of complicated cases with 3 or more lesions. The method

allows to study possible outcomes and choose the most cost efficient strategy of treatment of multiple stenoses.

Acknowledgement. The research was supported by Russian Science Foundation (RSF) grant 14-31-00024.

REFERENCES

[1] P.D. Morris, D. Ryan,A.C. Morton, R. Lycett, P.V. Lawford, D.R. Hose, J.P. Gunn. Virtual fractional flow reserve from coronary angiography: Modeling the significance of coronary lesions. Results from the VIRTU-1 (VIRTUal fractional flow reserve from coronary angiography) study, *JACC: Cardiovascular Interventions*, **62**, 149–157, 2013.

[2] E.S. Bernad, S.I. Bernad, M.L. Craina. Hemodynamic parameters measurements to assess the severity of serial lesions in patient specific right coronary artery, *Bio-Medical Materials and Engineering*, **24(1)**, 323–334, 2014.

[3] Ph. Kopylov, A. Bykova, Yu. Vassilevski, S. Simakov. Role of measurement of fractional flow reserve (FFR) in coronary artery atherosclerosis, *Therapeutic archive*, **87(9)**, 106–113, 2015.

[4] C.K. Zarins, C.A. Taylor, J.K. Min. Computed fractional flow Reserve (FFTCT) derived from coronary CT angiography, *Journal of Cardiovascular Translational Research*, **6(5)**, 708–714, 2013.

[5] E. Boileau, P. Nithiarasu. One-dimensional modelling of the coronary circulation. Application to noninvasive quantification of fractional flow reserve (FFR), *Computational and Experimental Biomedical Sciences: Methods and Applications*, **21**, 137–155, 2015.

[6] T.M. Gamilov, P.Y. Kopylov, R.A. Pryamonosov, S.S. Simakov. Virtual fractional flow reserve assessment in patient-specific coronary networks by 1D hemodynamic model, *RJNAMM*, **30(5)**, 269–276, 2015.

[7] A.A. Danilov, Yu.A. Ivanov, R.A. Pryamonosov, Yu.V. Vassilevski. Methods of graph network reconstruction in personalized medicine. *Int.J.Numer.Meth.Biomed.Engng.*, e02754, 2016.

[8] S.E. Razavi, R. Zanbouri, O. Arjmandi-Tash. Simulation of blood flow coronary artery with consecutive stenosis and coronary-coronary bypass, *Bioimpacts*, **1(2)**, 99–104, 2011.

[9] N. Saito, H. Matsuo, Y. Kawase, S. Watanabe, B. Bao, E. Yamamoto, H. Watanabe, K. Nakatsuma, K. Ueno, T. Kimura. In vitro assessment of mathematically-derived fractional flow reserve in coronary lesions with more than two sequential stenoses, *Journal of Invasive Cardiology*, **25 (12)**, 642–649, 2013.

[10] Yu.V. Vassilevski, A.A. Danilov, T.M. Gamilov, Yu.A. Ivanov, R.A. Pryamonosov, S.S. Simakov. Patient-specific anatomical models in human physiology, *RJNAMM*, **30(3)**, 185–201, 2015.

[11] S.S. Simakov, T.M. Gamilov TM, Y.N. Soe. Computational study of blood flow in lower extremities under intense physical load. *RJNAMM*, **28(5)**, 485–504, 2013.

[12] I. B. Wilkinson, J. R. Cockcroft, D. J. Webb. Pulse wave analysis and arterial stiffness *J. Cardiovasc. Pharmacol.*, **32, Suppl.3**, S33–7, 1998.

[13] W.F. Ganong. *Review of Medical Physiology*, Stamford, CT, Appleton and Lange, 1999.

[14] M.A. Vis, P.H. Bovendeerd, P. Sipkema, N. Westerhof. Effect of ventricular contraction, pressure, and wall stretch on vessels at different locations in the wall, *Am J Physiol Heart Circ Physiol*, **272**, H2963–H2975, 1997.

[15] K.M. Magometov, A.S. Kholodov. *Grid Characteristic Numerical Methods*. Nauka, Moscow, 1988 (In Russian).

[16] O.C. Zienkiewicz, R.C. Taylor. *The finite element method, Vol. I, 4th Edition*. McGraw Hill, 1989.

[17] J.T. Oden, T. Belytschko, I. Babuska, T.J.R. Hughes. Research directions in computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, **192**, 913–922, 2003.

[18] J.H. Argyris, M. Papadrakakis, L. Karapitta. Elastoplastic analysis of shells with the triangular element TRIC. M. Papadrakakis, A. Samartin, E. Oñate eds. *4th International Colloquium on Computation of Shell and Spatial Structures (IASS-IACM 2000)*, Chania, Crete, Greece, June 4-7, 2000.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

BLOOD FLOW IN THE COMMON CAROTID ARTERY WITH STENOSIS

Helena A.M. Henriques[1], Luísa C. Sousa[2] , Catarina F. Castro[2], , Carlos C. António[2], Rosa Santos[3], Pedro Castro[3] e Elsa Azevedo[3]

[1]INEGI, Instituto de Ciência e Inovação Em Engenharia Mecânica e Engenharia Industrial
Campus da FEUP, Rua Dr. Roberto Frias, 400, 4200-465 Porto, Portugal
andreina-monteiro@hotmail.com

[2]INEGI and DEMec, Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
lcsousa@fe.up.pt, ccastro@fe.up.pt, cantonio@fe.up.pt

[3]Departamento de Neurologia, Hospital de S. João e Faculdade de Medicina da Universidade do Porto
Rua Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
rosampsantos2@gmail.com, pedromacc@gmail.com, elsaazevedo1@gmail.com

Keywords: Ultrasound, Carotid Bifurcation, Segmentation, Hemodynamics.

Abstract. *This work presents the development of a computational methodology based on ultrasound (US) data and able to simulate the hemodynamics of a carotid bifurcation with stenosis. A semi-automatic lumen segmentation methodology of longitudinal and cross-sectional images is presented. The algorithm, implemented in Matlab is based on the hypoecogenic characteristic of the lumen and allows the lumen contour extraction in 2D B-mode images. Patient-specific Womersley velocity profiles, derived from the pulsatile velocity waveforms obtained by pulsed Doppler images, were considered for the definition of the boundary conditions. An image based computational fluid dynamics model of a middle-aged patient was reconstructed from both segmented transversal and longitudinal US images. The blood flow simulation was performed using the Ansys/Fluent software. Pulsatile conditions of the hemodynamic analysis were validated by the carotid US scan. Hemodynamic analysis was performed based on wall shear stress (WSS) descriptors. High values of time-averaged wall shear stress (TAWSS) were captured within stenosis. Low values of TAWSS where found at the carotid bulb, upstream and downstream stenosis and also at external carotid artery. In these regions the oscillatory shear index (OSI) and the relative residence time (RRT) present high values, in agreement with the fact these hemodynamic descriptors are able to capture abnormal flow conditions witch present an import role in the local development of atherosclerotic plaques. This study shows to be clinically useful in the diagnosis and management of the treatment of carotid stenosis as it is able to show a complex hemodynamic behaviour during the cardiac cycle allowing the correlation between the carotid wall pathophysiology and the local influence in artery hemodynamics.*

1 INTRODUCTION

Atherosclerosis is a degenerative disease of the arteries resulting in plaques that can cause stenosis, embolization and thrombosis. In the past decades, several studies based on computational fluid dynamics (CFD) showed that hemodynamic factors are associated with the development and progression of the disease, determining the local distribution of plaques [1-4] However, as hemodynamic phenomena is not yet fully understood, detailed study of disturbed flow during the cardiac cycle may give additional insight to understanding the progression of atherosclerosis being useful in clinical diagnosis.

Due to the anatomy of the head neck region, the use of sonography for diagnostic is integrated in daily clinical routine; it is a fast and inexpensive technique, extremely useful in the initial evaluation of symptomatic patients and enables the acquisition of the carotid artery bifurcation geometry. Automatic segmentation of US images is a complex work due to poor imaging quality as low contrast, speckle, echo shadows and artifacts. Segmentation methods based on active contours, and parametric snakes are not the best choice for an automatic and accurate segmentation of the carotid artery wall; they are based on intensity gradients and require manual initialization, with the design of the initial contour [5, 6].

Previous works on patient-specific hemodynamic combining manual or semi-automatic image segmentation with finite element blood flow simulation have been performed [7, 8]. Like those studies this research presents a patient-specific study, considering a three-dimensional (3D) model of the bifurcation wall reconstructed from segmented B-mode US and blood flow velocity spectral waveforms obtained from PW Doppler images. Carotid artery bifurcation reconstruction based on a set of segmented longitudinal and transversal US images was considered as input geometry of the 3D model for computational fluid dynamics. The assumptions of an incompressible and Newtonian fluid, rigid vascular wall and non-slip condition were considered. Pulsatile inlet boundary conditions obtained from Doppler US measurements are applied at the entrance of the common carotid artery (CCA). This study presents a new segmentation methodology based on pixel intensity distribution and region growing and hemodynamic simulation is performed using ANSYS Fluent 14.5 package.

2 METHODS

The patient-specific hemodynamic study presented is done using Doppler US imaging data; five sequential steps are required [7, 8]: data acquisition, lumen segmentation of longitudinal and transversal carotid images, 3D carotid bifurcation surface reconstruction, mesh generation and hemodynamic simulation.

Data was obtained at São João Hospital Centre, a university hospital in Porto, Portugal. Carotid artery examinations were performed according to a proposed created protocol; this study was approved by the institutional ethical committee and informed consent of each volunteer was obtained.

2.1 Data aquisition

Experimental data necessary for the present study were obtained by the same experienced sonographer (R.S.) dedicated to neurovascular ultrasound at the Neurosonology Unit of the Department of Neurology of São João Hospital Centre [7, 8].

For a middle aged patient of 57 years old man with a high grade stenosis, B-mode and pulsed-wave (PW) Doppler images were collected along the CCA, its bifurcation and proximal segments of internal carotid artery (ICA) and external carotid artery (ECA) including the bifurcation entrance (APEX). For this purpose a commercial colour US scanner (Vivid e; GE, Milwaukee, WI, USA) and a linear array probe (GE 8L-RS) were used. The longitudinal and

transversal set of B-mode images was recorded at end-diastole, to control physiologic variations of diameter within cardiac cycle. Cross-sectional locations were registered at the longitudinal bifurcation image, to allow the correct reconstruction of the CCA bifurcation luminal surface. In order to minimize flow modelling inaccuracies, tracking of the US probe was done by marking positions along the artery bifurcation relying on the ability to manually guide the US probe. During medical examination, ICA stenosis was measured according to European Carotid Surgery Trial (ECST), the percentage of luminal diameter narrowing at the most stenotic region.

Using US PW mode with a 2 mm sample volume, axial flow velocity waveforms were obtained at several specific locations along CCA and distal region of ICA and ECA arteries that could be measured with angle of insonation $\leq 60^{o}$. Considering waveform data from three cardiac cycles Fourier analysis allowed the determination of the first seven harmonics coefficients used to set Womersley profiles.

2.2 Imaging segmentation and carotid bifurcation geometry definition

Segmentation of acquired B-mode images is a complex work due to poor imaging quality as low contrast, speckle, echo shadows and artifacts. Segmentation of the transversal images is more difficult because contour lumen is discontinuous as they present more artifacts caused by speckle noise. A semi-automatic algorithm was implemented using MATLAB software. After converting the colour image to gray levels the selection of the four corners delimiting the rectangular region of interest is performed and the region of interest is selected as shown in figure 1a).

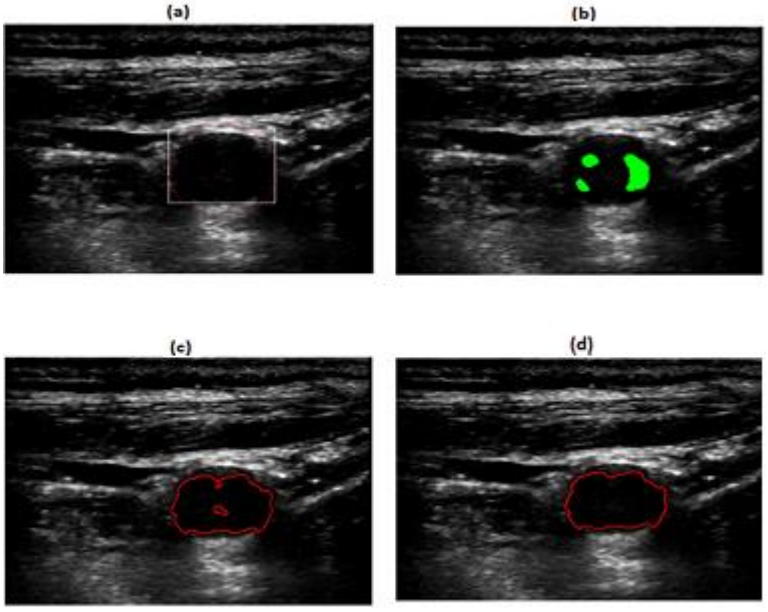


Figure 1: Transversal CCA lumen segmentation: (a) Selection of the region of interest and increased overall image contrast; (b) Identification of lumen central region; (c); (d) Lumen segmentation by region growing.

The identification of the lumen central region is made applying the histogram equalization (CLAHE) and a Gaussian filter smoothing. To identify the central lumen, the seed points for the lumen segmentation algorithm, the method proposed by Molinari et al. (2007) [9], was followed; like in this method, carotid features in an image are addressed as a distribution model of variable intensity along the carotid regions and 2D histograms for mean and stand-

ard deviation were calculated for the establishment of lumen selection criteria. The identified lumen central region is presented in figure 1b).

Region growing segmentation methods are based on grouping sets of neighbouring pixels that meet defined similarity criteria, figures 1c) and 1d). For this study, the considered similarity criterion was taken as the average intensity of a neighbourhood window of 10 × 10 pixels after applying the DSFSRAD filter. An iterative process consisting in checking every pixel just outside the border of the lumen central region was performed; if the homogeneity criterion is validated (intensity is smaller than 0.027) the new pixels will join the newly enlarged lumen central region. The iterative process finishes when enlargement of the region becomes impossible.

After imaging segmentation, 2D smooth lumen contours were stacked in the axial direction according to each image location obtained during data acquisition. The reconstructed lumen surface was smoothed in order to reduce misalignment errors due to patient's involuntary movements during scan.

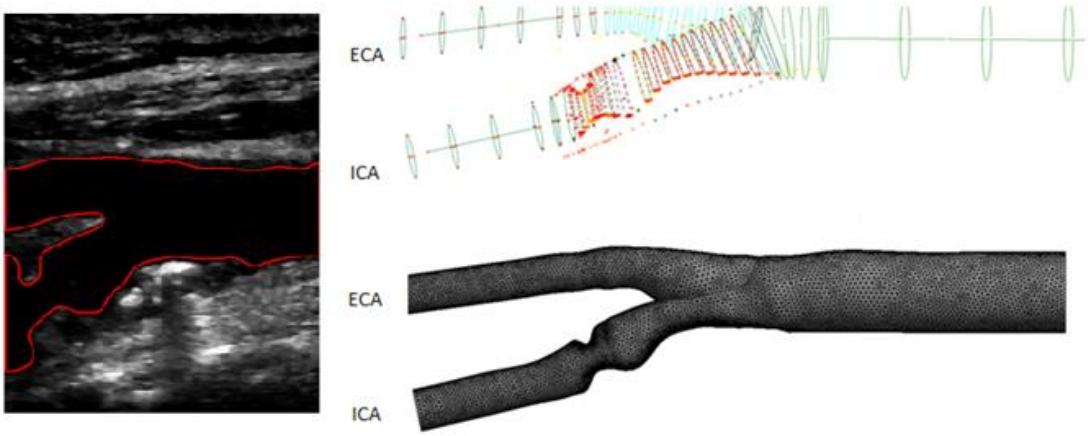


Figure 2: 3D model: segmented longitudinal image, wall surface reconstruction and fluid volume mesh.

The obtained carotid bifurcation surface was not directly usable for generating a suitable computational mesh. Cylindrical flow extensions with a length of four times the local diameters were added at the inlet (before bifurcation) and outlet (after bifurcation) locations, in the direction of the centreline, in order to ensure fully developed velocity profiles at the inlet and to minimize the influence of outlet boundary conditions.

2.3 Meshing and blood flow simulation

Tetrahedral mesh was generated using the software Ansys Workbench (figure 2). The quality of the mesh was tested under steady-state conditions. Using meshes with increasing resolution, the adopted mesh was chosen when a relative error of nodal WSS lower than 5% was reached.

Commercial package Ansys Fluent 14.5 was used to simulate pulsatile blood flow. The first order implicit backward Euler method was considered for the temporal discretization of the Navier-Stokes equations and converged solutions searched using the SIMPLE algorithm and the second-order upwind scheme with a maximum of 40 iterations for each time step. Blood was modelled as a viscous Newtonian fluid [4], incompressible and isotropic with a density of 1060 kg/m3 and a dynamic viscosity of 0.0035 kg/(m.s). At the entrance of the CCA, Womersley velocity profiles derived from pulsatile velocity waveforms extracted from patient PW Doppler images were imposed [7, 8]; a zero wall motion was assumed as in dis-

eased vessels wall motion is further reduced. Outflow boundary conditions were imposed considering a constant flow ratio ICA/ECA = 70%/30%. Three cardiac cycles were simulated considering a constant time step equal to 0.008 seconds. The flow velocities and the WSS vectors from the last cycle were recorded for post-processing.

3 RESULTS AND DISCUSSION

The distribution of velocities at systolic peak is shown in Figure 3. As expected, high velocities were detected at ICA stenosis. Stagnation zones and high velocity gradients can be noticed near the outer bulb wall, opposite to the bifurcation divider wall and downstream stenosis, due to the high variability of the lumen arterial sectional area.

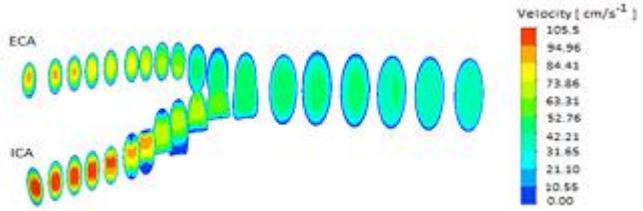


Figure 3: Velocity field at systolic peak (top).

In this work the most widely used WSS-based descriptors, the time-averaged WSS (TAWSS), the oscillating shear index (OSI) and the relative residence time (RRT) were adopted to measure low and oscillating shear stress at the carotid bifurcation [10].

Figure 4 presents the distribution of the wall shear stress based descriptors, shown on the anterior and posterior side of the carotid artery. Low values of TAWSS can be found at ICA origin and upstream and downstream stenosis identifying flow disturbances. Low TAWSS values appeared also on the distal inner wall of ECA. TAWSS values, lower than 0.4 Pa, are associated with intima/media thickening and moderate TAWSS values, larger than 1.5 Pa, induce quiescence and atheroprotective gene expression profile.

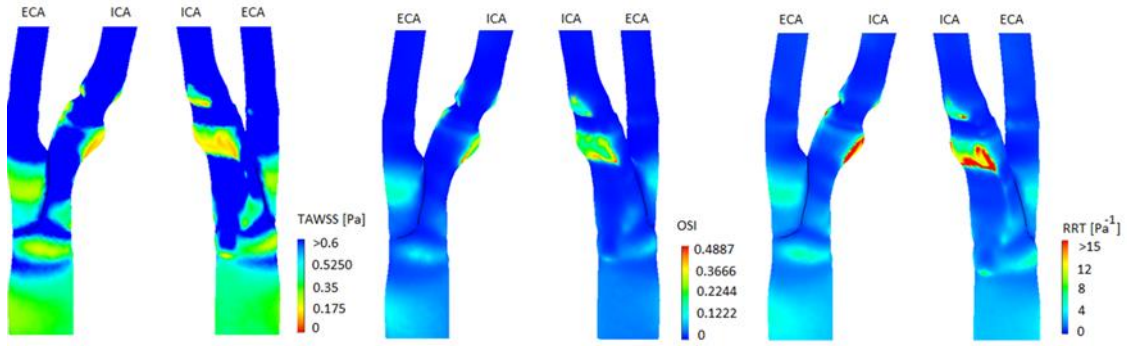


Figure 4: Wall shear stress descriptors: TAWSS (left), OSI (center) and the RRT (right).

In figure 4, OSI and RRT high values regions coincide with TAWSS low values sites; all the hemodynamic indexes captured apparent abnormal flow disturbances at the same sites, ICA origin, upstream and downstream stenosis and in distal inner wall of ECA. Previous studies identify the same regions corresponding to recirculation zones, areas of greater turbulence, and consequently a high shear oscillation [2, 3, 11]. In this work WSS-based descriptors show to be correlated as sites of extremes for TAWSS coincide with sites of extremes of the two other descriptors. Some authors recommended the RRT as a robust single metric of low and oscillatory shear stress [2-4].

4 CONCLUSIONS

This study combined computational 3D geometry reconstruction and simulation of blood flow characteristics in a carotid artery bifurcation based on Doppler images. Using patient-specific carotid artery US data collected in hospital practice, a semi-automatic reconstruction of a carotid bifurcation geometry was presented in this work. A new semiautomatic segmentation algorithm for the lumen contour extraction in US longitudinal 2D B-mode images was presented. This semiautomatic algorithm is robust with respect to factors that degrade this type of images, including appearance of artifacts, lots of speckle noise and occlusions of the lumen caused by plaque region. The algorithm provides a good estimate of the lumen contour when compared with a manual contour and showed to be able to segment the carotid lumen of a stenosed bifurcation.

The simulation demonstrated several complex flow features near the stenosis associated with the diseased vessel geometry. All WSS-based descriptors, assigned highly disturbed flows at the same artery surface regions, as they were able to capture abnormal flow at carina and upstream and downstream ICA stenosis.

Further research will be needed to understand the effect of wall compliance on the blood flow patterns of specific severely diseased carotid arteries. In order to elucidate the role of carotid hemodynamics on plaque development and plaque vulnerability, further larger studies should be done, providing clinically relevant information for risk assessment and surgical planning.

5 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding by FCT, Portugal, of the Research Unit of LAETA-INEGI, Faculdade de Engenharia da Universidade do Porto.

REFERENCES

[1] G. De Santis, M.Conti, B. Trachet, T. De Schryver, M. De Beule, J. Degroote, J. Vierendeels, F. Auricchio, P.Segers, P. Verdonck, B. Verhegghe, Haemodynamic impact of stent-vessel (mal)apposition following carotid artery stenting: mind the gaps! *Computer Methods Biomech Biomed Engin*, **16**(6), 648-59, 2013.

[2] S. Lee, L. Antiga, D. Steinman, Correlations among indicators of disturbed flow at the normal carotid bifurcation. *Journal of Biomechanics Engineering*, **131**(6), 061013, 2009.

[3] U. Morbiducci , D. Gallo, D. Massai, F. Consolo, R. Ponzini, L. Antiga, C. Bignardi, M.A. Deriu, A. Redaelli, Outflow conditions for image-based hemodynamic models of the carotid bifurcation: implication for indicators of abnormal flow. *Journal of Biomechanics Engineering,* **132**(9), 091005, 2010.

[4] U. Morbiducci , D. Gallo, D. Massai, R. Ponzini, H.A. Deriu, L. Antiga, L. Redaelli, F. Montevecchi, On the importance of blood rheology for bulk flow in hemodynamic models of the carotid bifurcation. *Journal of Biomechanics,* **44**(13), 2427–2438, 2011.

[5] C.P. Loizou, C. Theofanous, M. Pantziaris, T. Kasparis, Despeckle filtering software toolbox for ultrasound imaging of the common carotid artery. *Computer Methods and Programs in Biomedicine,* **114**(1), 109-24, 2014.

[6] Z. Ma, R.N. Jorge, T. Mascarenhas, J.M.R.S. Tavares, A Review of Algorithms for Medical Image Segmentation and their Applications to the Female Pelvic Cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, **13**, 235–246, 2010.

[7] L.C. Sousa, C.F. Castro, C.C. António, A. Santos, R. Santos, P. Castro, E. Azevedo, J.M.R.S. Tavares, Towards hemodynamic diagnosis of carotid artery stenosis based on ultrasound image data and computational modelling. *Medical & Biologcal Engineering & Computing*, **52**(11), 971-83, 2014.

[8] L.C. Sousa, C.F. Castro, C.C. António, F. Sousa, R. Santos, P. Castro, E. Azevedo, Computational simulation of carotid stenosis and flow dynamics based on patient ultrasound data - A new tool for risk assessment and surgical planning. *Advances in Medical Sciences*, **61** (1), 32–39, 2016.

[9] F. Molinari, W. Liboni, E. Pavanelli, P. Giustetto, S. Badalamenti, J.S. Suri, Accurate and automatic carotid plaque characterization in contrast enhanced 2D ultrasound images. Conf. Proc. IEEE Eng. Med. Biol. Soc. 335–338 (2007)

[10] D. Ku, D. Giddens, C. Zarins, S. Glagov, Pulsatile flow and atherosclerosis in the human carotid bifurcation—positive correlation between plaque location and low and oscillating shear-stress. *Arteriosclerosis*, **5**(3), 293–302, 1985.

[11] Y.P. Lian, Y. Liu, X. Zhang, Coupling of membrane element with material point method for fluid–membrane interaction problems. *International Journal Mechanics and Materials in Design.***10** (2), 199–211, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A FENICS-HPC FRAMEWORK FOR MULTI-COMPARTMENT BLOCH-TORREY MODELS

D. V. NGUYEN[1], J. JANSSON[1,2], J. HOFFMAN[1]

[1] Computational Technology Laboratory, Department of Computational Science and Technology, School of Computer Science and Communication, KTH Royal Institute of Technology
[2] BCAM - Basque Center for Applied Mathematics, Bilbao Basque Country - Spain
e-mail: {vdnguyen, jjan, jhoffman}@kth.se

Keywords: FEniCS-HPC, diffusion NMR, diffusion MRI, simulation.

Abstract. *In diffusion nuclear magnetic resonance (NMR) and diffusion magnetic resonance imaging (MRI), the multi-compartment Bloch-Torrey equation plays an important role in probing the diffusion characteristics from a nanometer scale to a macroscopic scale. The signal attenuation can be computed by solving the equation. If the volume of interest is composed by multiple compartments, interface conditions with permeability are imposed. Depending on applications, different gradient strengths can be used to capture the signal attenuation. In probing porous media, for instance, high gradient strengths are used. In diffusion MRI, since water molecules enter and exit the computational domain in realistic cases, pseudo-periodic boundary conditions are used. These conditions cause difficulties in solving the equation efficiently and many efforts have been made to develop an efficient numerical method. However, large-scale problems for supercomputers with realistic applications have not been considered yet. We propose a framework for the multi-compartment Bloch-Torrey models based on the FEniCS-HPC platform, a part of the FEniCS project that allows for automated discretization, automated error control with mesh adaptivity and high performance computing. The framework runs on supercomputers with near optimal weak and strong scaling. Our work includes two parts. First, we simplify the multi-compartment Bloch-Torrey model used in diffusion MRI by proposing an approximation to the pseudo-periodic boundary conditions to derive a general form for the interface and boundary conditions. The second part includes implementation and numerical validation of our method on the FEniCS-HPC platform. This simplified model is straightforward to implement and to parallelize and shows promise in validation against more realistic models.*

1 Introduction

In diffusion NMR and diffusion MRI, the signal attenuation S represents the diffusion characteristic of molecular motion through the use of magnetic field gradient pulses and it is computed as the average of the transverse magnetization $M(\mathbf{r}, t)$ over a volume of interest Ω. The signal attenuation is measured at the echo time T by integrating the complex transverse magnetization, i.e

$$S = \int_{\mathbf{r}\in\Omega} M(\mathbf{r}, T)\, d\mathbf{r} \tag{1}$$

The water diffusion is encoded by using time-varying linear spatial magnetic field gradient pulses $f(t)$.

The complex transverse magnetization at a spatial posion $\mathbf{r} = (r_1, \cdots, r_d)$ (here $d = 2, 3$ is the problem dimensions) can be modeled by the Bloch-Torrey equation [15, 19]

$$\frac{\partial}{\partial t} M(\mathbf{r}, t) = -\mathcal{I}\,\gamma f(t)\mathbf{g}\cdot\mathbf{r}\, M(\mathbf{r}, t) + \nabla\cdot\left(\mathbf{D}(\mathbf{r})\nabla M(\mathbf{r}, t)\right), \quad \mathbf{r}\in\Omega, \tag{2}$$

where $\mathcal{I}$ is the complex unit ($\mathcal{I}^2 = -1$), $\gamma = 2.67513\times 10^8\,\mathrm{rad\,s^{-1}T^{-1}}$ is the gyromagnetic ratio of the water proton, $\mathbf{D}(\mathbf{r})$ is the diffusion tensor, and $\mathbf{g} = (g_1, \cdots, g_d)$ is the diffusion gradient including gradient strength and gradient direction.

Eq. (2) needs an initial condition:

$$M(\mathbf{r}, 0) = \rho, \qquad \mathbf{r}\in\Omega \tag{3}$$

where ρ is the initial transverse magnetization in Ω.

The most commonly used temporal profile is called the pulsed-gradient spin echo (PGSE) sequence [18], with two rectangular pulses of duration δ, separated by a time interval $\Delta - \delta$, for which the profile $f(t)$ is

$$f(t) = \begin{cases} 1, & 0 \le t \le \delta, \\ -1, & \Delta < t \le \Delta + \delta, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The signal attenuation is usually computed and plotted against the gradient strength $q = \|\mathbf{g}\|$ or a quantity called the b-value which is computed as

$$b = \gamma^2\|\mathbf{g}\|^2 \int_0^T \left(\int_0^u f(s)\, ds\right) du \tag{5}$$

For the PGSE, it is

$$b = \gamma^2 \|\mathbf{g}\|^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right) \tag{6}$$

If Ω is composed by multiple compartments, i.e $\Omega = \cup^l \Omega^l$, the following interface conditions need to be imposed

$$\mathbf{D}^l(\mathbf{r}) \nabla M^l(\mathbf{r},t) \cdot \mathbf{n}^l = -\mathbf{D}^n(\mathbf{r}) \nabla M^n(\mathbf{r},t) \cdot \mathbf{n}^n = \kappa^{ln} \Big(M^n(\mathbf{r},t) - M^l(\mathbf{r},t) \Big), \tag{7}$$

where $\mathbf{r} \in \Gamma^{ln}, \forall\, l, n$ and $\mathbf{n}^k$ is a normal vector pointing outward Ω^k.

In diffusion MRI, to allow the water molecules to enter and exit the volume of interest $\Omega = \prod_{i=1}^{d} [a_i, b_i]$, the pseudo-periodic boundary conditions are applied

$$\begin{aligned} M(\mathbf{r},t)|_{r_i=a_i} &= M(\mathbf{r},t)|_{r_i=b_i} \exp\Big(\mathcal{I}\, \theta_i(t) \Big), \\ \frac{\partial M(\mathbf{r},t)}{\partial r_i}\bigg|_{r_i=a_i} &= \frac{\partial M(\mathbf{r},t)}{\partial r_i}\bigg|_{r_i=b_i} \exp\Big(\mathcal{I}\, \theta_i(t) \Big), \end{aligned} \tag{8}$$

where

$$\theta_i(t) := \gamma\; g_i \,(b_i - a_i) \int_0^t f(s)\, ds, i = 1, \cdots, d$$

An advantage of using Eq. (8) is significantly reducing the size of computational domains. This is good in terms of saving computer resources but it is challenging to impose and to generate valid meshes.

Solving the full model, i.e Eqs. (2, 7, 8), is challenging and many efforts have been made [1, 10, 13, 14, 16, 20]. However, large-scale problems for supercomputers with realistic applications have not been considered yet.

We propose a framework for solving the multi-compartment Bloch-Torrey equation based on the FEniCS-HPC which is an open source framework for automated solution of PDE on massively parallel architectures, providing automated evaluation of variational forms given a high-level description in mathematical notation, duality-based adaptive error control, implicit parameter-free turbulence modeling by use of stabilized FEM and strong linear scaling up to thousands of cores [4–9]. The FEniCS-HPC is a branch of the FEniCS [2, 11] framework based on components with clearly defined responsibilities:

- FIAT: finite element spaces, basis functions
- FFC+UFL: Automated evaluation of weak forms in mathematical notation.
- DOLFIN-HPC: Automated high performance assembly of weak forms and interface to linear algebra of discrete systems and mesh refinement.

Our work includes two parts. First, we simplify the multi-compartment Bloch-Torrey model by proposing an approximation to the pseudo-periodic boundary conditions (8) to derive a general form for the interface and boundary conditions. In the second part, we implement the approximated model on the FEniCS-HPC platform.

The main contributions of this paper include:

- Proposing a framework on supercomputers for solving the muti-compartment Bloch-Torrey equations applied to diffusion NMR and diffusion MRI.
- Solving the difficulties of mesh generation caused by the pseudo-periodic boundary conditions that enables us to approach more realistic geometries.
- Implementing the method on the FEniCS-HPC platform.
- Validating the correctness and demonstrating the efficiency of the model and the method for some example geometries.

The paper is organized as the follows. First, we introduce the multi-compartment Bloch-Torrey models used in diffusion NMR and diffusion MRI with equations and quantities used in this field. In section 2, we propose an approximation of the pseudo-periodic boundary conditions. Then, we introduce the discretizations and implementations of the approximated model on the FEniCS-HPC platform in section 3. In section 4, we first validate the correctness of the model through two examples: a two-layer infinitely-long cylinder and a two-layer spherical cell. Then, we show speedup ratio and parallel efficiency. Finally, in section 5 we conclude our work and propose future directions.

2 An approximation of the pseudo-periodic boundary conditions

To impose the pseudo-periodic boundary conditions (8) for the r_i-direction, vertices belonging to opposite planes $r_i = a_i$ and $r_i = b_i$ should be aligned, i.e $\mathbf{r}_{a_i} = \mathbf{r}_{b_i} - (b_i - a_i)\, e_i$, here $\mathbf{r}_{a_i}, \mathbf{r}_{b_i}$ are coordinates of vertices belonging to opposite planes $r_i = a_i$ and $r_i = b_i$, and e_i is a unit vector in r_i-direction. However, it is usually challenging to generate a mesh with such the enforcement in realistic cases since the interfaces usually cut the exterior boundaries. We propose an approximation of the pseudo-periodic boundary conditions that allows water molecules to enter and exit the boundaries but does not require vertices belonging to the opposite planes to be aligned.

The approximation comes from a combination of the interface conditions between two compartments and the pseudo-periodic conditions for the exterior boundaries. Since the computational domain Ω is extended periodically, in some cases the cell interfaces touch the exterior boundary $\partial\Omega$. To simplify the explanation, we consider a computational domain $\Omega = [a_1, b_1] \times [a_2, b_2]$ in which a cell is placed touching $r_1 = a_1$ (Fig. 1). The boundary needs to be periodic at $r_1 = a_1$ and $r_1 = b_1$. The cell interior is characterized by a diffusion tensor $\mathbf{D}^c$. The extra-cellular space is the remaining part and is

characterized by a diffusion tensor $\mathbf{D}^e$. The cell touches the boundary at the interface Σ, which is the intersection between the cell boundary and $\partial\Omega$.

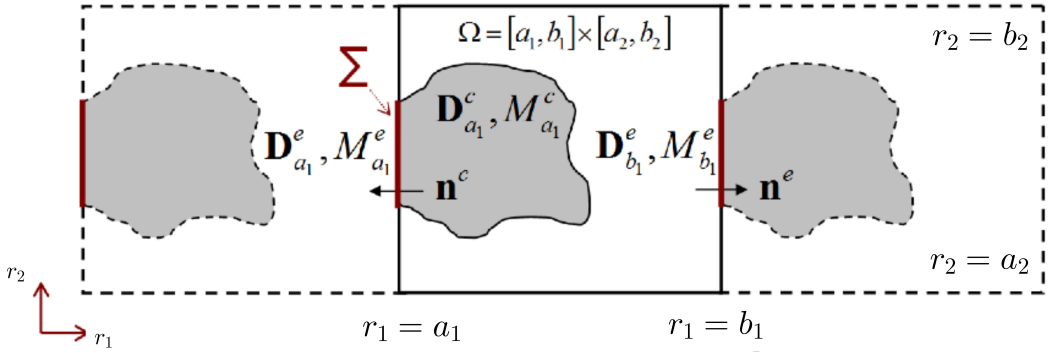


Figure 1: When the cell interface touches $\partial\Omega$, the interface conditions and periodic boundary conditions are combined.

We recall the pseudo-periodic boundary conditions for this specific domain

$$\begin{aligned} M^e_{a_1} &= M^e_{b_1} \exp(\mathcal{I}\,\theta), \\ \mathbf{D}^e_{a_1} \nabla M^e_{a_1} \cdot \mathbf{n}^c &= -\mathbf{D}^e_{b_1} \nabla M^e_{b_1} \cdot \mathbf{n}^e \exp(\mathcal{I}\,\theta). \end{aligned} \tag{9}$$

where

$$\theta = \gamma\, g_1\, (b_1 - a_1) \int_0^t f(s)\, ds,$$

and the interface conditions at Σ are

$$\begin{aligned} \mathbf{D}^c_{a_1} \nabla M^c_{a_1} \cdot \mathbf{n}^c &= \kappa^e (M^e_{a_1} - M^c_{a_1}) \\ \mathbf{D}^e_{a_1} \nabla M^e_{a_1} \cdot \mathbf{n}^e &= \kappa^e (M^c_{a_1} - M^e_{a_1}) \end{aligned} \tag{10}$$

The combination of (9) and (10) with $\mathbf{n}^c = -\mathbf{n}^e$ gives

$$\begin{aligned} \mathbf{D}^c_{a_1} \nabla M^c_{a_1} \cdot \mathbf{n}^c &= \kappa^e \Big(M^e_{b_1} \exp(\mathcal{I}\,\theta) - M^c_{a_1} \Big), \\ \mathbf{D}^e_{b_1} \nabla M^e_{b_1} \cdot \mathbf{n}^e &= \kappa^e \Big(M^c_{a_1} \exp(-\mathcal{I}\,\theta) - M^e_{b_1} \Big). \end{aligned} \tag{11}$$

Theoretically, when $\kappa^e \to +\infty$, Eq. (11) becomes the pseudo-periodic boundary conditions (9). However, we see that practically when $\kappa^e = 1.0\,\mathrm{m/s}$, Eq. (11) is a good approximation of Eq. (9) for free diffusion up to $b = 2000\,\mathrm{s/mm^2}$ corresponding to $S = 2.479\text{e-}3$. If the computed signal is stronger, $\kappa^e = 0.1\,\mathrm{m/s}$ is good enough. Eq. (11) is interesting for the following reasons

- The interface conditions (7) are only a special case of Eq. (11). Their general form is Eq. (12).

$$\begin{aligned}\mathbf{D}^l(\mathbf{r}^l)\nabla M^l(\mathbf{r}^l,t)\cdot\mathbf{n}^l &= \kappa^{ln}\Big(M^n(\mathbf{r}^n,t)\exp(\mathcal{I}\,\theta^{ln}) - M^l(\mathbf{r}^l,t)\Big),\\ \mathbf{D}^n(\mathbf{r}^n)\nabla M^n(\mathbf{r}^n,t)\cdot\mathbf{n}^n &= \kappa^{nl}\Big(M^l(\mathbf{r}^l,t)\exp(\mathcal{I}\,\theta^{nl}) - M^n(\mathbf{r}^n,t)\Big),\end{aligned} \tag{12}$$

  where $\kappa^{ln} = \kappa^{nl}$, $\mathbf{r}^n \in \Gamma^n$, $\mathbf{r}^l \in \Gamma^l$ and

$$\theta^{ln} = -\theta^{nl} = \gamma\ \mathbf{g}\cdot(\mathbf{r}^n - \mathbf{r}^l)\int_0^t f(s)\,ds.$$

  Further on, instead of solving the complete Bloch-Torrey PDE problem involving the PDE (2), the interface conditions Eqs. (7), the two boundary conditions Eqs. (8), we will solve an approximated model consisting of the Bloch-Torrey equation (2) and the general conditions (12).

- If $M^c_{a_1}$ and $M^e_{b_1}$ are known from the previous time step, the exterior boundaries of the computational domain need not to be treated as periodic. It is also applicable to non confirming interfaces. The constraints of periodicity and confirming interfaces which are usually the most challenging issues in generating meshes can be avoided.

- Since Eq. (12) becomes standard Robin boundary conditions, it is straightforward to implement and parallelize.

3 Discretizations and Implementations on the FEniCS-HPC platform

Let V^l be a function space and $v \in V^l$ be a test function, then the weak form of Eq. (2) on a compartment Ω^l is

$$\frac{\partial}{\partial t}\int_{\Omega^l} v\,M^l\,d\mathbf{r} = -\mathcal{I}\gamma f(t)\int_{\Omega^l} v\,\mathbf{g}\cdot\mathbf{r}\,M^l d\mathbf{r} - \int_{\Omega^l}\nabla v\cdot\mathbf{D}\nabla M^l d\mathbf{r} + \int_{\partial\Omega^l\cap\Gamma^{ln}} v\,\mathbf{D}\nabla M^l\cdot\mathbf{n}^l\,ds \tag{13}$$

Combining with Eq. (12), we get

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Omega^l} v\,M^l\,d\mathbf{r} = &-\mathcal{I}\gamma f(t)\int_{\Omega^l} v\,\mathbf{g}\cdot\mathbf{r}\,M^l d\mathbf{r} - \int_{\Omega^l}\nabla v\cdot\mathbf{D}\nabla M^l d\mathbf{r}\\ &+\int_{\partial\Omega^l\cap\Gamma^{ln}} v\,\kappa^{ln}\Big(M^n(\mathbf{r}^n,t)\exp(\mathcal{I}\,\theta^{ln}) - M^l(\mathbf{r}^l,t)\Big)\,ds\end{aligned} \tag{14}$$

Assuming that V^l is spanned by basis functions φ_α, i.e $V^l = span\{\varphi_1, \ldots, \varphi_N\}$, and $\xi^l_\alpha(t) \approx M^l(\mathbf{r}_\alpha, t)$, $M^l(\mathbf{r}, t)$ can be presented in a linear combination of basis function, i.e $M^l(\mathbf{r}, t) = \sum_{k=1}^N \varphi_\alpha \xi^l_\alpha$, where N is the number of degrees of freedom. Eq. (14) takes $v = \varphi_\beta, \beta = 1, \ldots, N$ to give

$$\begin{aligned}\sum_{\alpha=1}^N \int_{\Omega^l} \varphi_\beta\, \varphi_\alpha\, d\mathbf{r} \frac{\partial}{\partial t}\xi^l_\alpha(t) = -\mathcal{I}\gamma f(t) \sum_{\alpha=1}^N \int_{\Omega^l} \varphi_\beta\, \mathbf{g}\cdot\mathbf{r}\, \varphi_\alpha d\mathbf{r}\, \xi^l_\alpha \\ - \sum_{\alpha=1}^N \int_{\Omega^l} \nabla\varphi_\beta \cdot \mathbf{D}\nabla\varphi_\alpha d\mathbf{r}\, \xi^l_\alpha + \kappa^{ln} \sum_{\alpha=1}^N \int_{\partial\Omega^l\cap\Gamma^{ln}} \varphi_\beta\varphi_\alpha\, ds\, \xi^{ln}_\alpha\end{aligned} \tag{15}$$

Here $\xi^{ln}_\alpha \approx \xi^n_\alpha(t)\exp(I\,\theta^{ln}) - \xi^l_\alpha(t) = M^n(\mathbf{r}^n_\alpha, t)\exp(I\,\theta^{ln}) - M^l(\mathbf{r}^l_\alpha, t)$. We note that $M^n(\mathbf{r}^n_\alpha, t)$ comes from $\Gamma^{nl} \subset \Omega^n$ but its values should be projected onto $\Gamma^{ln} \subset \Omega^l$. The spatial position $\mathbf{r}^n_\alpha$ should be computed such that it aligns with $\mathbf{r}^l_\alpha$. We Assume that $\mathbf{r}^l_\alpha = (r^l_i)_{i=1,\ldots,d}$ and $\mathbf{r}^n_\alpha = (r^n_i)_{i=1,\ldots,d}$. If $\mathbf{r}^l_\alpha$ belongs to the interfaces between two compartments, then $\mathbf{r}^n_\alpha = \mathbf{r}^l_\alpha$. If $\mathbf{r}^l_\alpha$ belongs to the plane $r_i = a_i$, i.e $r^l_i = a_i$, then we will set $r^n_i = b_i$ and $r^n_k = r^l_k$ for all $k \neq i$. In reverse, if $\mathbf{r}^l$ belongs to the plane $r_i = b_i$, i.e $r^l_i = b_i$, then we will set $r^n_i = a_i$ and $r^n_k = r^l_k$ for all $k \neq i$. Since $\mathbf{r}^n_\alpha$ is computed with respect to $\mathbf{r}^l_\alpha$, the value of ξ^n_α at $\mathbf{r}^n_\alpha$ may not be available and interpolation will be used.

Let ξ and φ is a column vectors of degrees of freedom and basis functions, i.e $\xi = (\xi_k)$ and $\varphi = (\varphi_k)$, Eq. (15) can be rewritten in a matrix form

$$\mathbf{M}^l \frac{\partial}{\partial t}\xi^l = -\mathcal{I}\gamma f(t)\, \mathbf{J}^l\, \xi^l - \mathbf{S}\, \xi^l + \kappa^{ln}\mathbf{F}^l\, \xi^{ln} \tag{16}$$

Since ξ^l is complex-valued, we can write $\xi^l = \xi^l_{\mathcal{R}} + \mathcal{I}\, \xi^l_{\mathcal{I}}$. Eq. (16) is then decoupled to two equations

$$\begin{aligned}\mathbf{M}^l \frac{\partial}{\partial t}\xi^l_{\mathcal{R}} &= \gamma f(t)\, \mathbf{J}^l\, \xi^l_{\mathcal{I}} - \mathbf{S}\, \xi^l_{\mathcal{R}} + \kappa^{ln}\mathbf{F}^l\, \xi^{ln}_{\mathcal{R}} \\ \mathbf{M}^l \frac{\partial}{\partial t}\xi^l_{\mathcal{I}} &= -\gamma f(t)\, \mathbf{J}^l\, \xi^l_{\mathcal{R}} - \mathbf{S}\, \xi^l_{\mathcal{I}} + \kappa^{ln}\mathbf{F}^l\, \xi^{ln}_{\mathcal{I}}\end{aligned} \tag{17}$$

where

$$\mathbf{M}^l = \int_{\Omega^l} \varphi\, \varphi^T\, d\mathbf{r}, \quad \mathbf{S}^l = \int_{\Omega^l} \nabla\varphi \cdot \nabla\varphi^T\, d\mathbf{r}, \quad \mathbf{J}^l = \int_{\Omega^l} \mathbf{g}\cdot\mathbf{r}\, \varphi\, \varphi^T\, d\,\mathbf{r},$$

$$\mathbf{F}^l = \int_{\partial\Omega^l} \chi\, \varphi\, \varphi^T\, ds, \quad \chi(\mathbf{r}) = \begin{cases} 1 & \mathbf{r} \in \partial\Omega \cap \Gamma^{ln} \\ 0 & \text{otherwise} \end{cases}.$$

Before writing a high-level "solver" in C++ using DOLFIN-HPC, we define the matrices in UFL "form files", closely mapping mathematical notation. The form files are then compiled to low-level C++ source code with FFC.

Matrices $\mathbf{M}$ and $\mathbf{J}$ can be defined in a same form with $f = 1$ for matrix $\mathbf{M}$ and $f = \mathbf{g} \cdot \mathbf{r}$ for matrix $\mathbf{J}$. Here f is a dolfin "Function" defined in the "solver".

```
V = FiniteElement("CG","tetrahedron" , 1)
u = TrialFunction(V)
v = TestFunction(V)
f = Coefficient(V)
a = f*u*v*dx
```

The stiffness matrix $\mathbf{S}$ with a diffusion coefficient D is defined as

```
V = FiniteElement("CG", "tetrahedron", 1)
u = TrialFunction(V)
v = TestFunction(V)
D = Coefficient(V)
a = D*dot(grad(u),grad(v))*dx
```

Finally, the matrix $\mathbf{F}$ with a boundary marker χ is defined as

```
V= FiniteElement("CG", "tetrahedron", 1)
v = TestFunction(V)
chi = Coefficient(V)
L = chi*v*ds
```

For the time discretization, similar to [10, 14] we apply the second-order adaptive Runge-Kutta Chebyshev (RKC) [17]. This method has second order in time and is especially suitable for parabolic PDEs. Since this method is explicit, the right-hand side of Eq. (17) is known and the two equation can be solved separately. With the same reason, each equation in turn is solved independently on each compartment. To adapt the FEniCS-HPC platform, the serial code of the RKC available at `http://www.netlib.org/ode/rkc.f` was rewritten and parallelized.

4 Results

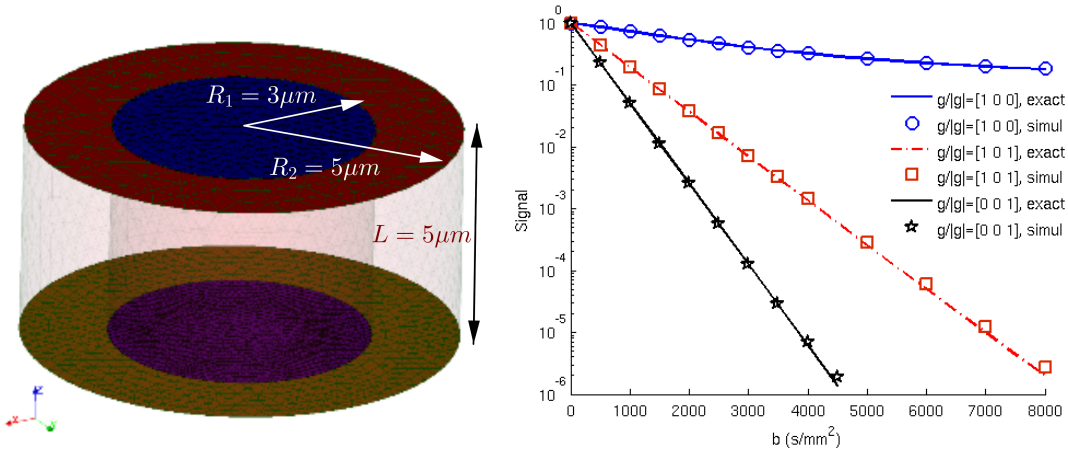
4.1 Validation of the framework

We validate the framework through two examples: a two-layer infinitely-long cylinder and a two-layer sphere. In the first example, we show that the pseudo-periodic boundary conditions can be well approximated by imposing high permeabilities at the exterior boundaries that solves difficulties of mesh generation and well facilitates for parallelization. In the second example, we show that the framework works efficiently in a wide range of gradient strength.

Signals for a two-layer infinitely-long cylinder

In this sample, we compute signals for a cylinder which is composed of two lay-

ers with $R_1 = 3\mu$m and $R_2 = 5\mu$m and $L = 5\mu$m. The diffusion coefficients are $D_1 = D_2 =$ 3e-3 mm^2/s. We allow water molecules to enter and exit the computational domain in the z-direction by imposing $\kappa^e = 0.1$ m/s. So, the signals represent the diffusion in an infinitely-long cylinder. Between two layers, a permeability of κ=5e-5 m/s is imposed and the diffusion time is $\delta = \Delta = 10$ms.
The mesh was generated in such a way that it is totally non conforming and non-periodic. The interior layer is twice as fine as the exterior one. The top base is twice as coarse as the bottom (Fig. 2a).
We performed simulations for three gradient directions $\frac{\mathbf{g}}{|\mathbf{g}|} = [1, 0, 0]$, $\frac{[1,0,1]}{\sqrt{2}}$ and $[0, 0, 1]$. The exact signals are computed by a combination between free diffusion and matrix formalism. In all three gradient directions, we obtain good approximations. For high b-values, the approximations are less good since the signals are too weak (Fig. 2b).



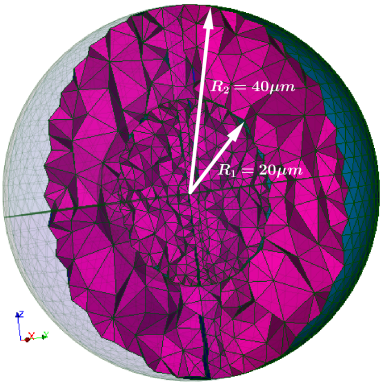
(a) (b)

Figure 2: Signal on a double-layer cylinder with $R_1 = 3\mu$m and $R_2 = 5\mu$m, $L = 5\mu$m. The diffusion coefficients are $D_1 = D_2 =$ 3e-3 mm^2/s. The permeability is κ=5e-5 m/s for the membrane between two layers. The approximation of the pseudo-periodic boundary conditions is applied in z-direction with $\kappa^e = 0.1$ m/s.

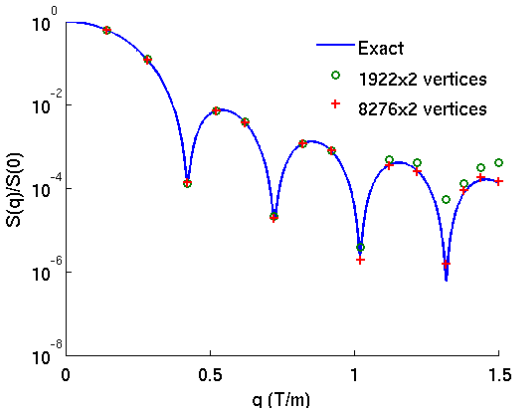
Signals for a two-layer sphere

In this example, we compute signals on a double-layer spherical cell with $R_1 = 20\mu$m and $R_2 = 40\mu$m (Fig. 3a). A wide range of gradient strengths [3, 13] is used. The signals are computed and plotted against b-values and q-values. In Fig. 3b, the gradient strength q is varied between 0 and 1.5 T/m for δ=1e-3s, Δ=1s. Two mesh sizes of 1922 and 8276 vertices are considered. For high gradient strength, finer mesh is needed to give good approximation. When computed against b-values, the gradient strength is

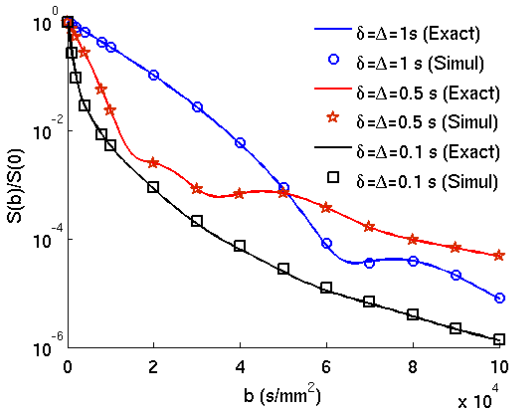
large for short-time diffusion and smaller for long-time diffusion. Three time sequences, $\delta = \Delta = 0.1, 0.5, 1.0$s with b varied between 0 and 1e5 s/mm^2 are considered (Fig. 3c). The largest gradient strength is 0.045 T/m for $\delta = \Delta = 0.1$s and b=1e5 s/mm^2. Numerical signals are compared to the reference ones computed by matrix formalism [3]. The diffusion coefficients are $D_1 = D_2 =$ 3e-3 mm^2/s, the permeability for the membrane between two layers is κ=5e-5 m/s, and $\kappa^e = 0$ is set for exterior boundaries.



(a)



(b)



(c)

Figure 3: Signal on a double-layer spherical cell with $R_1 = 20\mu$m and $R_2 = 40\mu$m. The diffusion coefficients are D_0=D_1=3e-3 mm^2/s, the permeability for the membrane between two layers is κ=5e-5 m/s, and $\kappa^e = 0$ is set for exterior boundaries. The gradient strength is varied between 0 and 1.5 T/m for δ=1e-3s, Δ=1s and two mesh sizes of 1922 and 8276 vertices are considered (a). When computed against b-values, the gradient strength is large for short-time diffusion and smaller for long-time diffusion (b). Three time sequences, $\delta = \Delta = 0.1, 0.5, 1.0$s with b varied between 0 and 1e5 s/mm^2 are considered.

Each b-value is computed in between 5 and 10 minutes using 4 MPI processes for the mesh size of 8276×2 vertices. Timing for each q is less than 2 minutes for the mesh size of 1922×2 and 15 minutes for the mesh size of 8276×2 vertices.

4.2 Speedup ratio and parallel efficiency

We clarify the timing, speedup ratio and parallel efficiency on the Hydra and Beskow computers for a sphere of radius $R = 80\mu$m with $b = 4000$ s/mm^2 and the diffusion time $\delta = \Delta = 100$ms. On the Hydra a moderate-scale problem with a tetrahedral mesh of 124886 vertices and on the Beskow a large-scale problem with 4307475 vertices are considered. The number of MPI processes vary between 1 and 42 on the Hydra and between 4 and 232 on the Beskow.

On the Hydra, the serial computation is about 1.7 hours and the fastest computation with 42 MPI processes is about 4 minutes. On the Beskow, the timing is about 14.7 hours for 4 MPI processes and 20 minutes for 232 MPI processes. See Fig. 4.

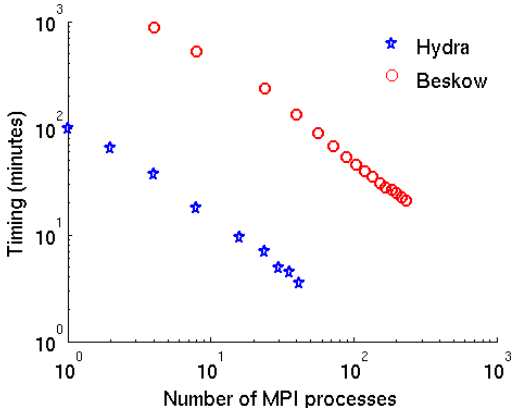


Figure 4: Timing tested on the Hydra and on the Beskow for a sphere of radius $R = 80\mu$m with $b = 4000$ s/mm^2 and the diffusion time $\delta = \Delta = 100$ms. The tetrahedral mesh has 124886 vertices on the Hydra and 4307475 vertices on the Beskow.

The speedup ratio is the ratio between timing for serial execution T_{serial} and timing for parallel execution $T_{parallel}$, i.e

$$\mathcal{S}(p) = \frac{T_{serial}}{T_{parallel}(p)} \tag{18}$$

where p is the number of MPI processes used in the parallel execution.

The ideal speedup is $\mathcal{S}_{ideal}(p) = p$, i.e when p MPI processes are used, the parallel execution will be p times faster than the serial execution.

The parallel efficiency is computed by

$$\mathcal{E} = \frac{\mathcal{S}}{p} \tag{19}$$

In Fig. 5, we show the speedup ratio and parallel efficiency corresponding to the timing in Fig. 4. We obtain a strong scaling on both computers and the parallel efficiency is always higher than 60%.

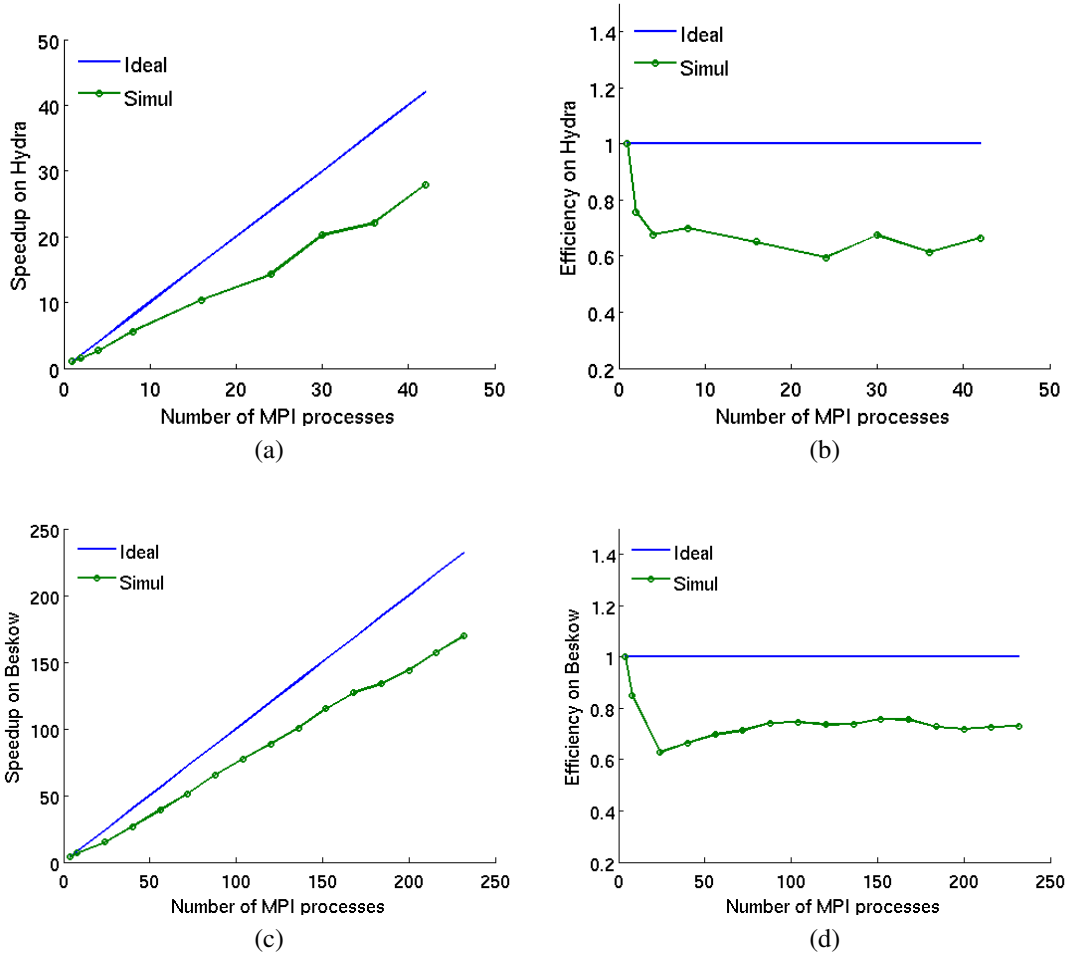


Figure 5: Speedup ratio and parallel efficiency on the Hydra and on the Beskow for a sphere of radius $R = 80\mu$m with $b = 4000\,$s/mm^2 and the diffusion time $\delta = \Delta = 100$ms. The tetrahedral mesh has 124886 vertices on the Hydra and 4307475 vertices on the Beskow. We obtain a strong scaling on both computers and the parallel efficiency is always higher than 60%.

5 Conclusion and future works

We proposed a FEniCS-HPC framework for the multi-compartment Bloch-Torrey models applied for diffusion NMR and diffusion MRI. This framework works efficiently for a wide range of gradient strengths and shows promise in validation against more realistic models. A comparison with other methods will be performed and more realistic applications will be considered in the future.

Acknowledgments

The first author would like to thank Jing-Rebecca Li and Denis Grebenkov for their teachings about the multi-compartment Bloch-Torrey models applied to diffusion NMR and diffusion MRI.

The authors would like to thank Niyazi Cem Degirmenci and Jeannette Spühler for their enthusiasm helps with the FEniCS-HPC platform.

This work has been funded by the Swedish Foundation for Strategic Research, the European Research Council, and the Swedish Research Council. The simulations were performed on resources provided by the SNIC.

References

[1] L. BELTRACHINI, Z. A. TAYLOR, AND A. F. FRANGI, *A parametric finite element solution of the generalised Bloch-Torrey equation for arbitrary domains*, Journal of Magnetic Resonance, 259 (2015), pp. 126 – 134.

[2] FENICS, *Fenics project*, `http://www.fenicsproject.org`, (2003).

[3] D. S. GREBENKOV, *Pulsed-gradient spin-echo monitoring of restricted diffusion in multilayered structures*, Journal of Magnetic Resonance, 205 (2010), pp. 181–195.

[4] J. HOFFMAN, J. JANSSON, R. V. DE ABREU, N. C. DEGIRMENCI, N. JANSSON, K. MÜLLER, M. NAZAROV, AND J. H. SPÜHLER, *Unicorn: Parallel adaptive finite element simulation of turbulent flow and fluid-structure interaction for deforming domains and complex geometry*, Computers and Fluids, (2012).

[5] J. HOFFMAN, J. JANSSON, C. DEGIRMENCI, N. JANSSON, AND M. NAZAROV, *Unicorn: a Unified Continuum Mechanics Solver*, Springer, 2012, ch. 18.

[6] J. HOFFMAN, J. JANSSON, N. JANSSON, C. JOHNSON, AND R. V. DE ABREU, *Turbulent flow and fluid-structure interaction*, in Automated Solutions of Differential Equations by the Finite Element Method, Springer, 2011.

[7] J. HOFFMAN, J. JANSSON, N. JANSSON, AND M. NAZAROV, *Unicorn: A unified continuum mechanics solver*, in Automated Solutions of Differential Equations by the Finite Element Method, Springer, 2011.

[8] N. JANSSON, J. HOFFMAN, AND J. JANSSON, *Framework for Massively Parallel Adaptive Finite Element Computational Fluid Dynamics on Tetrahedral Meshes*, SIAM J. Sci. Comput., 34 (2012), pp. C24–C41.

[9] R. C. KIRBY, *FIAT: Numerical Construction of Finite Element Basis Functions,*, Springer, 2012, ch. 13.

[10] J.-R. LI, D. CALHOUN, C. POUPON, AND D. L. BIHAN, *Numerical simulation of diffusion mri signals using an adaptive time-stepping method*, Physics in Medicine and Biology, 59 (2014), p. 441.

[11] A. LOGG, K.-A. MARDAL, G. N. WELLS, ET AL., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, 2012.

[12] ——, *Automated Solution of Differential Equations by the Finite Element Method*, Springer, 2012.

[13] B. F. MORONEY, T. STAIT-GARDNER, B. GHADIRIAN, N. N. YADAV, AND W. S. PRICE, *Numerical analysis of NMR diffusion measurements in the short gradient pulse limit*, Journal of Magnetic Resonance, 234 (2013), pp. 165–175.

[14] D. V. NGUYEN, J.-R. LI, D. GREBENKOV, AND D. L. BIHAN, *A finite elements method to solve the Bloch - Torrey equation applied to diffusion magnetic resonance imaging* , Journal of Computational Physics, 263 (2014), pp. 283 – 302.

[15] W. S. PRICE, *NMR studies of translational motion: principles and applications*, Cambridge University Press, 2009.

[16] G. RUSSELL, K. D. HARKINS, T. W. SECOMB, J.-P. GALONS, AND T. P. TROUARD, *A finite difference method with periodic boundary conditions for simulations of diffusion-weighted magnetic resonance experiments in tissue*, Physics in Medicine and Biology, 57 (2012), p. N35.

[17] B. P. SOMMEIJER, L. F. SHAMPINE, AND J. G. VERWER, *RKC: An explicit solver for parabolic PDEs*, Journal of Computational and Applied Mathematics, 88 (1998), pp. 315–326.

[18] E. O. STEJSKAL AND J. E. TANNER, *Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient*, The Journal of Chemical Physics, 42 (1965), pp. 288–292.

[19] H. TORREY, *Bloch equations with diffusion terms*, Physical Review Online Archive (Prola), 104 (1956), pp. 563–565.

[20] J. XU, M. DOES, AND J. GORE, *Numerical study of water diffusion in biological tissues using an improved finite difference method*, Physics in medicine and biology, 52 (2007).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

THERMODYNAMICS OF IRREVERSIBLE AGGREGATION

Themis Matsoukas

150 Fenske Laboratory, Department of Chemical Engineering, Pennsylvania State University, University Park, PA 16802, United States

Keywords: Population balances, giant component, ensemble theory, thermodynamics, information theory.

Abstract. *The emergence of a giant component from a population of finite-size clusters is a problem in mathematical physics that is encountered in fields as diverse as polymer gelation, percolation of networks, and the spread of epidemics. But while such processes have long invited thermodynamic analogies, a formal connection to thermodynamics that goes beyond the qualitative has not been established. Here we develop a thermodynamic theory for a generic mathematical object, a population of individuals that cluster into groups. The theory views the distribution in the scaling limit as the one that is most probable among all distributions that satisfy the physical constraints of the problem. In this context, the emergence of the giant cluster is a phase transition that is governed by equilibrium criteria analogous to those in molecular systems.*

1 INTRODUCTION

Binary clustering is a process by which two elements (molecules, particles, aggregates or other) combine to form a new one. It is one of the most basic and ubiquitous growth mechanisms in nature, one that describes a wide range physical phenomena over a vast scale that encompasses molecular systems, social networks, and stars [1, 2]. It is industrially relevant in granulation, a unit operation designed to increase the size of granules and homogenize their composition through the aggregation of primary particles of the pure components. Roughly 80% of pharmaceutical formulations [3] and 60% of all industrial products [4] are processed at some point as solid materials. Despite such widespread application, the design, modelling and control of particulate processes has yet to achieve the same level of accuracy and predictability as other common industrial unit operations such as separations, for example. This unmet challenge in particle technology calls for new approaches to this problem. The main theoretical tool in the study of population balances is Smoluchowski's theory of aggregation, which formulates the governing equation for the size distribution. In the past 100 years since Smoluchowski's work, another view is emerging, one that treats populations from a probabilistic standpoint. This approach, which originated in the polymer literature [5, 2], has been migrating, albeit slowly, to particulate systems [6, 7]. We recently developed a new approach to the study of generic populations based on the notion of the cluster ensemble [8]. The theory considers the ensemble of all distributions that can be formed with a fixed number of primary particles placed into a fixed number of clusters and implements a selection rule to sample distributions from the ensemble. The macroscopic observable in the thermodynamic limit –the information accessible to the observer– is the most probable distribution in this ensemble. By proper construction of the selection rule, any conceivable distribution can be obtained as the most probable distribution of the ensemble. We derived the statistics of the ensemble and showed it is described by a mathematical calculus analogous that of statistical thermodynamics. Thus we have a general theory of populations formulated in the language of thermodynamics that is in principle applicable to generic populations. Here we propose to apply this theory to the study of population balances.

Why is this direction worth pursuing? First, the ensemble theory encompasses the Smoluchowski approach, but goes beyond it. For example, it is well known that the Smoluchowski equation fails with gelling kernels, a problem that has received great attention in the literature of mathematical physics [1, 2, 9, 10, 11]. Gelation in the cluster ensemble emerges naturally and coexistence conditions of the sol and the gel are obtained in direct analogy to those in phase equilibrium of molecular systems [12]. Second, The theory generates a new set of mathematical relationships between the primary inputs and outputs of the problem (aggregation rate, size distribution). These relationships are unavailable to the Smoluchowski theory and provide new ways to study and analyze population balance problems. And third, the theory makes available the language and toolbox of thermodynamics to the study of population balances. Many problems in this area could benefit from a thermodynamic insight but one that is particularly interesting and of practical significance is aggregative mixing. In granulation, we begin with a segregated state that consists of primary particles of the pure components, and end up with a population that is hopefully well-mixed at the granule level. How long does it take for components to reach an acceptable level of mixing? What if similar or dissimilar components tend to aggregate preferentially? More importantly, if it is found experimentally that granules remain poorly mixed, what does this tell us about the magnitude of unfavorable interactions between the components? The analogy to solution thermodynamics cannot be missed. Thermodynamics teaches us how quantify such interactions in solution using activity coefficients. Is there some

analogous metric that would allow us to interpret deviations from random mixing in granules in a similar manner? These are questions that we may now begin to address, not merely by qualitative analogy to thermodynamics, but by rigorous application of the cluster-ensemble theory [8, 12, 13].

2 THE CLUSTER ENSEMBLE

Suppose we have M identical particles ("primary particles" or monomers) which we distribute into N clusters, such that each cluster contains at least one particle. The cluster ensemble consists of all distributions that can be formed with fixed M and N. All distributions of the ensemble satisfy the conditions,

$$\sum_{i=1}^{\infty} n_i = N, \quad \sum_{i=1}^{\infty} i n_i = M, \tag{1}$$

where n_i is the number of clusters that contain i particles. We envision a stochastic process the pick distributions out of that ensemble. Given a distribution $\mathbf{n} = (n_1, n_, \cdots)$, we write its probability in the canonical form,

$$P(\mathbf{n}) = \mathbf{n}! \frac{W(\mathbf{n})}{\Omega_{M,N}}. \tag{2}$$

Here, $\mathbf{n}!$ is the multinomial coefficient of the list $(n_1, n_2, \cdots)$, $W(\mathbf{n})$ is the selection bias, and $\Omega_{M,N}$ is the partition function. These quantities are discussed in more detail below.

The multinomial coefficient is

$$\mathbf{n!} = \frac{N!}{n_1! n_2! \cdots},$$

and represents the natural multiplicity of the distribution in the ensemble. If we imagine that we sample a distribution one cluster at a time, $\mathbf{n}!$ is the number of different ordered sequences in which the clusters may appear. Taking the logarithm of the multinomial coefficient and applying the Stirling formula $x! \approx x \log x - x$, we obtain

$$\log \mathbf{n}! = -\sum_i n_i \log \frac{n_i}{N} = N S(\mathbf{p}), \tag{3}$$

where $\mathbf{p} = (p_1, p_2, \cdots) = (n_1/N, n_2/N, \cdots)$ is the probability of cluster size i in the distribution and $S[\mathbf{p}]$ is the Shannon entropy of the discrete probability distribution $\mathbf{p}$,

$$S = -\sum_i p_i \log p_i. \tag{4}$$

In other words, the log of the natural multiplicity is the Shannon entropy of the distribution.

The selection bias $W(\mathbf{n}) = W(n_1, n_2, \cdots)$ is a functional of the distribution that biases its selection. With $W(\mathbf{n}) = 1$ we have the special case of the unbiased ensemble, namely an ensemble in which the probability of distribution depends only on its natural multiplicity. The selection bias is the quantity of the ensemble that embodies the physics of the problem that governs the evolution of the population. The bias will remain for the moment unspecified and the only requirement we impose is that its logarithm must be homogeneous in n_i with degree 1. Accordingly, $\log W$ satisfies the condition [8]

$$\log W(\mathbf{n}) = \sum n_i \left(\frac{\partial \log W(\mathbf{n})}{\partial n_i} \right)_{n_j} \equiv \sum n_i \log w_i, \tag{5}$$

which follows from Euler's theorem for homogeneous functions. Here, $\log w_i$ is the partial derivative of the log bias with respect to n_i. It represents the contributions cluster size i to the log bias and we refer to it as the cluster bias.

The partition function is the normalization factor of probabilities and satisfies the normalizing condition,

$$\Omega_{M,N} = \sum_{\mathbf{n}} \mathbf{n}! W(\mathbf{n}). \tag{6}$$

The summation goes over all distributions that can be formed by M monomers assembled into N clusters. Accordingly, the partition function depends on M and N. It also depends on the selection bias, but to the extent that the bias is determined by the physics that govern the population, for a given process Ω is a function of M and N only.

2.1 Most probable distribution

The premise of the ensemble theory is that for large M, N, ensemble reduces to its most probable distribution (mpd). This distribution is obtained by maximizing the probability $P(\mathbf{n})$, or equivalently its logarithm, under the constraints in Eq. (1). We perform this constrained maximization by the method of Lagrange multipliers. The objective function to be maximized is

$$\mathcal{F} = \log \mathbf{n}! + \log W(\mathbf{n}) - \alpha \left(\sum n_i - N \right) - \beta \left(\sum i n_i - M \right). \tag{7}$$

Using Eq. (3) for $\log \mathbf{n}!$ and setting the derivative of $\mathcal{F}$ with respect to n_i equal to zero we find

$$\log \frac{\tilde{n}_i}{N} = -(\alpha + 1) - \beta i + \left(\frac{\partial W(\tilde{\mathbf{n}})}{\partial \tilde{n}_i} \right) \tag{8}$$

where $\tilde{n}_i$ refers to the most probable distribution. Solving for $\tilde{n}_i$, the most probable distribution takes the form [8]

$$\frac{\tilde{n}_i}{N} = \tilde{w}_i e^{-\beta i}/q, \tag{9}$$

where $q = \exp(\alpha + 1)$ and $\log \tilde{w}_i$ are the partial derivatives of $\log \tilde{W} = \log W(\tilde{\mathbf{n}})$ evaluated at the most probable distribution.

2.2 The partition function in the thermodynamic limit

According to the maximum term method, as the ensemble converges to the most probable distribution, the log of the partition function converges to the log of the maximum term:

$$\log \Omega_{M,N} = \sum_{\mathbf{n}} \log \mathbf{n}! W(\mathbf{n}) \to \log \tilde{\mathbf{n}}! W(\tilde{\mathbf{n}}) = -\sum n_i \log \frac{n_i}{N} + \log \tilde{W}. \tag{10}$$

We use Eq. (9) to calculate the right-hand side of the above equation:

$$\log \Omega_{M,N} = \beta \sum i n_i + \log \sum n_i + \sum n_i \log \tilde{w}_i, \tag{11}$$

and this, by virtue of Eq. (1) and the homogeneity of the log bias in Eq. (5), is written in the more compact form [8],

$$\log \Omega_{M,N} = \beta M + (\log q) N. \tag{12}$$

In the thermodynamic limit $\tilde{n}_i/N$ is an intensive property. It follows from Eq. (9) that β, $\log q$ and $\tilde{w}_i$ are intensive, and then from Eq. (12) that $\log \Omega_{M,N}$ is extensive, i.e., it is homogeneous in M and N with degree 1. By Euler's theorem $\log \Omega$ is of the form

$$\log \Omega_{M,N} = M \left(\frac{\partial \log \Omega}{\partial M} \right)_N + N \left(\frac{\partial \log \Omega}{\partial N} \right)_M . \tag{13}$$

Direct comparison of this result with Eq. (12) leads to the identifications

$$\beta = \left(\frac{\partial \log \Omega}{\partial M} \right)_N , \quad \log q = \left(\frac{\partial \log \Omega}{\partial N} \right)_M . \tag{14}$$

Equations (9), (12) and (14) summarize the results of the cluster ensemble: the most probable distribution depends on three intensive variables: β and q, which are given by the partial derivatives of the partition function, and $\tilde{w}_i$, which are the partial derivatives of the selection bias.

The differential of $\log \Omega$ with respect to M and N is

$$d \log \Omega = \beta dM + (\log q) dN. \tag{15}$$

The result bears a direct analogy to the thermodynamic differential,

$$d \log \Omega = \beta dE + \beta P dV - \beta \mu dn, \tag{16}$$

where Ω is the microcanonical partition function, E is energy, V is volume and n is the number of particles. M and N in the cluster ensemble may be taken to be analogous to E and n (other analogies can be drawn as well [8]). Accordingly, β in the cluster ensemble is analogous to inverse temperature and $\log q$ to chemical potential. There is one difference: in the thermodynamic ensemble the log of the partition function is entropy; in the cluster ensemble the corresponding relationship is

$$\log \Omega_{M,N} = \tilde{S} + \log \tilde{W}.$$

In the special case $W = 1$ (unbiased ensemble), we obtain $\log \Omega_{M,N} = \tilde{S}$, as in molecular systems, and the most probable distribution is the familiar exponential distribution of the canonical ensemble. In fact, $W = 1$ represents the mathematical statement of the postulate of equal a priori probabilities, which forms the basis for the derivation of the canonical distribution in statistical mechanics. The cluster ensemble is a generalization to systems that are not limited to unbiased selection. Whereas the most probable distribution of of thermodynamic microstates is exponential, the most probable distribution in the cluster ensemble can be *any* distribution. Conversely, any distribution may be viewed as the most probable distribution of the cluster ensemble under appropriate selection bias.

3 LINEAR ENSEMBLES

The cluster bias w_i, defined in Eq. (5), is a functional of the distribution, i.e., it depends on all n_i. An important special case is the selection bias is of the form,

$$W(\mathbf{n}) = a_1^{n_1} a_2^{n_2} \cdots \tag{17}$$

where a_i functions of i. Its logarithm is

$$\log W(\mathbf{n}) = a_1 n_1 + a_2 n_2 + \cdots , \tag{18}$$

from which we obtain the cluster bias,

$$w_i = \left(\frac{\partial \log W}{\partial n_i} \right)_{n_j} = a_i. \tag{19}$$

In this special type the cluster bias w_i is an intrinsic function of i, i.e., it is independent of the distribution. We call this type of selection functional linear because its logarithm is a linear combination with fixed coefficients. The corresponding ensemble, which we call linear, has several interesting properties. Combining this selection bias with Eq. (6) we obtain

$$\Omega_{M,N} = N! \sum_{n_i} \frac{a_1^{n_1}}{n_1!} \frac{a_2^{n_2}}{n_2!} \cdots , \tag{20}$$

with the summation going over all clusters of the ensemble. The derivative with respect to a_k gives

$$\frac{\partial \Omega_{M,N}}{\partial a_k} = N \left((N-1) \sum_{n_i} \cdots \frac{a_k^{n_k-1}}{(n_k-1)!} \cdots \right). \tag{21}$$

Notice that the differentiation amounts to removing a cluster of mass k from the summation, therefore producing the partition function of the ensemble with mass $M-k$ and number of particles $N-1$:

$$\frac{\partial \Omega_{M,N}}{\partial n_i} = N \Omega_{M-i,N-1}. \tag{22}$$

An equivalent form to write Eq. (21) is

$$\frac{\partial \Omega_{M,N}}{\partial a_k} = \frac{\Omega_{M,N}}{a_k} \left(\frac{1}{\Omega_{M,N}} \sum_{n_i} n_k \cdots \frac{a_k^{n_k-1}}{n_k!} \cdots \right) = \frac{\Omega_{M,N}}{a_k} \langle n_k \rangle , \tag{23}$$

whose result on the right is obtained by recognizing the quantity in brackets as the ensemble average of the number of clusters with k particles. The ensemble average $\langle n_k \rangle$ then is

$$\langle n_k \rangle = N a_k \frac{\Omega_{M-k,N-1}}{\Omega_{M,N}}. \tag{24}$$

Therefore, the mean cluster distribution of the linear ensemble can be calculated directly from the partition function and the cluster bias a_k.

4 DISCRETE BINARY AGGREGATION

The theory of the cluster ensemble discussed above is very general and applies to any population. We will now demonstrate the theory by applying it to a physical process, binary aggregation. In binary aggregation two clusters with mass i and j, respectively, combine to form a new cluster with mass $i+j$. The process represented by the reaction

$$(i) + (j) \xrightarrow{k_{ij}} (1+j)$$

whose rate is characterized by the aggregation kernel, k_{ij}. Given a cluster distribution $\mathbf{n}$, the probability that cluster masses i and j combine into a cluster is proportional to the aggregation

kernel and the number of (i,j) pairs. In discrete finite systems the number of pairs is $n_i n_j$ if $i \neq i$, and $n_i(n_i-1)/2$ if $i=j$. Thus the general case can be expressed in the form,

$$P_{(i)+(j)\to(i+j)} = C_{\mathbf{n}} \frac{n_i(n_j-\delta_{i,j})}{1+\delta_{i,j}}\, k_{i,j}, \tag{25}$$

with

$$C_{\mathbf{n}} = \frac{2}{N(N-1)\bar{k}(\mathbf{n})}, \tag{26}$$

with the constant C_n is such that the sum of the transition probabilities are normalized. The factor $\bar{k}(\mathbf{n})$ on the right-hand side of (26) is the mean kernel within distribution $\mathbf{n}$ and is calculated by averaging over all possible aggregation events of the $N(N-1)/2$ pairs of clusters in the distribution:

$$\bar{k}(\mathbf{n}) = \frac{2}{N(N-1)} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{n_i(n_j-\delta_{i,j})}{1+\delta_{i,j}}\, k_{i,j}. \tag{27}$$

The aggregation event between two cluster masses may be viewed as a reaction that converts a parent distribution into a new distribution that has the same total mass but one less particle. This parent-offspring relationship establishes a network of connections that link the distributions of the (M,N) ensemble to those of the $(M,N-1)$ ensemble. This network is shown in Fig. 1 for $M=7$. At the top of the network we have $M=7$, $N=7$, corresponding to a population of monomers that represents generation (fully dispersed state). In the next step, two monomers combine to form a dimer. In the third step a monomer may add to the dimer to produce a trimer, or two monomers combine to form a new dimer. From there on the possibilities increase, until at the end we reach the fully gelled system in which all monomers are part of the same cluster $(M=7,N=1)$. The transition between parents and offsprings can be conveniently tracked by the generation g, which we define as $g=M-N+1$. The completely dispersed state corresponds to generation $g=1$ and the completely gelled state to $g=N$. The distributions in generation g contain the complete ensemble of distributions that can be formed by M monomers into N clusters. For example, with $M=7$, $N=2$, the possible distributions are monomer + hexamer ($n_1=1$, $n_6=1$), dimer + pentamer ($n_2=1$, $n_5=1$) and trimer+tetramer ($n_3=1$, $n_4=1$). All are present in generation $g=6$. The aggregation network may them be viewed as a process that acts on the distributions of the (M,N) ensemble to produce the distributions of the $(M,N-1)$ ensemble. The problem now becomes to find the equations that govern this evolution of the ensemble.

4.1 Parent-offspring relationship

There is a systematic way to construct all parents of $\mathbf{n}$. A cluster of mass $i \geq 2$ is formed by the aggregation of masses $i-j$ and j. Accordingly, if we break a cluster i of distribution $\mathbf{n}$ into fragments $i-j$ and j, the resulting distribution is a parent of $\mathbf{n}$, specifically, the $(i-j,j)$-parent. Using unprimed variables for the offspring and primed for the parent, the equations that define the parent of distribution $\mathbf{n}$ are

$$\begin{aligned} n'_i &= n_i - 1; \\ n'_{i-j} &= n_{i-j} + 1 + \delta_{i-j,j}; \\ n'_j &= n_j + 1 + \delta_{i-j,j}; \end{aligned} \tag{28}$$

Clearly, only clusters with $i \geq 2$ produce a parent (the monomer cannot be produced by aggregation of smaller particles), The complete set of the parents of $\mathbf{n}$ is obtained by letting i, j, span

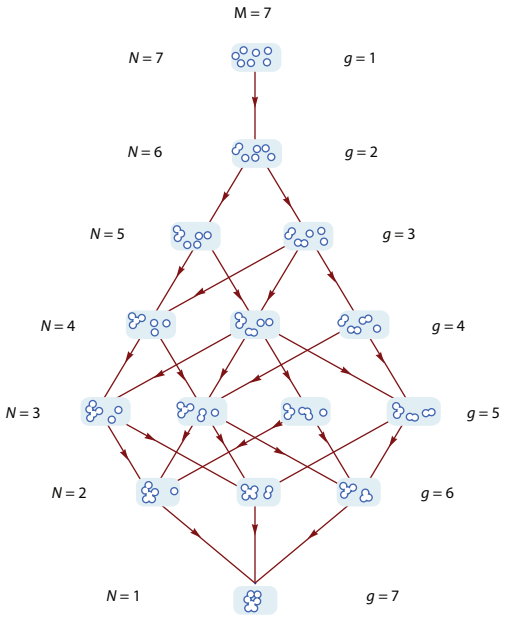


Figure 1: The network for discrete aggregation of $M = 7$ identical particles.

the range

$$i = 2, \cdots \infty; \quad j = 1, \cdots i/2. \tag{29}$$

where $i/2$ is understood as integer division. For example, the distribution that contains one monomer, one dimer and one tetramer has three parents: one corresponds to breaking the dimer into monomers (the result is three monomers and a trimer); the second corresponds to breaking the tetramer into a monomer and a trimer (the result is two monomers, a dimer and a trimer); and the third corresponds to breaking the tetramer into two dimers (the result is a monomer and three dimers).

4.2 The Master Equation

The probability of distribution propagates from parent to offspring via the Master Equation,

$$P(\mathbf{n}) = \sum_{\mathbf{n}'} P(\mathbf{n}') P_{\mathbf{n}' \to \mathbf{n}}. \tag{30}$$

Here, $P(\mathbf{n})$ is the probability of the offspring, $P(\mathbf{n}')$ the probability of the parent, and $P_{\mathbf{n}' \to \mathbf{n}}$ is the transition probability from parent to offspring. The transition probability from parent $\mathbf{n}'$ to offspring $\mathbf{n}$ is given by Eq. (31) with the understanding that the right-hand side must be expressed in terms of the parent. This amounts to replacing $\mathbf{n}$ by $\mathbf{n}'$ and N by $N + 1$. We now take $\mathbf{n}'$ to be the $(i - j, j)$-parent, namely, the parent that produces a new cluster of size i via the aggregation of cluster masses $(i - j)$ and (j). The transition probability then becomes

$$P_{\mathbf{n}' \to \mathbf{n}} = \frac{2}{N(N+1)} \frac{n_i'(n_j' - \delta_{i,j})}{1 + \delta_{i,j}} \frac{k_{i,j}}{\bar{k}(\mathbf{n}')}. \tag{31}$$

Finally, we express the probabilities $P(\mathbf{n})$, $P(\mathbf{n}')$ in the canonical form of Eq. (2),

$$P(\mathbf{n}) = \mathbf{n}! \frac{W(\mathbf{n})}{\Omega_{M,N}}; \quad P(\mathbf{n}') = \mathbf{n}'! \frac{W(\mathbf{n}')}{\Omega_{M,N+1}}, \tag{32}$$

and substitute these expressions in to the Master Equation. The resulting equation produces two separate recursions, one for the partition function and another for the selection bias. The recursion for the partition function is

$$\frac{\Omega_{M,N}}{\Omega_{M,N+1}} = \frac{M-N}{M}\frac{1}{\langle k_{M,N+1}\rangle}, \tag{33}$$

and is easily solved to produce the closed-form result

$$\Omega_{M,N} = \binom{M-1}{N-1}\left(\prod_{L=N+1}^{M}\langle k_{M,L}\rangle\right). \tag{34}$$

The recursion for the selection bias does not have a similar closed-form expression and is given by

$$W(\mathbf{n}) = \sum_{i=2}^{\infty}\frac{n_i(i-1)}{M-N}\left\{\frac{1}{i-1}\sum_{j=1}^{i-1}\frac{k_{i-j,j}\,\langle k_{M,N+1}\rangle}{\bar{k}(\mathbf{n}')}W(\mathbf{n}')\right\}. \tag{35}$$

The ensemble average kernel that appears in these equation is the average of $\bar{k}(\mathbf{n})$ overall distributions,

$$\langle k_{M,N}\rangle = \sum_{\mathbf{n}\in(M,N)} P(\mathbf{n})\,\bar{k}(\mathbf{n}) \tag{36}$$

with the summation running over all distributions of the (M,N) ensemble. $\langle k_{M,N}\rangle$ and $\bar{k}(\mathbf{n})$ are generally different: $\bar{k}(\mathbf{n})$ is the mean kernel within distribution $\mathbf{n}$ and $\langle k\rangle$ is the average of these mean kernels over all distributions of the ensemble. Equation (35) is a recursion that gives the bias of distribution $\mathbf{n}$ in terms of the bias $W(\mathbf{n}')$ of its parents. The double summation goes over all parents of $\mathbf{n}$: the inner summation goes over all possible ways to form a cluster of size i through the aggregation of sizes $(i-j)$ and (j), and the outer summation goes over all $i>1$ (the monomer cannot be formed by aggregation).

Having obtained the partition function, we derive expressions for the parameters β and $\log q$ in the most probable distribution from Eq. (14),

$$\beta = \frac{\Omega_{M+1,L}}{\Omega_{M,L}} = \frac{M}{M-N+1}\prod_{L=N+1}^{M}\frac{\langle k_{M+1,N+l}\rangle}{\langle k_{M,N+l}\rangle} \tag{37}$$

$$\log q = \frac{\Omega_{M,N+1}}{\Omega_{M,N}} = \frac{M-N}{N}\frac{1}{\langle k_{M,N+1}\rangle}, \tag{38}$$

which follow by applying the discrete equivalent of the derivatives $\partial\Omega/\partial M$ and $\partial\Omega/\partial N$. Equations (34)–(38) are a closed set of equations for β, $\log q$ and W, which together allow us to calculate the most probable distribution. To obtain the final distributions closed form we must first solve the recursion for W. Certain special cases where this can be done analytically are discussed below.

5 EXACT SOLUTIONS

In general, the mean kernel in distribution, $\bar{k}(\mathbf{n})$, and the ensemble average kernel, $\langle k_{M,N}\rangle$, are not the same. A special class of kernels is when these two averages are the same:

$$\langle k_{M,N}\rangle = \bar{k}(\mathbf{n}); \quad \text{for all } \mathbf{n} \text{ in the } (M,N) \text{ ensemble.} \tag{39}$$

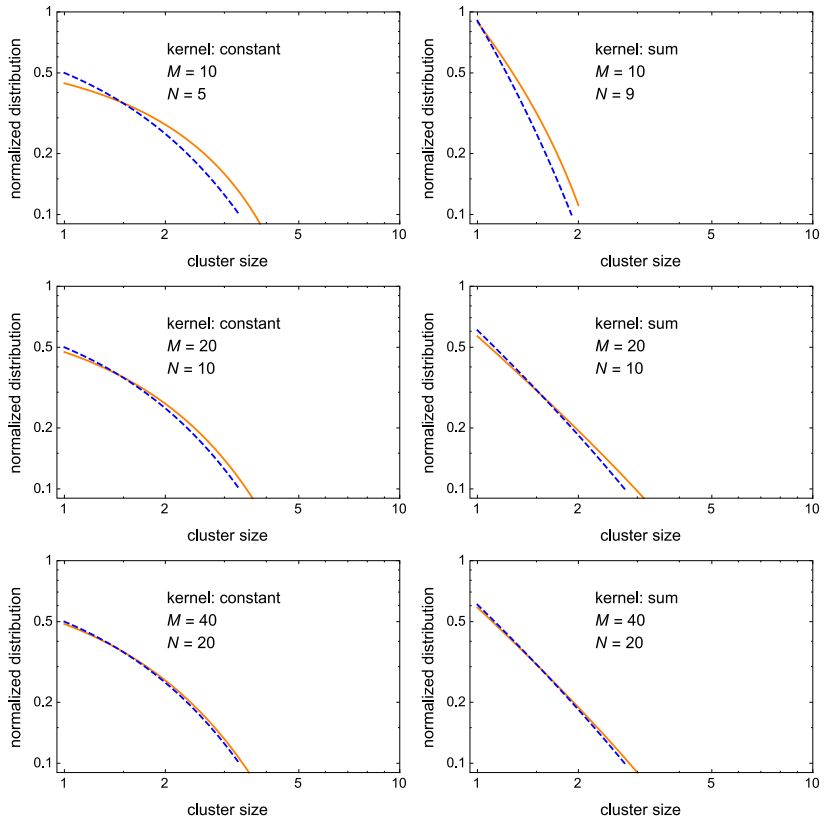


Figure 2: Distributions for the constant kernel (left column) and the sum kernel (right column) for various values of N, N, at $M/N = 2$. The solid line is the discrete solution (Eqs. 45 and 55) and the dashed line is the most probable distribution in the thermodynamic limit (Eqs. 48, 58). The two distributions converge as M and N increase.

If we insert this condition into Eq. (35) we obtain the following simplified recursion,

$$W(\mathbf{n}) = \sum_{i=2}^{\infty} \frac{n_i(i-1)}{M-N} \left\{ \frac{1}{i-1} \sum_{j=1}^{i-1} k_{ij} W(\mathbf{n}') \right\}, \tag{40}$$

whose solution is a linear bias with coefficients

$$a_i = \frac{1}{i-1} \sum_{j=1}^{i} a_{i-j} a_j k_{i-j,j}; \quad a_1 = 1. \tag{41}$$

The only kernels that satisfy Eq. (39) and produce linear ensembles are the constant kernel ($k_{ij} = 1$), the sum kernel ($k_{ij} = (i+j)/2$) and their linear combinations. Below we derive the results for the constant and sum kernels.

5.1 Constant kernel

The constant kernel, $k_{ij} = 1$, satisfies the condition in Eq. (39) trivially,

$$\langle k_{ij} \rangle_{M,N} = \bar{k}(\mathbf{n}) = k_{ij} = 1; \quad \text{for all } M, N, i, j, \mathbf{n} \tag{42}$$

and with this result Eq. (41) gives

$$a_i = 1; \quad \text{for all } i. \tag{43}$$

The corresponding cluster bias is $W(\mathbf{n}) = 1$ for all $\mathbf{n}$, which means that the constant kernel produces an unbiased ensemble, one in which all distributions are equally probable. Various properties of the ensemble can now be calculated. First, the partition function, which follows from Eq. (34):

$$\Omega_{M,N} = \binom{M-1}{N-1}. \tag{44}$$

Next, the mean distribution from Eq. (24):

$$\frac{\langle n_k \rangle}{N} = \binom{M-k-1}{N-2} \Big/ \binom{M-1}{N-1}. \tag{45}$$

This result is exact for any finite M and N. For the most probable distribution we return to Eq. (9). Before we proceed we recall that the derivation treats the most probable distribution as a continuous function, which implicitly assumes large M and N. For this reason we will derive the results in the limit $M, N \to \infty$ at fixed $M/N = \bar{x}$ (thermodynamic limit). Using Eq. (14), the parameters β and q are

$$\beta = \frac{\Omega_{M+1,N}}{\Omega_{M,N}} = \log \frac{M}{M-N} \to \log \frac{\bar{x}}{\bar{x}-1}, \tag{46}$$

$$\log q = \frac{\Omega_{M,N+1}}{\Omega_{M,N}} = \log \frac{M-N}{N} \to \log(\bar{x}-1). \tag{47}$$

The most probable distribution is obtained by substituting these results in Eq. (9),

$$\frac{\tilde{n}_k}{N} = \frac{(1-1/\bar{x})^k}{\bar{x}-1} \to \frac{e^{-k/\bar{x}}}{\bar{x}}. \tag{48}$$

The limiting form on the right-hand side is obtained for $\bar{x} \gg 1$. In the unbiased ensemble the most probable distribution is the distribution that maximizes entropy, and as is well known, of all distribution with the same mean $\bar{x}$, the distribution that maximizes entropy is exponential. This is also the well-known solution of the Smoluchowski equation for the constant kernels, therefore we have also made contact with the classical literature on population balances.

5.2 Sum kernel

The sum kernel is proportional to the sum of the cluster masses,

$$k_{ij} = \frac{i+j}{2}. \tag{49}$$

In this case the mean aggregation kernel in any distribution of the (M, N) ensemble is

$$\langle k_{12} \rangle_{M,N} = \frac{M}{N}, \tag{50}$$

from which it follows that $\langle k \rangle_{M,N} = M/N$. The product of kernels that appears in Eq. (34) is

$$\prod_{l=N+1}^{M} \langle k_{12} \rangle_{M,l} = \frac{M}{M} \cdot \frac{M}{M-1} \cdots \frac{M}{N+1} = \frac{N!}{M!} M^{M-N}, \tag{51}$$

and the partition function becomes

$$\Omega_{M,N} = \frac{M^{M-N}}{M!}\binom{M-1}{N-1} \tag{52}$$

The cluster bias requires inversion of the recursion

$$a_i = \frac{i}{2(i-1)}\sum_{j=1}^{i-1} a_{i-j}a_j, \quad a_1 = 1. \tag{53}$$

The same recursion was obtained by Spouge [14] in a combinatorial study of aggregation, and also in [1] in a treatment of the same problem based on Smoluchowski equation; its solution is [1]

$$a_i = \frac{i^{i-1}}{i!}. \tag{54}$$

By application of Eq. (24), the mean distribution is

$$\frac{\langle n_k\rangle_{M,N}}{N} = \frac{(M-k)^{M-N-k}}{(M-N-k+1)!}\cdot\frac{(N-1)(M-N)!}{M^{M-N-1}} \tag{55}$$

The parameters β and $\log q$ are

$$\beta = \frac{\Omega_{M+1,N}}{\Omega_{M,N}} \to \frac{x-1}{x} - \log\left(\frac{x-1}{x}\right) \tag{56}$$

$$\log q = \frac{\Omega_{M,N+1}}{\Omega_{M,N}} \to \log\left(\frac{\bar{x}-1}{\bar{x}}\right) \tag{57}$$

and the most probable distribution is

$$\frac{\tilde{n}_k}{N} = \frac{k^{k-1}}{k!}\left(\frac{x-1}{x}\right)^{k-1} e^{-k(x-1)/x}. \tag{58}$$

In contrast to Eq. (55), which is valid for any M, N, Eq. (58) is appropriate in the thermodynamic limit. The classical solution for the number concentration c_i as a function of time, t, is [1]

$$\frac{n_k}{N} = \frac{k^{k-1}}{k!}\left(1-e^{-t}\right)^{k-1} e^{-k(1-e^{-t})}. \tag{59}$$

where t is time. With the substitution $e^t = \bar{x}$, this reverts to (58).

Figure 2 shows graphs of the mean and the most probable distribution at fixed $/N = 2$ for $M = 10, 20, 40$. The mean distribution is exact for all M and N. The most probable distribution approaches the mean as M increases. The convergence between the mean and the most probable distribution demonstrates the basic premise of the cluster ensemble: in the thermodynamic limit the ensemble converges into a single distribution. In this limit the mean and the most probable distributions are identical.

6 THE GIANT COMPONENT

The convergence of the mean distribution and the most probable distribution is a trait of single-phase systems. This equality is destroyed when multiple phases are present. Since each phase converges to its own mean and most probable distribution, the two-phase system does not consist of a single distribution but is a linear combination of the distributions in each phase. Is it possible for an irreversible system of aggregating clusters to exhibit phase equilibrium? Yes. The classical example is the product kernel,

$$k_{ij} = ij, \tag{60}$$

which is known to result in gelation [1, 10]. The physical manifestation of gelation comes to us from polymer science and is manifested by the formation of a giant network, a single "cluster" that contains a finite fraction of the total particles in the system. This network coexists with a phase of dispersed finite-size clusters (the sol). This process has been long viewed as a phase transition by qualitative analogy to vapor/liquid systems. We now have the tools to address this problem in rigorous thermodynamic terms.

The product kernel asymptotically gives

$$\bar{k}(\mathbf{n}) \approx \left(\frac{M}{N}\right)^2, \tag{61}$$

a result that we obtain by replacing the diagonal elements $k_{ii}n_i(n_i-1)/2$ with $k_{11}n_i^2$. Except for distributions that contain a sizeable fraction of mass in very large clusters, this replacement is inconsequential. This approximation implicitly constraints the above relationship to the sol phase only, but as we will see, this will not prevent us from applying the theory to the two-phase system. It only means that the above expression can be used only for the properties of the sol, whereas the properties of the gel must be calculated from the the total mass balance. Since $\bar{k}(\mathbf{n})$ is (asymptotically) the same in all distributions of the ensemble, the product kernel satisfies Eq. (39) and the theory of linear ensembles applies to the sol distribution.

6.1 Sol distribution

The product of kernels in Eq. (34) is easily obtained in closed form

$$\prod_{L=N+1}^{M} \langle k_{ij}\rangle_{M,L} = \left(\frac{M}{M}\right)^2 \cdots \left(\frac{M}{N+1}\right)^2 = \left(\frac{N!}{M!}M^{M-N}\right)^2, \tag{62}$$

and leads to the following expression for the partition function [12]

$$\Omega_{M,N} = \left(\frac{N!}{M!}M^{M-N}\right)^2 \binom{M-1}{N-1} \tag{63}$$

The recursion for a_i in Eq. (53) becomes

$$a_i = \frac{1}{i-1}\sum_{j=1}^{i-1}(i-j)ja_{i-j}a_j$$

and its solution is

$$a_k = 2\frac{(2i)^{i-2}}{i!}. \tag{64}$$

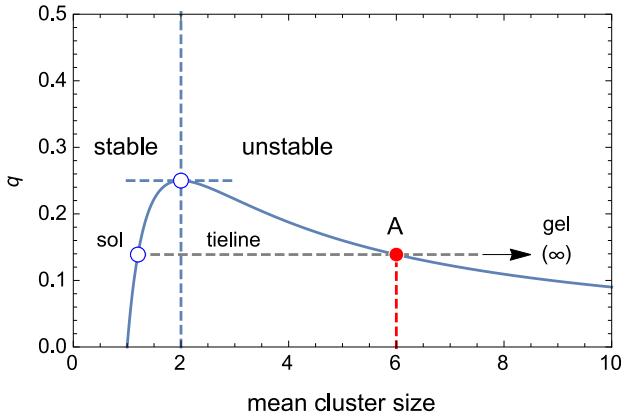


Figure 3: The parameter q for the product kernel. Stability requires $dq/d\bar{x} > 0$. Accordingly, the sol can exist only in the region $1 < \bar{x}_{\text{sol}} < 2$.

For the parameters β and q we find [12]

$$\beta = \frac{2(M-N)}{M} - \log\frac{M-N}{M} \to \frac{2(\bar{x}-1)}{\bar{x}} - \log\left(\frac{\bar{x}-1}{\bar{x}}\right)$$
$$\log q = \log\frac{N(M-N)}{M^2} \to \log\left(\frac{x-1}{x^2}\right).$$

With these results, the most probable distribution of the sol phase is

$$\frac{\tilde{n}_k}{N} = \frac{2^{k-1}k^{k-2}\bar{x}}{k!}\left(\frac{\bar{x}-1}{\bar{x}}\right)^{k-1} e^{-2k(x-1)/\bar{x}}. \tag{65}$$

6.2 Sol distribution

Let us return to the fundamental equation of the cluster ensemble, Eq. (12). The most probable distribution maximizes the partition function ($d\Omega = 0$) with respect to all distributions $\mathbf{n}$ with fixed M and N. Accordingly, Ω (as well as $\log\Omega$) is concave in M and N, which further implies that the second derivatives of $\log\Omega$ in M and N are both negative:

$$\left(\frac{\partial^2\Omega}{\partial M^2}\right) = \left(\frac{\partial\beta}{\partial M}\right)_N < 0, \quad \left(\frac{\partial^2\Omega}{\partial N^2}\right) = \left(\frac{\partial\log q}{\partial N}\right)_M < 0. \tag{66}$$

Using $\bar{x} = M/N$, the derivative with respect to N is

$$\left(\frac{\partial\log q}{\partial N}\right)_M = \left(\frac{\partial\log q}{\partial\bar{x}}\right)_M\left(-\frac{\bar{x}}{M}\right) < 0, \tag{67}$$

which requires the derivative $(\partial\log q/\partial\bar{x})$ to be positive. In aggregation M is constant, therefore, stability requires q to be a monotonically increasing function of $\bar{x}$. As we see in Fig. 3, q plotted against $\bar{x}$ has a maximum at $\bar{x} = x_* = 2$ that divides the domain of $\bar{x}$ into two regions: a stable region to the left of x_* and an unstable region to the right. If $M/N < 2$, the entire system exists as a stable single-phase sol with $\bar{x}_{\text{sol}} = M/N$. When $M/N > 0$, the single sol is unstable and the system splits into two phases, a sol with $x_{\text{sol}} < 2$ and a gel with $\bar{x}_{\text{gel}} > 2$. To

construct the tie line, we recall Eq. (33), which governs the evolution of the partition function during aggregation. This result states that the $\log q$ of the sol satisfies

$$\log q_{\text{sol}} = \left. \frac{M-N}{N} \frac{1}{\langle k_{M,N+1} \rangle} \right|_{\text{sol}}, \tag{68}$$

which establishes the condition that defines the tie line. Consider for example state A corresponding to an overall mean cluster size $\bar{x} = 6$. This state cannot exist as a single sol but splits into a stable sol (point S), and a gel cluster, whose size cannot be represented on this graph, which we recall is based on equations that apply only to the sol phase. The size of the gel cluster can be calculated if we know the M and N values of the overall system. The two phase system consists of a single gel cluster [8] with mass m_{gel} and $N-1$ sol clusters with average size $\bar{x}_{\text{sol}}$. By mass balance,

$$\bar{x}_{\text{sol}}(N-1) + m_{\text{gel}} = M. \tag{69}$$

The gel fraction $\phi_{\text{gel}} = m_{\text{gel}}/M$ is

$$\phi_{\text{gel}} = 1 - \bar{x}_{\text{sol}} \frac{N-1}{M}. \tag{70}$$

For large M and N this becomes

$$\phi_{\text{gel}} = 1 - \frac{\bar{x}_{\text{sol}}}{\bar{x}}. \tag{71}$$

Thus we have the gel fraction in terms of he overall mean cluster size $\bar{x}$ and the equilibrium sol size $\bar{x}_{\text{sol}}$, determined graphically from the tie line.

7 CONCLUSIONS

Here is what we have accomplished. We have formulated the cluster ensemble, a new theory for generic populations. The theory builds on Gibbs's notion of ensemble, with an added feature, a selection functional that establishes the selection rule among distributions of the ensemble. With the only requirement of homogeneous behavior, the ensemble is seen to obey regular thermodynamics. By "thermodynamics" we mean that it possesses the following properties:

1. It is governed by a partition function whose maximization with respect to the primitive stochastic variable of the ensemble (distribution $\mathbf{n}$) as well as with respect to its extensive independent variables (M and N) determines the equilibrium state.

2. It is equilibrated when its distribution relaxes to the most probable distribution.

3. The partition function satisfies the homogeneous condition

$$\log \Omega_{M,N} = M \left(\frac{\partial \log \Omega}{\partial M} \right)_N + N \left(\frac{\partial \log \Omega}{\partial N} \right)_M.$$

   and the stability conditions

$$\left(\frac{\partial^2 \Omega}{\partial M^2} \right), \left(\frac{\partial^2 \Omega}{\partial N^2} \right) < 0.$$

This formalism describes *any* population. More generally, it describes the probability density function of any stochastic variable. Thus we recognize thermodynamics not as a physical theory, but as a probabilistic calculus whose applicability extends beyond molecular systems. The crux of the mathematical problem is the determination of the selection functional. In some cases this may be given as part of the model [8]. In more typical situations, this selection bias must be determined from the laws that govern the population. The case of binary aggregation discussed here is an example of how this would work in the general case. Start with the Master Equation of the problem, obtain the transition probabilities from the governing laws, express the probabilities of distribution in canonical form, and work out a recursive solution for the partition function and the selection bias.

All of this can be done much more easily using the Smoluchowski equation. Why venture into the complexities of the ensemble? The Smoluchowski equation is fine as long as (a) the only information we seek is the average distribution (in most cases adequate) and (b) the population consists of a single phase. When the second requirement is not met, the Smoluchowski equation breaks down. There are several examples, apart from gelation, where this is the case: Network connectivity, spread of epidemics, financial mergers and the emergence of monopolies, species extinction and dominance, economic inequality, are all systems in which a single element of the population, infinitesimal in terms of number fraction, possesses a finite fraction of the whole. For these types of problems, the cluster ensemble provides a rigorous tool, one whose success is well established in physics and chemistry, and whose applicability, we now recognize, is far more general.

REFERENCES

[1] F. Leyvraz, "Scaling theory and exactly solved models in the kinetics of irreversible aggregation," *Physics Reports*, vol. 383, no. 2-3, pp. 95–212, 2003.

[2] W. H. Stockmayer, "Theory of molecular size distribution and gel formation in branched-chain polymers," *The Journal of Chemical Physics*, vol. 11, no. 2, pp. 45–55, 1943.

[3] CSDD, "Tufts center for the study of drug development impact report," tech. rep., Tufts, 2012.

[4] S. M. Iveson, J. D. Litster, K. Hapgood, and B. J. Ennis, "Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review," *Powder Technology*, vol. 117, no. 1-2, pp. 3 – 39, 2001.

[5] P. J. Flory, "Molecular size distribution in three dimensional polymers. II. trifunctional branching units," *Journal of the American Chemical Society*, vol. 63, no. 11, pp. 3091–3096, 1941.

[6] D. Ramkrishna, "Analysis of population balance—iv: The precise connection between Monte Carlo simulation and population balances," *Chemical Engineering Science*, vol. 36, no. 7, pp. 1203 – 1209, 1981.

[7] D. Ramkrishna, *Population Balances*. San Diego: Academic Press, 2000.

[8] T. Matsoukas, "Statistical thermodynamics of clustered populations," *Phys. Rev. E*, vol. 90, p. 022113, Aug 2014.

[9] E. M. Hendriks, M. H. Ernst, and R. M. Ziff, "Coagulation equations with gelation," *J. Stat. Phys.*, vol. 31, pp. 519–563, 1983.

[10] A. A. Lushnikov, "Exact kinetics of the sol-gel transition," *Phys. Rev. E*, vol. 71, p. 046129, Apr 2005.

[11] A. A. Lushnikov, "Gelation in coagulating systems," *Physica D: Nonlinear Phenomena*, vol. 222, pp. 37–53, 10 2006.

[12] T. Matsoukas, "Statistical thermodynamics of irreversible aggregation: The sol-gel transition," *Sci. Rep.*, vol. 5, p. 8855, 2015.

[13] T. Matsoukas, "Abrupt percolation in small equilibrated networks," *Phys. Rev. E*, vol. 91, p. 052105, May 2015.

[14] J. L. Spouge, "Equilibrium polymer size distributions," *Macromolecules*, vol. 16, no. 1, pp. 121–127, 1983.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

RELATIVE ROLES OF MECHANICS AND BIOCHEMISTRY IN THE INITIATION AND PROGRESSION OF CEREBRAL ANEURYSM THROMBOSIS

Malebogo N. Ngoepe[1,2], and Yiannis Ventikos[3]

[1] Department of Mechanical Engineering
University of Cape Town, Cape Town, South Africa
malebogo.ngoepe@uct.ac.za

[2] Centre for High Performance Computing
CSIR, Cape Town, South Africa
malebogo.ngoepe@uct.ac.za

[3] Department of Mechanical Engineering
University College London, London, United Kingdom
y.ventikos@ucl.ac.za

Keywords: thrombosis, computational fluid dynamics, biochemistry, cerebral aneurysms

Abstract. Clotting in cerebral aneurysms can either serve to stabilise an aneurysm, or can accelerate the time to rupture. In the former case, it is thought that a fully occlusive clot is likely to reduce complex haemodynamic conditions within the aneurysm sac, thereby reducing the chances of rupture. In cases where the clot fills the aneurysm sac partially, further vascular wall degradation has been observed, thus accelerating the time to rupture. The aim of this study is to determine the relative contributions of mechanical processes and biochemistry in the process of cerebral aneurysm clot development. An idealised model of an aneurysm was developed and used as the geometry of interest. This made it possible to make comparisons without a strong dependence on the peculiarities of a specific patient's case. The CFD-ACE+ Multiphysics Package (ESI Group, Paris, France) was used to calculate flow and account for the transport and reaction of biochemical species that contribute to the clotting process. In addition, user-defined functions were used to write a code that coupled the flow, biochemistry and level-set methods, which made it possible to track the surface of the growing clot, and distinguish between normal and clotted blood. The model was then used to examine the relative contributions of mechanics and biochemistry to the initiation and progression of clotting. It was found that the process of clot initiation is governed mainly by mechanical processes. Biochemistry plays a greater role in the progression of the clot where thrombin, the key player in thrombosis contributes to the formation of fibrin and thus, to the stability of the clot.

1 INTRODUCTION

Thrombosis is a condition that is often observed in cerebral aneurysms [1]. In some instances, the presence of a clot can advance vascular wall degradation and thereby accelerate the processes that lead to rupture [2]. In other cases, the presence of a clot in the aneurysm can lead to a more stable situation [3]. Owing to the highly patient-specific nature of aneurysms and of coagulation, it has proven rather challenging to fully identify the factors that make one clot desirable and another not. What has been observed is that a fully occlusive clot in the sac tends to prevent complex flow conditions that are often thought to be a high risk factor for rupture [4], [5]. An additional layer of complexity is introduced by the observation of both spontaneous and device-induced thrombosis in aneurysms [6]–[8]. In order to understand how best to treat cerebral aneurysms, it is necessary that steps be taken in an effort to better understanding clotting in cerebral aneurysms. A key aspect to deciphering clotting is exploring the contributions of mechanics and biochemistry to the process.

Given the elusive timing of clot formation and general difficulty of observing the role of the various miniscule components that make up the coagulation system *in vivo*, the use of computational models may be of benefit in an exploratory study aiming to examine the mechanisms that contribute to clot formation. Various computational models have been developed in effort to elucidate features of aneurysm thrombosis [9]–[13]. Even though they have been developed with different goals in mind, they all make a contribution to the effort of elucidating aneurysm thrombosis. Some of these models have focused on aortic aneurysm thrombosis while others have aimed to understand cerebral aneurysm thrombosis. A distinction is made between these pathologies as it is thought that different mechanisms contribute to thrombus development in the two contexts. In this paper, we use a previously developed computational model to explore clotting patterns that emerge for different mechanical and biochemical inputs [13]. The study aims to examine how these different inputs lead to different outcomes, and which of those outcomes correlate with clot patterns observed *in vivo*.

2 METHODS

Two main tasks are described in the methodology of this work. The first pertains to the creation of a suitable, idealised, cerebral aneurysm geometry for examining the features of interest. The second describes the computational model used to simulate the clotting process. Both of these are detailed in the paragraphs that follow. A section detailing the variation in the models is included at the end of the methods section.

2.1 Idealised cerebral aneurysm geometry and meshing

An idealised aneurysm geometry was adapted from one described in a paper by Mulder et al [14]. The main motivation for using an idealised geometry in this study was to avoid obtaining results which would be unique to a specific aneurysm case only. Given that the work was carried out in an attempt to derive general principles and observations, an “idealised” aneurysm was deemed most suitable, while bearing in mind that there is a fairly broad range of aneurysm shapes and sizes. The geometry was created as a surface in CFD-GEOM (ESI Group, Paris, France), the geometry creation tool of the CFD-ACE+ Multiphysics Package (ESI Group, Paris, France), and can be seen in Figure 1. The geometry consists of the parent vessel with an inlet and an outlet, and a spherical cerebral aneurysm that is attached to the vessel. The neck of the aneurysm is located at the intersection between the parent vessel and the sphere. This surface was than exported as an STL file to CFD-VisCART (ESI Group, Par-

is, France), where a non-conforming, Cartesian mesh was created. In addition, boundary regions were defined in this step.

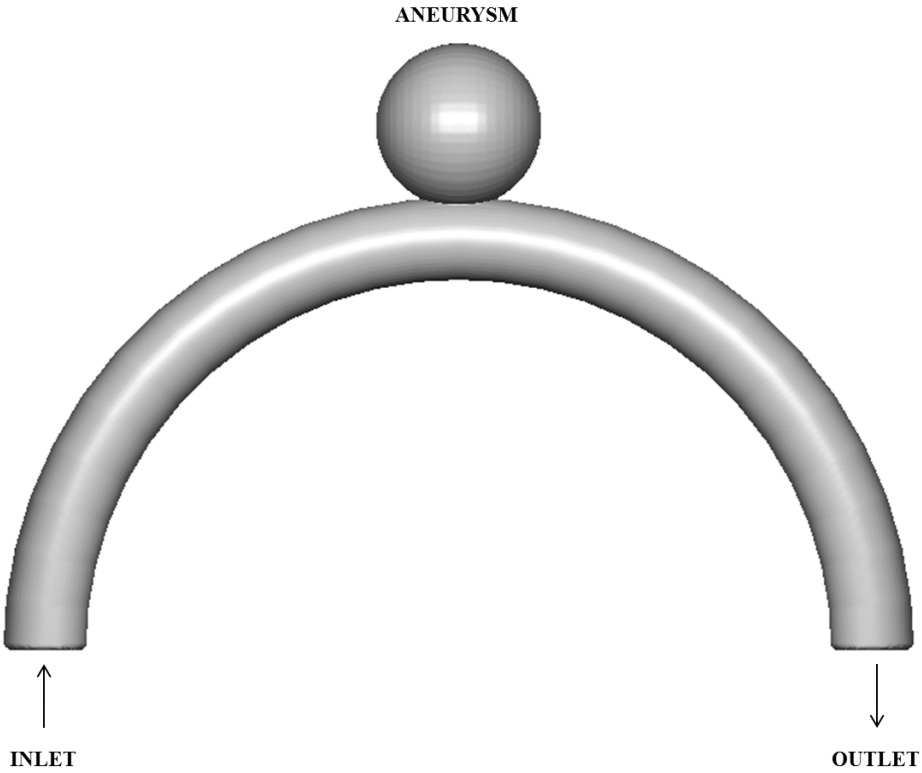


Figure 1: Idealised aneurysm geometry adapted from Mulder et al [14].

2.2 Integrated computational model

An integrated computational model capable of accounting for flow, transport of coagulation proteins, reactions between these proteins, the surface of the growing clot and the different properties of the clot region was used. This model consists of four main subsystems, namely flow, porosity and permeability, biochemistry, and level set, which were coupled to ensure that the results of the different systems had bearing on each other. The systems are described in the paragraphs that follow but greater detail can be found in the original paper [13].

The flow submodule describes the flow of blood through the parent vessel and the aneurysm, and accounts for the transport of the coagulation proteins. The continuity and momentum equations are given by (1) and (2), respectively.

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \boldsymbol{U}) = 0 \tag{1}$$

$$\rho \frac{\partial \boldsymbol{U}}{\partial t} + \rho \boldsymbol{U}.\nabla \boldsymbol{U} + \nabla P = \mu \nabla^2 \boldsymbol{U} \tag{2}$$

Where, ρ is density, t is time, $\boldsymbol{U}$ is the velocity vector, P is pressure and μ is dynamic viscosity. The transport equation which accounts for the transport of the different coagulation proteins is given by (3).

$$\frac{\partial C_j}{\partial t} + \boldsymbol{U}\nabla C_j = \nabla\left(\Gamma_j \nabla C_j\right) + S_j \tag{3}$$

Where, C_j is the concentration of a specific protein, Γ_j is the diffusivity of that protein, and S_j is the source of that protein. The diffusivity was adjusted in the clot region to account for the porosity of that region.

In the model, the clot region assumed different porosity and permeability values to the rest of the fluid region. This made it possible for the growing clot to have an impact on the flow region, and vice versa. Using values taken from Diamond and Anand, computational cells which fell within the clot region assumed porosity and permeability values of $\varepsilon = 0.75$ and $\kappa = 1.0\times10^{-12}m^2$, respectively [15].

The biochemistry submodule accounts for the reactions that take place during the coagulation process, and the subsequent changes in the species' concentrations as a result. The coagulation process was described using equations derived by Wagenvoord et al [16]. These were adapted from a six-reaction system to a four-reaction system, as seen in Table 1. Reactions were described using Michaelis-Menten kinetics, or first order kinetics. A stiff equation solver developed by Hairer and Wanner was coded into the user-defined subroutines using FORTRAN [17].

Step	Reaction Pathway	Forward Reaction Pathway ($nM^{-1}s^{-1}$)	kcat (s^{-1})	kM (nM)
1	X + TF → Xa-Va + TF		5.00E-02	1.00E+02
2	PT + Xa-Va → TH+ Xa-Va		8.00E-02	1.00E+03
3	V + TH → Xa-Va + TH		2.50E+01	5.50E+04
4	TH + AT → INACT	5.00E-06		

Table 1: Reaction constants for the adapted Wagenvoord model [16].

The level set submodule enables the tracking of the clot surface [18]. This is important as the careful distinction of the clot region from the rest of the fluid region enables the assignment of different properties to the different regions. The level set equation is given by (4).

$$\phi_t + V_n.|\nabla\phi| = 0 \tag{4}$$

Where ϕ_t is the partial derivative of the signed distance function with respect to time, V_n is the propagation velocity of the clot surface and ϕ is the signed distance function. In this study, V_n varies and details of the variation are given in the next section.

2.3 Variation in models

In order to explore the contributions of mechanics and biochemistry to clot development, it is necessary that input factors be varied. The first set of variations relates to the initiation of clot growth, while the second focuses on clot propagation. In the tables that follow, the propagation velocity column details the equation used for Vn for that case. Initial [TF] gives the starting tissue factor concentration applied at the initiating surface. For cases where tissue factor expression is dependent on a threshold strain rate value, the critical strain rate gives the value above or below which tissue factor is expressed on the initiating surface. Finally, the initiation surface specifies the surface on which tissue factor is expressed for clot initiation. The experiments detailed in Tables 2 and 3 are designed to explore clot initiation. Table 2 illustrates the variation in shear rate threshold for tissue factor expression. Below or above a

certain threshold, tissue factor is expressed, enabling clot initiation. Table 3 details the variations in initial tissue factor concentration required to initiate clotting in the aneurysm sac.

Case	Propagation velocity ($m.s^{-1}$)	Initial [TF] (M)	Critical Strain Rate (s^{-1})	Initiation surface
SR100	$V_n = a[TH]$	1.00E-17	100	SR<100s^{-1} on wall
SR500L	$V_n = a[TH]$	1.00E-17	500	SR<500s^{-1} on wall
SR500G	$V_n = a[TH]$	1.00E-17	500	SR>500s^{-1} on wall
SR800	$V_n = a[TH]$	1.00E-17	800	SR>800s^{-1} on wall

Table 2: Parameters for different critical strain rate models.

Case	Propagation velocity ($m.s^{-1}$)	Initial [TF] (M)	Critical Strain Rate (s^{-1})	Initiation surface
TFconc18	$V_n = a[TH]$	1.00E-18	N/A	Entire wall
TFconc17	$V_n = a[TH]$	1.00E-17	N/A	Entire wall
TFconc15	$V_n = a[TH]$	1.00E-15	N/A	Entire wall
TFconc12	$V_n = a[TH]$	1.00E-12	N/A	Entire wall

Table 3: Parameters for models with different initial tissue factor concentrations.

For clot propagation, the propagation velocity, V_n, bears relation to the quantity of interest as described in Tables 4 and 5. In Table 4, relationships with mechanical variables are detailed while in Table 5, relationships with biochemical factors are shown.

Case	Propagation velocity ($m.s^{-1}$)	Initial [TF] (M)	Critical Strain Rate (s^{-1})	Initiation surface
Mech_Const	$V_n = a$	N/A	N/A	Entire wall
Mech_SRDir	$V_n = a \times StrainRate$	N/A	N/A	Entire wall
Mech_SRInv	$V_n = a/StrainRate$	N/A	N/A	Entire wall

Table 4: Parameters for models linking propagation velocity to mechanical variables.

Case	Propagation velocity ($m.s^{-1}$)	Initial [TF] (M)	Critical Strain Rate (s^{-1})	Initiation surface
Bio_ThromDir	$V_n = a[TH]$	1.00E-12	N/A	Entire wall
Bio_ThromInv	$V_n = a/[TH]$	1.00E-12	N/A	Entire wall
Bio_ExtTenase	$V_n = a[XaVa_E]$	1.00E-12	N/A	Entire wall
Bio_TotTenase	$V_n = a[XaVa_T]$	1.00E-12	N/A	Entire wall

Table 5: Parameters for models linking propagation velocity to biochemical species concentrations

3 RESULTS

A comparison of clotting outcomes for different models is presented in this section. Distinction is made between results for clot initiation and those for progression.

3.1 Clot initiation

Clot growth for different strain rate thresholds is illustrated in Figure 2. For this model, a strain rate of 500s^{-1} is the defining value above which no clot development is observed. Below this value, clot development is uniform, regardless of the threshold value.

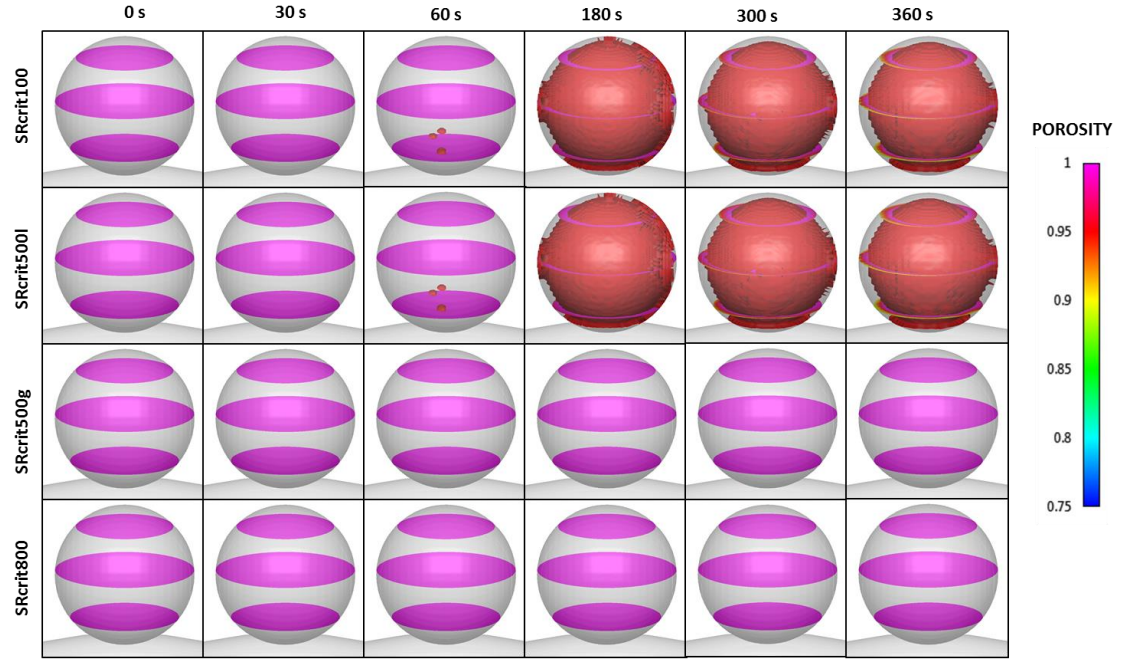


Figure 2: Clot growth for different initiation strain rate thresholds. Below or above a given threshold on the aneurysm wall, tissue factor is expressed.

The clotting process can be initiated for different tissue factor concentrations as seen in Figure 3. An increase in tissue factor concentration results in faster clot growth, however the difference in clot growth becomes less significant as concentration increases.

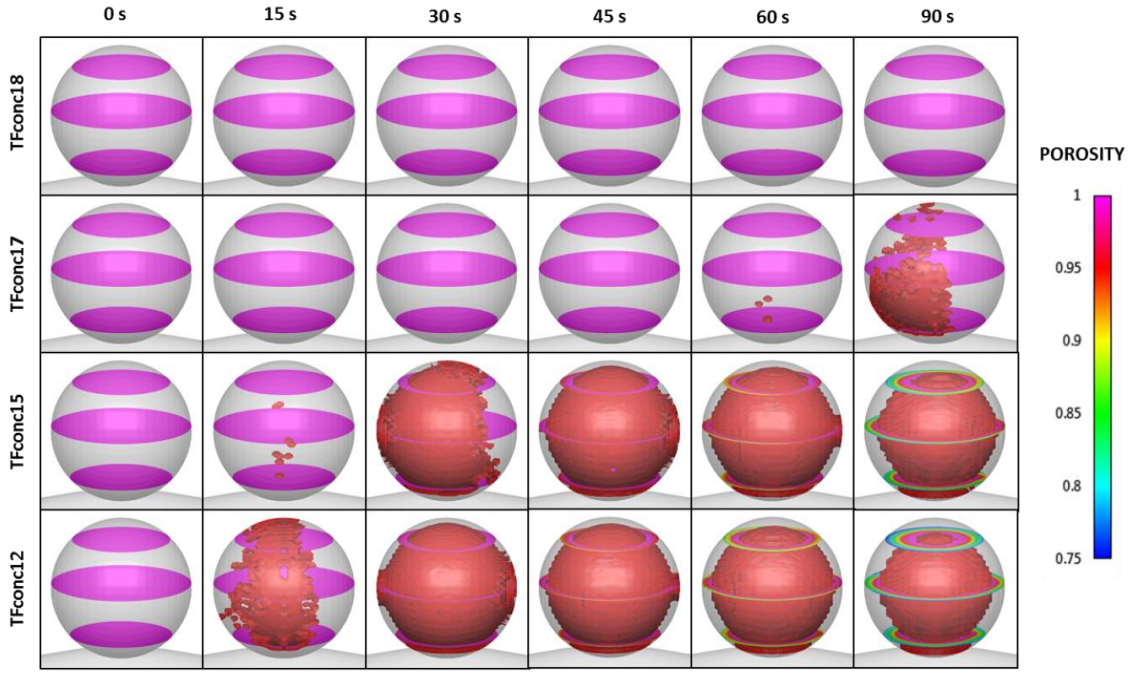


Figure 3: Clot growth for different starting concentrations of tissue factor.

3.2 Clot progression

Clot propagation velocity can be linked to mechanical variables as illustrated in Figure 4. Linking direct strain rate results in very rapid occlusion of the aneurysm sac, but also results in clot growth in the aneurysm sac. An inverse relationship between propagation velocity and strain rate results in an asymmetrical clot in the aneurysm sac.

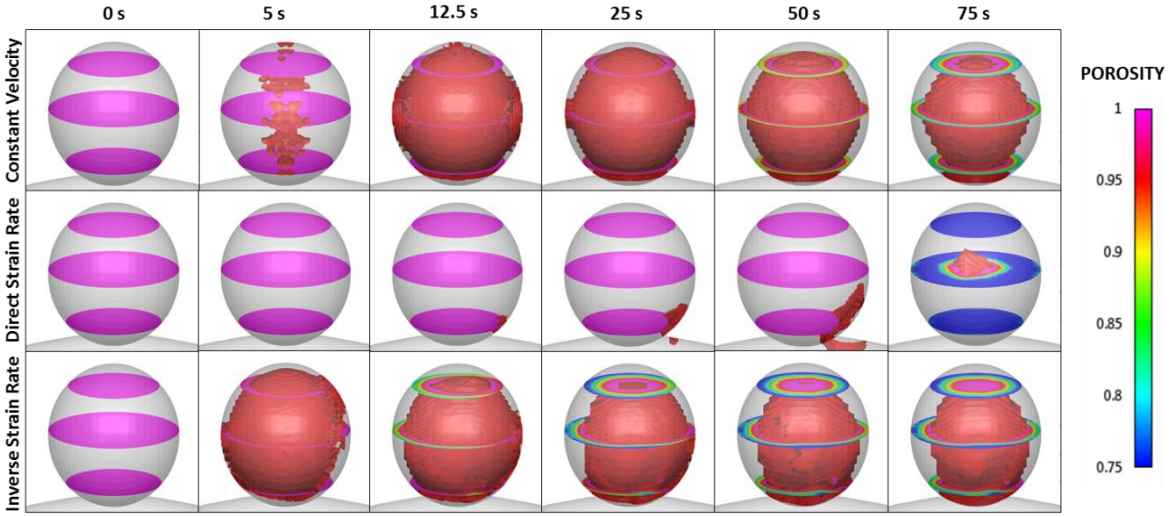


Figure 4: Clot growth when propagation velocity is linked to different mechanical variables.

When clot propagation velocity is linked to the concentrations of different biochemical species, the result produced, regardless of the species, is fairly uniform. This is illustrated in Figure 5. A time lag can be seen when comparing the results of thrombin and tenase because thrombin can only form once tenase is formed.

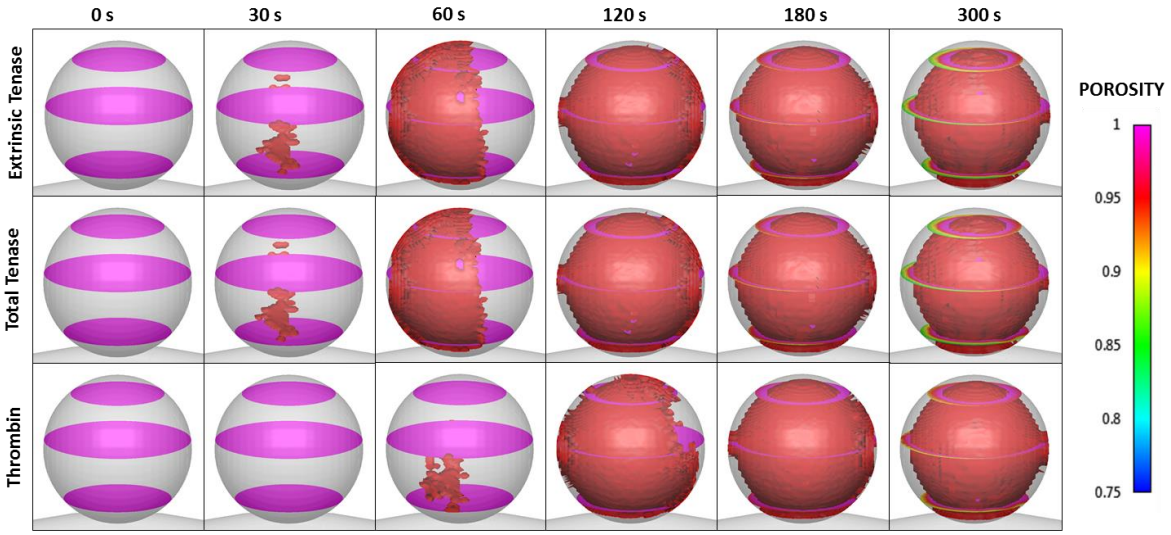


Figure 5: Clot growth when propagation velocity is linked to the concentrations of different biochemical species.

4 DISCUSSION

Clot initiation can be regulated either by controlling the initial tissue factor concentration or by regulating the strain rate at which tissue factor is expressed. From the results presented

in the previous section, it seems that the strain rate threshold has a much greater impact on whether or not a region is able to support clot growth. For strain rates greater than $500s^{-1}$, no clotting is observed. Savage et al found that clot deposition governed by fibrin typically occurred at lower shear rates [19]. In this study, the clotting process is governed by the biochemical cascade which eventually results in the formation of fibrin. The results observed therefore support the assertion made by Savage et al. The initial tissue factor concentration seems to have a greater impact on the rate at which the clot subsequently develops, but does not seem to influence whether or not a clot forms. Other studies have also observed that clot initiation in aneurysms seems to correlate with regions of low strain rate [11], [12].

Once clot growth has been initiated, propagation of the clot seems to be governed more by biochemical variables. The results presented show a large variation in clotting outcome for different mechanical variables. The type of clot growth into the parent vessel as illustrated by the direct strain rate model is something that is rarely observed *in vivo*. On the contrary, the biochemical models illustrate typical clotting patterns [1]. The formation and consumption of specific biochemical species in specific places has a significant impact on the stability of the clot [20]. From their studies, Panteleev et al found that extrinsic tenase played a key role in earlier clot propagation, while intrinsic tenase was more instrumental in advanced clot propagation [21]. Thrombin is known to play a key role in thrombus development as it has the ability to activate platelets and amplify its own production [22], [23].

5 CONCLUSION

Mechanical and biochemical factors contribute to clot initiation and propagation in cerebral aneurysms. From this exploratory qualitative study, it seems that clot initiation is more closely linked to mechanical factors, while propagation is more dependent on biochemical influences. More work would need to be conducted to verify these observations and to examine the combined effect of mechanical and biochemical factors in both clot initiation and propagation.

REFERENCES

[1] M. T. Lawton, A. Quiones-Hinojosa, E. F. Chang, and T. Yu, "Thrombotic Intracranial Aneurysms: Classification Scheme and Management Strategies in 68 Patients," *Neurosurgery*, vol. 56, no. 3, pp. 441–454, Mar. 2005.

[2] L. Roccatagliata, P. Guédin, S. Condette-Auliac, S. Gaillard, F. Colas, A. Boulin, A. Wang, S. Guieu, and G. Rodesch, "Partially thrombosed intracranial aneurysms: symptoms, evolution, and therapeutic management.," *Acta Neurochir. (Wien).*, vol. 152, no. 12, pp. 2133–42, Dec. 2010.

[3] J. De Vries, J. Boogaarts, A. Van Norden, and A. K. Wakhloo, "New generation of Flow Diverter (surpass) for unruptured intracranial aneurysms: a prospective single-center study in 37 patients.," *Stroke.*, vol. 44, no. 6, pp. 1567–77, Jun. 2013.

[4] J. E. Cohen, E. Itshayek, E. Yitshayek, J. M. Gomori, S. Grigoriadis, G. Raphaeli, S. Spektor, and G. Rajz, "Spontaneous thrombosis of cerebral aneurysms presenting with ischemic stroke.," *J. Neurol. Sci.*, vol. 254, no. 1–2, pp. 95–8, Mar. 2007.

[5] J.-N. Vallée, L. Pierot, A. Bonafé, F. Turjman, P. Flandroy, J. Berge, G. Rodesch, and S. Bracard, "Endovascular treatment of intracranial wide-necked aneurysms using

three-dimensional coils: predictors of immediate anatomic and clinical results.," *AJNR. Am. J. Neuroradiol.*, vol. 25, no. 2, pp. 298–306, Feb. 2004.

[6] J. V Byrne, R. Beltechi, J. a Yarnold, J. Birks, and M. Kamran, "Early experience in the treatment of intra-cranial aneurysms by endovascular flow diversion: a multicentre prospective study.," *PLoS One*, vol. 5, no. 9, pp. 1–8, Jan. 2010.

[7] V. M. Gonçalves, N. Cristino, and M. C. Sá, "Case Report Spontaneous Thrombosis in Giant Aneurysm of the Anterior Communicating Artery Complex in Pediatric Age : Five-Year Follow-Up," *HindawiiCase Reports Vasc. Med.*, vol. 2014, 2014.

[8] I. R. Whittle, N. W. Dorsch, and M. Besser, "Spontaneous thrombosis in giant intracranial aneurysms.," *J. Neurol. Neurosurg. Psychiatry*, vol. 45, no. 11, pp. 1040–7, Nov. 1982.

[9] P. Di Achille, G. Tellides, C. A. Figueroa, and J. D. Humphrey, "A haemodynamic predictor of intraluminal thrombus formation in abdominal aortic aneurysms," *Proc. R. Soc. London A Math. Phys. Eng. Sci.*, vol. 470, no. 2172, Oct. 2014.

[10] J. Biasetti, P. G. Spazzini, J. Swedenborg, and T. C. Gasser, "An integrated fluid-chemical model toward modeling the formation of intra-luminal thrombus in abdominal aortic aneurysms," *Front. Physiol.*, vol. 3, no. July, pp. 1–16, Jan. 2012.

[11] V. L. Rayz, L. Boussel, L. Ge, J. R. Leach, a J. Martin, M. T. Lawton, C. McCulloch, and D. Saloner, "Flow residence time and regions of intraluminal thrombus deposition in intracranial aneurysms," *Ann. Biomed. Eng.*, vol. 38, no. 10, pp. 3058–69, Oct. 2010.

[12] R. Ouared, B. Chopard, B. Stahl, D. a. Rüfenacht, H. Yilmaz, and G. Courbebaisse, "Thrombosis modeling in intracranial aneurysms: a lattice Boltzmann numerical algorithm," *Comput. Phys. Commun.*, vol. 179, no. 1–3, pp. 128–131, Jul. 2008.

[13] M. Ngoepe and Y. Ventikos, "Computational modelling of clot development in patient-specific cerebral aneurysm cases," *J. Thromb. Haemost.*, vol. 14, no. February 2016, pp. 262–272, 2016.

[14] G. Mulder, A. C. B. Bogaerds, P. Rongen, and F. N. Vosse, "On automated analysis of flow patterns in cerebral aneurysms based on vortex identification," *J. Eng. Math.*, vol. 64, no. 4, pp. 391–401, Mar. 2009.

[15] S. L. Diamond and S. Anand, "Inner clot diffusion and permeation during fibrinolysis.," *Biophys. J.*, vol. 65, no. 6, pp. 2622–43, Dec. 1993.

[16] R. Wagenvoord, P. W. Hemker, and H. C. Hemker, "The limits of simulation of the clotting system.," *J. Thromb. Haemost.*, vol. 4, no. 6, pp. 1331–8, Jun. 2006.

[17] E. Hairer and G. Wanner, *Solving ordinary differential equations II. Stiff and differential-algebraic problems*. Springer, 1993.

[18] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, 1st ed. New York: Springer, 2003.

[19] B. Savage, E. Saldívar, and Z. M. Ruggeri, "Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor," *Cell*, vol. 84, no. 2, pp. 289–97, Jan. 1996.

[20] S. Falati, P. Gross, G. Merrill-Skoloff, B. C. Furie, and B. Furie, "Real-time in vivo imaging of platelets, tissue factor and fibrin during arterial thrombus formation in the

mouse.,” *Nat. Med.*, vol. 8, no. 10, pp. 1175–81, Oct. 2002.

[21] M. a Panteleev, M. V Ovanesov, D. a Kireev, A. M. Shibeko, E. I. Sinauridze, N. M. Ananyeva, A. a Butylin, E. L. Saenko, and F. I. Ataullakhanov, “Spatial propagation and localization of blood coagulation are regulated by intrinsic and protein C pathways, respectively,” *Biophys. J.*, vol. 90, no. 5, pp. 1489–500, Mar. 2006.

[22] B. Furie and B. C. Furie, “In vivo thrombus formation,” *J. Thromb. Haemost.*, vol. 5, no. Supplement S1, pp. 12–17, 2007.

[23] M. Hoffman, “A cell-based model of coagulation and the role of factor Vlla,” *Blood Rev.*, vol. 17, pp. S1–S5, 2003.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

TECHNIQUES TO INTEGRATE PATIENT-SPECIFIC SIMULATION OF ANEURYSMAL BLOOD FLOW INTO THE CLINICAL WORKFLOW

Gábor Závodszky[1], Roland Joó-Kovács[1], György Paál[1] and István Szikora[2]

[1]Budapest University of Technology and Economics
Műegyetem rkp. 3., Budapest, 1111
e-mail: {gzavodszky,rjoo-kovacs, gypaal}@hds.bme.hu

[2] National Institute of Clinical Neurosciences
Amerikai str. 57, Budapest, 1145
e-mail: szikoraistvan@me.com

Keywords: CFD, GPU, Blood flow, Aneurysm, Clinical workflow

Abstract. *The CFD simulation of the emergent flow field inside vessel malformations is a generally employed technique during research processes. The results of an accurate numeric calculation might be used for risk assessment or for predictive purposes, such as the analysis of the possible pathogenesis. Even though several generally accepted numerical methods exist which are applied frequently during research, the current mainstream methods carry specific properties that significantly limit their application in everyday medical practice. We propose a well-selected set of methodologies to provide a complete solution that is able overcome several of these shortcomings by making the preparation of the numerical mesh more automated while reducing the average computational time to a clinically relevant time-scale on a single desktop workstation. A study involving 10 patients with intracranial aneurysm was carried out using the presented fast simulation method. The sole human interaction required for the process, that is, the selection of the boundary conditions might add a few minutes at most to the total investigation time required for one patient. The possibility of modelling the effects of a virtually implemented stent without losing the performance benefits is also discussed.*

1 INTRODUCTION

Intracranial aneurysms are focal dilatations of branches of the Circle of Willis, most frequently appearing around bifurcation points. The pathogenesis, growth, and rupture of this malformation is still not fully understood. The analysis of the flow parameters (e.g., wall shear stress, pressure, oscillatory shear index) may play a critical role in understanding these processes. The application of computational fluid dynamics (CFD) in vascular research is often considered to be the best method to non-invasively extract information about the flow field. Amongst others, the research of cerebral aneurysms relies heavily on such techniques [1, 2, 3]. A further improvement would be to incorporate these methods not only in research activities but also during clinical practice. However, there are two major difficulties preventing an effective solution. The first one is the setup of the patient-specific numerical model, including the proper description of the lumen surface geometry, the creation of the numerical mesh, and finally the selection and initialisation of the appropriate boundary conditions. Although the steps to perform such operations accurately are well-established [4, 5], they certainly demand the interaction of a specialist at several points. The second problem that follows is the significant amount of required computational time to carry out an appropriately detailed transient simulation. A possible solution to the latter one is the application of super-computers, however, that can in return further augment the first problem. While these are quite feasible in research processes, they usually do not fit the requirements of an everyday clinical protocol. In the current work a feasible solution is presented based on the combination of a carefully selected set of existing technologies.

2 METHODS

In this study the triangulated lumen surface geometry is considered as first input. Most of the recent clinical workstations are capable of exporting vessel sections in this format. Note that this segmentation step usually requires human interaction and presents the scientifically most questionable part of the pipeline since the quality of the resulting surface depends strongly on subjective decisions. The automation of this step is the subject of several current researches. The vessel geometries for our study were obtained from DSA-Angiography records.

2.1 The CFD method

Since both performance and automation are high priorities, the lattice Boltzmann method was chosen as a CFD numerical scheme, which is a highly parallel numerical scheme known to be able to simulate aneurysmal flows accurately [6]. Its regular numerical grid allows for an automatic setup of the simulation domain without the need for a specialist to verify the quality of the generated numerical mesh. The LBM method also carries a few properties that makes it very appealing in today's computational environments: it is inherently parallel due to being explicit in time and local in space. It was shown to be capable of simulating large-scale cardiovascular systems [7] while exhibiting linear scaling behaviour over several GPUs [8]. To take full advantage of the parallelism, the applied solver was based on an open source LBM code [9] implemented for graphics processing units (GPUs) which excel at high performance parallel code execution. This solver relies heavily on meta-programming through the usage of a templating language to generate the GPU specific code part at run-time. This approach helps to keep the implementation of the physical description of the dynamics closer to the mathematical definition and hide some of the complications of the actual hardware specific implementation. An indirect addressing for lattice cells was used, which increased the computational time mainly

due to the overhead added by pointer dereferencing, however, it was necessary to reduce the memory requirements of storing the LBM distribution functions. The MRT relaxation model [10] was chosen to represent the collision dynamics which is known to be able to produce accurate results in the case of highly curved geometries [6]. The performance and the scaling of the code was first tested using a single computational node equipped with four Tesla C2070 cards, using double floating point precision, two Intel Xeon 5650 CPUs and 48GB or RAM. Figure 1 shows the performance results in 10^6 lattice updates per second (MLUPS). The visible drop of performance using three cards is the result of the simple domain decomposition algorithm which caused some imbalance in the computational distribution.

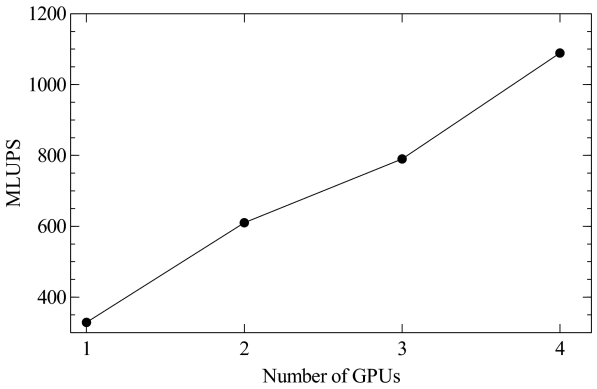


Figure 1: The strong scaling of the GPU implementation over four Tesla cards in a single computing node.

2.2 Model preparation

The surface description is turned into a regular grid by means of voxelisation. In this work the algorithms of the open source CVMLCPP library [1] have been used for this purpose. During the process one can prescribe the required resolution of the numerical grid. A higher resolution naturally leads to higher accuracy although care should be taken not to choose orders of magnitude smaller spatial steps than the original resolution of the angiographic record. In our cases the spatial steps were between 3 - 6×10^{-5}m. The solid wall and fluid cell array can be efficiently generated from the voxelised geometry using circular shifts and basic Boolean algebra.

The Newtonian rheology model was used since it was shown [11] that in most cases it does not lead to significant differences compared to the non-Newtonian rheology described by the Casson model. However, the utilisation of non-Newtonian rheology introduces only a small computational overhead, and as such, the findings of this works remain valid in this case as well.

2.3 Boundary conditions

A rigid wall model was imposed during the simulations. From one hand, the main motivation of this decision was the nature of the available medical records as they represent a time-averaged information on the vessel lumen. On the other hand, a simulation with flexible walls would significantly increase the computational costs. Currently, this options seems to be out of bounds

[1] http://tech.unige.ch/cvmlcpp/ - version 2012-04-22

for a single-workstation computational environment when a clinically relevant time-scale is necessary (that is, the time-scale of minutes rather than hours). For representing the walls in the CFD solver, the half-way bounce back algorithm was used but it is possible to change it to Guo's off-lattice boundaries [12] for higher accuracy.

The prescribed shape of the velocity profile on the inlet can also have significant influence on the resulting flow field. In the current work the same artificial volumetric inflow function was used as in [6]. The Reynolds number was set to the value of 450 at the inlet in all the cases. To recover the correct inlet shape in vessel geometries it is common to create an elongated artificial lead-in vessel section in which the accurate velocity profile can develop. This, however, leads to increased computational costs. To circumvent this, the inlet velocity vector profiles can be calculated as follows: the direction of the inlet velocity vector is chosen so that it is parallel with the beginning section of the centreline of the vessel; the length of the vector is calculated using the sum of the Womersley components. The inlet function is usually given as a periodic volume flow function representing the flow generated by the cardiac system, therefore it can be expanded using the Fourier series:

$$Q(t) = \sum_{n=0}^{N} q_n e^{in\omega t}. \tag{1}$$

Combining these, the Womersley velocity profile of every Fourier mode can be calculated and their linear sum produces the answer of the fluid system for the inlet excitation [13]:

$$u(r,t) = \frac{2q_0}{\pi R^2}\left[1-(\frac{r}{R})^2\right] + \sum_{n=1}^{N} \frac{q_n}{\pi R^2}\left[\frac{1-\frac{J_0(\beta_n \frac{r}{R})}{J_0(\beta_n)}}{1-\frac{2J_1(\beta_n)}{\beta_n J_0(\beta_n)}}\right] e^{in\omega t}, \tag{2}$$

where

$$\beta_n = i^{\frac{3}{2}}\alpha_n = i^{\frac{3}{2}} R\sqrt{\frac{n\omega}{\nu}}. \tag{3}$$

In this expression u is the length of the axial velocity vector at a given r radius, R is the radius of the pipe, J_0 and J_1 denotes the zeroth and first order Bessel functions, ν is the kinematic viscosity, ω is the angular frequency, α is the Womersley number, and finally n denotes the ordinal number of the harmonic component from the Fourier expansion.

On the outlets constant pressure is imposed on a surface perpendicular to the outlet centerline section. This might be improved by imposing a resistance model [14] based on the idea that the capillary level can be regarded as a common equipressure point and the vessel network leading to it can be represented as hydrodynamic resistance.

To model the effect of an implanted stent is generally a computationally intensive task. A numerically effective solution is to represent it from a macroscopic viewpoint as a porous layer representing some resistance on the flow [15]. The Darcy equation describes such porous materials. Since the intra-aneurysmal flow can be regarded as a high velocity flow, the equation should be extended with the Forcheimer term [16] to account for the nonlinear behaviour:

$$\bigtriangledown p = -\frac{\mu}{\kappa}\vec{v} - \frac{\rho}{\kappa_1}|\vec{v}|\vec{v}, \tag{4}$$

where κ denotes the permeability, κ_1 is the inertial permeability and μ is the dynamic viscosity. Naturally, to acquire a proper value for κ and κ_1 for a real implanted stent is not an easy task. The final resistance is dependent on many factors, such as the structure of the stent, the implantation technique, and the local curvature. The authors have an ongoing research in this direction.

3 RESULTS

The workflow being investigated consists of several steps. Firstly, an automatic voxelisation of the input lumen surface geometry is carried out. The run-time of this step represents a few seconds time at most, thus it can be regarded as negligible. The following stage is to select the appropriate inlet and outlet boundary conditions. This can be done in any 3D editor supporting voxel geometries (such as 3DSlicer[2]) by labelling the openings. This is the only step requiring human interaction and based on the number of openings might take up to a few minutes. Finally, the CFD simulation of the model takes place. Generally, this is the most time-consuming part, therefore, the run-times of several cases were evaluated. Depending on the use-case, either a stationary velocity snapshot or the full transient cardiac cycle might be required. Both cases were tested for the expected run-times. Since the LBM method is inherently transient, the stationary simulations were actual transient computations executed until the flow inside the whole domain reached its equilibrium. Every computation starts with a "warm-up" phase during which the fluid is slowly accelerated from the initial zero velocity to the stationary velocity that corresponds to the beginning of the inlet volume flow function. This ensures a good initial approximation of the velocity profiles within the domain without having initial transients. In our case, the run-times for both stationary and transient situations were evaluated on a high-end portable notebook (Intel i7-4980HQ, 16GB RAM, nVidia Geforce 980M) and on a desktop workstation (Intel i7-3770K, 32GB RAM, nVidia Geforce Titan Black) for ten intracranial aneurysmal geometries corresponding to different patients. Two different numerical grid resolutions were generated for every geometry. For the lower resolution one the target number of numerical mesh elements were 1 million while for the high resolution it was 9 million (this is the number of useful cells, on average accounting for around 20% of the total number of cells residing in the encompassing prismatic domain). To access the GPUs the CUDA 7.5 API was used. Table 1 shows the average execution times required by the CFD computations.

	Notebook [mm:ss]	Workstation [mm:ss]	
		Single	Double
low resolution (1M)	1:17 ($\sigma = 0.27$)	1:01 ($\sigma = 0.25$)	4:58 ($\sigma = 0.31$)
high resolution (9M)	10:05 ($\sigma = 0.61$)	8:13 ($\sigma = 0.63$)	23:05 ($\sigma = 0.75$)

Table 1: Run-time averages of the transient computations of one full cardiac cycle for the ten different geometries with the sample standard deviations in minutes. The hardware of the notebook allowed for single precision only while the computations on the workstation were run using both single (32 bit) and double (64 bit) floating point representation precisions.

The stationary computations typically required about 60 % of the transient run-times, since the "warm-up" phase is necessary here as well. Finally, the user interface of the solver is implemented in Python language[3]. For visualisation purposes, it embeds an instance of a scripted MayaVi [17] application. During the computation of a transient simulation, the already calculated time steps are moved to the host memory from the GPU. This makes them immediately available for vector or streamline presentation, thus, it is not necessary to wait for the full computation to begin evaluating the flow field properties. Figure 2 shows a snapshot of the working application.

[2]https://www.slicer.org/

[3]https://www.python.org/

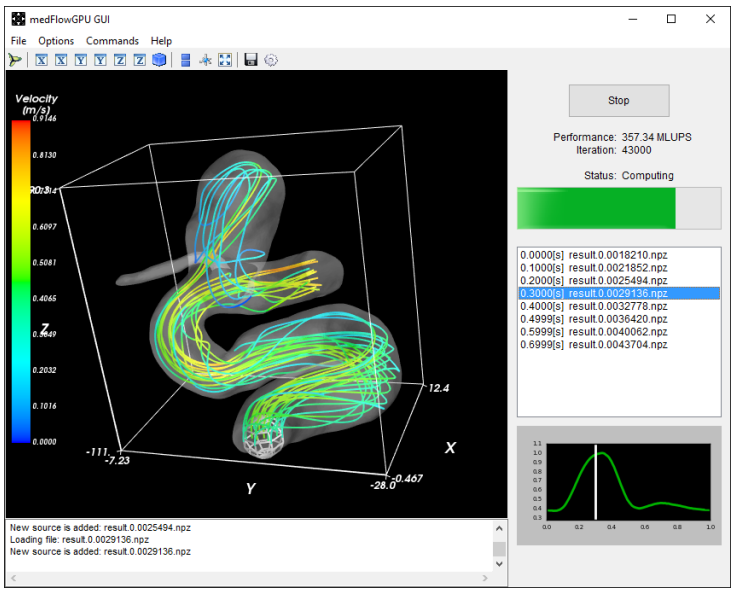


Figure 2: The software in work on a notebook with the already computed time segments available for evaluation.

4 CONCLUSIONS

The resulting run-time values of around 10 minutes for the high-resolution models using single precision are well within the clinically relevant time-scale. We should point out here that even though the presented set of techniques are evaluated in the case of aneurysmal blood flows, they do not build on any information specific to these cases, henceforth, they are applicable to a wider range of fluid flows (e.g., carotid and coronary arterial flows and other vascular flows where the same modelling approximations are valid). Thus, it can be concluded that with the adequate set of methods the fluid dynamics computation of a real aneurysm can be carried out as part of a regular clinical examination workflow even on a portable notebook. The only phase depending on human interaction is confined to the selection of the appropriate boundary conditions for the given case. This step, however, can be carried out using general clinical 3D image processing tools and does not build on the knowledge of any specialised software.

5 ACKNOWLEDGEMENT

The project was supported by the KTIA_NAP_13-1-2013-0001 Hungarian Brain Research Program.

REFERENCES

[1] J. R. Cebral, M. A. Castro, J. E. Burgess, R. S. Pergolizzi, M. J. Sheridan, and C. M. Putman, "Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models," *American Journal of Neuroradiology*, vol. 26, no. 10, pp. 2550–2559, 2005.

[2] G. Paál, A. Ugron, I. Szikora, and I. Bojtár, "Flow in simplified and real models of intracranial aneurysms," *International Journal of Heat and Fluid Flow*, vol. 28, no. 4, pp. 653–664, 2007.

[3] I. Szikora, G. Paal, A. Ugron, F. Nasztanovics, M. Marosfoi, Z. Berentei, Z. Kulcsar, W. Lee, I. Bojtar, and I. Nyary, "Impact of aneurysmal geometry on intraaneurysmal flow: a computerized flow simulation study," *Neuroradiology*, vol. 50, no. 5, pp. 411–421, 2008.

[4] D. M. Sforza, C. M. Putman, and J. R. Cebral, "Hemodynamics of cerebral aneurysms," *Annual review of fluid mechanics*, vol. 41, p. 91, 2009.

[5] L. Campo-Deaño, M. S. Oliveira, and F. T. Pinho, "A review of computational hemodynamics in middle cerebral aneurysms and rheological models for blood flow," *Applied Mechanics Reviews*, vol. 67, no. 3, p. 030801, 2015.

[6] G. Závodszky and G. Paál, "Validation of a lattice Boltzmann method implementation for a 3D transient fluid flow in an intracranial aneurysm geometry," *International Journal of Heat and Fluid Flow*, vol. 44, no. 0, pp. 276 – 283, 2013.

[7] S. Melchionna, M. Bernaschi, S. Succi, E. Kaxiras, F. J. Rybicki, D. Mitsouras, A. U. Coskun, and C. L. Feldman, "Hydrokinetic approach to large-scale cardiovascular blood flow," *Computer Physics Communications*, vol. 181, no. 3, pp. 462–472, 2010.

[8] M. Bisson, M. Bernaschi, S. Melchionna, S. Succi, and E. Kaxiras, "Multiscale Hemodynamics Using GPU Clusters," *Communications in Computer Physics*, vol. 11, pp. 48–64, 2012.

[9] M. Januszewski and M. Kostur, "Sailfish: a flexible multi-gpu implementation of the lattice boltzmann method," *Computer Physics Communications*, vol. 185, no. 9, pp. 2350–2368, 2014.

[10] D. D'Humières, I. Ginzburg, M. Krafczyk, P. Lallemand, and L.-S. Luo, "Multiple-relaxation-time lattice Boltzmann models in three dimensions.," *hilosophical Transactions of the Royal Society - Series A: Mathematical, Physical and Engineering Sciences*, vol. 360, no. 1792, pp. 437–451, 2002.

[11] M. A. Castro, M. C. A. Olivares, C. M. Putman, and J. R. Cebral, "Unsteady wall shear stress analysis from image-based computational fluid dynamic aneurysm models under newtonian and casson rheological models," *Medical & biological engineering & computing*, vol. 52, no. 10, pp. 827–839, 2014.

[12] Z. Guo, C. Zheng, and B. Shi, "An extrapolation method for boundary conditions in lattice boltzmann method," *Physics of Fluids (1994-present)*, vol. 14, no. 6, pp. 2007–2010, 2002.

[13] J. R. Cebral, M. A. Castro, S. Appanaboyina, C. M. Putman, D. Millan, and A. F. Frangi, "Efficient pipeline for image-based patient-specific analysis of cerebral aneurysm hemodynamics: technique and sensitivity," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 4, pp. 457–467, 2005.

[14] G. Janiga, P. Berg, O. Beuing, M. Neugebauer, R. Gasteiger, B. Preim, G. Rose, M. Skalej, and D. Thévenin, "Recommendations for accurate numerical blood flow simulations of stented intracranial aneurysms," *Biomedizinische Technik/Biomedical Engineering*, vol. 58, no. 3, pp. 303–314, 2013.

[15] L. Augsburger, P. Reymond, D. Rufenacht, and N. Stergiopulos, "Intracranial stents being modeled as a porous medium: flow simulation in stented cerebral aneurysms," *Annals of biomedical engineering*, vol. 39, no. 2, pp. 850–863, 2011.

[16] D. Ruth and H. Ma, "On the derivation of the forchheimer equation by means of the averaging theorem," *Transport in Porous Media*, vol. 7, no. 3, pp. 255–264, 1992.

[17] P. Ramachandran and G. Varoquaux, "Mayavi: 3d visualization of scientific data," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 40–51, 2011.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PROXIMAL ISOVELOCITY SURFACE FOR DIFFERENT MITRAL VALVE HOLE GEOMETRIES

A. This[1], H. G. Morales[1], and O. Bonnefous[1]

[1]Philips Research Paris
Medisys
33, Rue de Verdun, 92156 Suresnes, France
e-mail: {alexandre.this,hernan.morales,odile.bonnefous}@philips.com

Keywords: Mitral regurgitation, Proximal isovelocity surface area, Computational Fluid Dynamics

Abstract. *The mitral valve can be subject to several pathologies resulting in blood regurgitation from the ventricle to the atrium. The proximal isovelocity surface area (PISA) method is widely used to assess the severity of the regurgitation. However, it relies on several assumptions on both fluid motion and morphology. We designed numerical simulations to assess the influence of breaking those assumptions on the PISA model. A simplified heart geometry was created and three different defects were introduced in the mitral valve. Four heart cycles were simulated using a fluid/structure interaction method. The simulated results were then compared with the estimation provided by the PISA method. In the simulation presented, the mitral defect shapes strongly affects the shape of the isovelocity surface and this results in important estimation error of clinically relevant quantities. Moreover, it was found that the aortic outflow influenced the regurgitation. These observations would justify the use of patient-specific models for clinical evaluations.*

1 Introduction and previous work

The mitral complex is located between the left atrium and the left ventricle. It is composed of the mitral annulus, the mitral leaflets, the chordae tendinae and the papillary muscles [13]. The physiological functionning of the mitral valve is a result of complex biomechanical interactions. Its role is to prevent blood from going back in the atrium during the ventricular systole. Due to various pathologies such as annulus calcification, damaged chordae or ventricle hypertrophy,for example, the mitral valve can behave unproperly thus allowing blood to leak back into the atrium at high velocities (up to several meters per seconds [4]). This behavior is called mitral valve regurgitation (**MR**). When the blood leak is too important, physicians have to choose an appropriate course of action for the patient treatment. Alternatives include for example the replacement of the valve, the placement of a mitral clip, or the repair of the chordae.

To assess the severity of the mitral regurgitation, quantitative measurements have been designed, namely the Effective Regurgitant Orifice Area (**EROA** , the area that is subject to regurgitation on the mitral valve), the Vena Contracta Width/Area (**VCW/VCA**, ie the width/area of the narrowest portion of the regurgitant jet of blood in the atrium) or the Regurgitant Fraction (**RF**, ie the percentage of blood that went to the ventricle and that was regurgitated). Guidelines have been given to physicians to assess the severity of the regurgitation depending on those measurements. For example, a regurgitation with an EROA of more than 0.40 cm^2 is associated with a severe MR [3] and it is recommended to perform a surgical intervention in this case [8].

Depending on the image modality, several techniques can be used to assess MR. One of the most commonly used technique is based on color-Doppler echocardiographic images and is called the proximal isovelocity surface area (**PISA**) method [15]. In particular, it allows the estimation of the EROA and RF based on several hydrodynamic principles and morphological simplifications. Assuming that the mitral valve defect is small, circular and lies on an infinite planar surface, the velocity field of an inviscid fluid forms hemispheric isovelocity concentric shells of increasing value and decreasing area when converging to the hole. The PISA method requires the physician to 1) locate the center of the holes in the mitral valve, 2) select a velocity $|\mathbf{u}_a|$ and 3) find the distance between the center of the hole and the isovelocity surfaces of the selected velocity. Using those parameters, the physician can then calculate the isovelocity surface area ($2\pi r^2$) and the instantaneous regurgitant flow ($2\pi r^2|\mathbf{u}_a|$). The purpose of this work is to evaluate the validity of the PISA model when the assumptions are violated by assuming more realistic scenarios such as a non circular hole or the presence of a neighboring aortic flow.

1.1 Previous work : CFD analysis of the MR

In 1992, [12] used a simple 2D CFD model to assess the validity of the PISA model when the regurgitant hole in the mitral valve is not infinitesimally small, breaking one of the assumptions of the PISA model. They observed that the expected hemispheric isovelocities were flattened in the vicinity of the hole and proposed a correction factor to take that deformation into account.

Yap et al. [17] presented a novel method to quantify mitral regurgitation. Using 3D echography, they proposed an automatic algorithm claimed to reduce the variability of the measures. The authors used a similar simple 2D CFD model to validate the flow velocity scheme assumed by the PISA model. They modeled a regurgitation with a finite sized hole and several valve leaflets configurations. They concluded that the flow velocity scheme of the PISA model was generally valid for the tested configurations although there is a strong error in the orientation of the flow in the vicinity of the hole.

3D CFD of simplified mitral valve geometries was performed by [14] to assess the possibility

of using CFD to help characterize mitral regurgitation and improve treatments. In vitro experiments were performed and a CFD model with equivalent geometry and boundary conditions was used to validate the CFD model. It was shown that realistic three dimensional data could be obtained. They also stressed that the use of CFD allows the investigation of data that are not obtainable using traditional echography.

A similar study was performed by Quaini et al. [11]. On top of validating a 3D CFD model with in vitro experiments, the PISA model was compared with the simulated data. It was shown that the reliability of the PISA measure is highly dependent of the shape of the holes. They also proposed the use of CFD to evaluate the limitations of the ultrasound (**US**) acquisition on the quantitative measures of mitral regurgitation. The same authors went further by introducing a 3D Fluide-Structure interaction (**FSI**) model where the mitral valves were considered as an elastic solid [10]. The model was validated using in vitro experiments. The goal of this last paper was solely focused on the validation of the model.

To summarize the current state of the art, it has been shown that reliable results can be obtained to perform flow analysis in the context of the mitral valve regurgitation using all kind of model complexity. Although PISA method is well known and has been used for decades, it is known that it leads to rough estimations. CFD simulations have helped quantify the drawbacks of the PISA method and provided useful data to build better estimation techniques for realistic holes and mitral valve configuration.

2 Materials and methods

A FSI simulation has been developed to simulate blood flow behavior in patients affected with MR. The heart model used consisted of three cylinders. The ventricle and the atrium were represented by two cylinders sharing the same diameter and were connected along their longitudinal direction. The connection (circular) between them represents the mitral valve. The third cylinder modeled the aorta and was intersecting the ventricle with an angle of 50°, which was similar to clinical measurements [7]. An internal circle in the aorta was placed to mimic the aortic valve. Figure 1 shows the simplified heart model that was used in this study. Diameters and length of each cylinder are presented in the table 1.

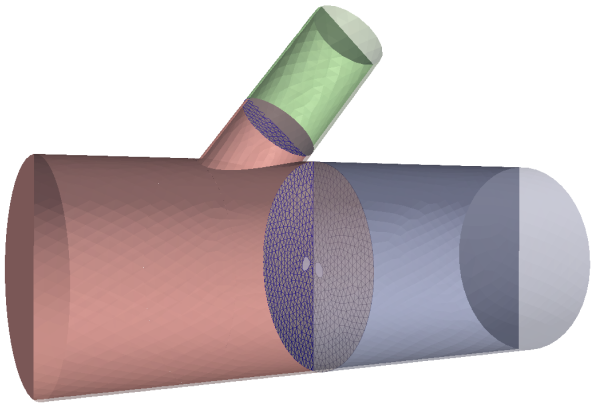


Figure 1: Heart model. The atrium, ventricle and aorta are represented by the blue, red and green cylinder respectively.

To model the impact of the hole shapes on the PISA estimation error, three different mitral valve were modeled (see figure 2). Sim #1 contains a round hole in the middle of the valve, Sim #2 contains an ellipsoid hole. Sim #3 contains a valve defect that was taken from segmented

ultrasound data of a human mitral valve. It contains two holes, which total area was coherent with reported values (≈ 0.3 cm^2). The holes of Sim #1 and Sim #2 holes have been designed so their total area was equal to the defect area of Sim #3 to facilitate the comparison among simulations.

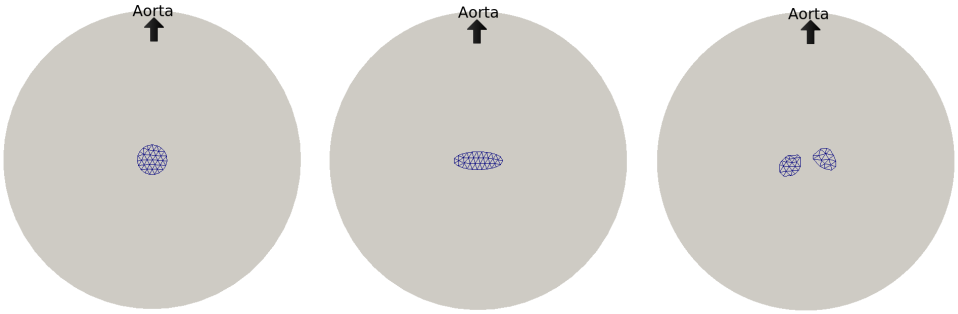


Figure 2: Hole configuration used in the presented work. The rightmost configuration is a mitral defect segmented from ultrasound images.

The heart geometry was then meshed with 3D tetrahedra using Salome (Version 7.7.1 : www.salome-platform.org) with specific refinements in the vicinity of the mitral valve defect. The fluid simulation was performed using the open-source software Openfoam (Version 2.4 : http://www.openfoam.com).

The kinematic viscosity ν was set to 3.5×10^{-6}m^2 s^{-1} [5]. It has been decided to use a Newtonian fluid model since blood can be modeled realistically with constant viscosity if the vessel diameter is greater than 0.5mm [2].

Physiological pressures were imposed as dirichlet boundary conditions for both the atrium and aortic outlet (11.5 mmHg and 100 mmHg respectively [9, 16]). The contraction and the relaxation of the ventricle was simulated by deforming the mesh and this solid boundary displacement was coupled with the CFD solver. To keep the deformation simple, the farmost wall was compressing the ventricle like a piston. The wall velocity was designed so that the ventricular volume evolution was coherent with physiological behaviors. As the amount of contraction was constrained by the aorta positioning, the fraction of blood ejected during systole was below physiological values (40% of the initial ventricular volume).

The aortic and mitral valves were opened and closed concurrently using user-defined condition. The aortic valves opens when the contraction of the ventricle starts and it closes when fluid is coming back into the ventricle. The mitral valve acted in the reverse way.

Simulations were initialized with zero velocity and pressure fields. To alleviate this erroneous initialization, four heart cycles were computed and only the results of the last cycle were considered for further analysis. The isovolumetric contraction and relaxation were not simulated. To ensure the independency of the results with respect to the mesh, four different meshes

Denomination	Reported value
Ventricle cylinder radius	3 cm
Aortic cylinder radius	1.2 cm
Atrium cylinder length	8 cm
End-diastolic ventricle length	8 cm
US-segmented defect area	0.3 cm^2

Table 1: Table of the geometrical dimensions of the model

of increased refinement (120k, 200k, 320k and 1300k elements) were investigated and the isovelocity surface area of each mesh was compared to the one of the most refined mesh. The third mesh was used to compute the results for all the cases presented in the next section as it showed convergence of the quantities of interest and was a good compromise as it was computationally accessible.

2.1 Assumption evaluation

For each simulations, two area measurements were made. First, the PISA estimation of the isovelocity surface area was obtained. A velocity magnitude $|\mathbf{u}_a| = 0.25$ m.s^{-1} was chosen to locate the isovelocity surfaces in all the PISA measurements, which was alined with clinical recommendations [3]. The second measurement was the isovelocity surface area that was provided by the CFD simulations for the same velocity magnitude. The chosen velocity $|\mathbf{u}_a|$ allows the simulated isovelocity surface to be in the region where the mesh was refined, thus allowing a more accurate estimation of this surface area.

3 Results

3.1 Simulation assessment and observations

The CFD simulations were able to reproduce physiological haemodynamics in the location of interests. During systole, blood convergence to the valve defects and blood jets in the atrium were observed. Key caracteristics of the mitral jet, such as the vena contracta, the region where the jet is the narrowest, were observed. As expected, the vena contracta occured just behind the mitral hole due to the fact that the fluid cannot abruptly change direction. The blood velocity at the mitral defect for the three simulations was approximately 3 m.s^{-1} which is comparable to reported clinical measurements [4]. In the case of the round mitral hole, the blood velocity is increasing as it approaches the hole and form hemispheric isovelocity surfaces, resulting in a good correlation between the CFD results and the hemispherical isovelocity surface assumption of the PISA methods. Nevertheless, several discrepancies between the PISA technique and the CFD results were found in the other defect configurations (Sim #2 and Sim #3).

3.2 Comparison of the simulation results with the pisa measurements

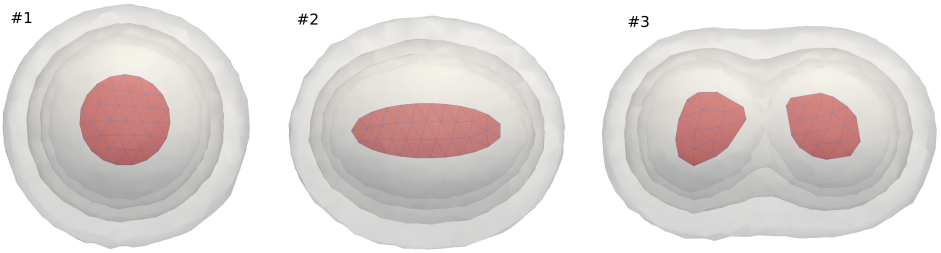


Figure 3: Isovelocity surfaces for the three hole configuration as a top view. Red surfaces represent the holes. The isovelocity area were hemispheric for Sim #1 but not for the two other cases.

Figure 3 depicts proximal isovelocity surfaces for the different hole configurations. Sim #1 exhibits successive hemispheric isovelocity surfaces. The PISA measurements was straightforward as the center of the round hole was not difficult to locate and the radius to perform the PISA measurement was easily picked. A good agreement between the simulations and the assumptions of the PISA method was found (less than 3% difference). For the sim #2, the

isovelocities were no longer hemispherical. As the distance between the center of the ellipsoid and the isovelocity surface are no longer constant, it is not possible to provide a single PISA estimation of the surface. It has been decided to bound the PISA measurement between its maximum and minimum possible values by taking the extremum distances between the center of the ellipsoid and the isovelocity surface. Finally, sim #3 involves two distinct holes in the mitral valve. In this case, the isovelocity surfaces separates in two distinct surfaces when approaching the holes. However, at the chosen isovelocity, a unique isovelocity surface is found for the two holes. As in the ellipse case, this isolvelocity surface is not hemispherical. Moreover, the choice of the center of the defect is harder because there are two holes. To perform the PISA estimation, a point centered between the two hole was taken. Table 2 reports the variation between the estimated PISA and the actual surface area.

Simulation	Surface area (cm^2)		Percentage difference
	Simulated	Pisa method	
#1	2.65	2.58	-3%
#2	2.61	2.18	-16%
		3.45	32%
#3	2.99	2.24	-25%
		4.47	49%

Table 2: Surface area and variation for a velocity $|\mathbf{u}_a| = 0.25$ cm.s^{-1}. The two reported PISA estimations for the sim #2 and #3 are values obtained by selecting the smallest and largest distance to the iso-surface. Percentage difference was calculated as $100 \times (1 - \frac{S_{Pisa}}{S_{Sim}})$

3.3 Aortic outflow influence

During the systole of an healthy heart, the blood is directed to the aorta. However, for a patient with mitral regurgitation, part of this blood is regurgitating back in the atrium through the mitral valve. It has been decided to used the simple round hole case to investigate the influence of the aortic outflow on the haemodynamics near the mitral valve. In particular, the PISA methods assumes that the blood is directed to the center of the hole. To verify this assumptions, the angle deviation between this assumption and the simulation has been computed on the simulated isovelocity surface at $|\mathbf{u}_a|$. As seen on figure 4, the distribution of this angle deviation is not uniform on the surface. At $|\mathbf{u}_a|$, a deviation of approximately 10° is exhibited while the distal part is closer to the PISA assumption. In the PISA method, this assumption is important because it is used to compute the flow through the isovelocity surface. If the velocity is not perpendicular to the surface, the relationship used to compute the flow through the surface doesn't hold anymore. In the specific case of a deviation $\theta = 10^\circ$, the error is less than 2% (the corrected flow was calculated as $2\pi r^2 |\mathbf{u}_a| \cos(\theta)$). Nevertheless, this effect should not put aside. If the defect was closer to the aorta, or if the isovelocity chosen was higher, this angle deviation could rise and thus impacting the PISA estimation even more.

4 Discussions and future work

In the presented cases, a significative difference between the isovelocity surface area of the holes and their respective PISA estimation were found. More precisely, it has been shown that the isovelocity surface area could be under or overestimated using the PISA method depending on the valve defects. Specifically, using an US segmented defect from a real mitral regurgita-

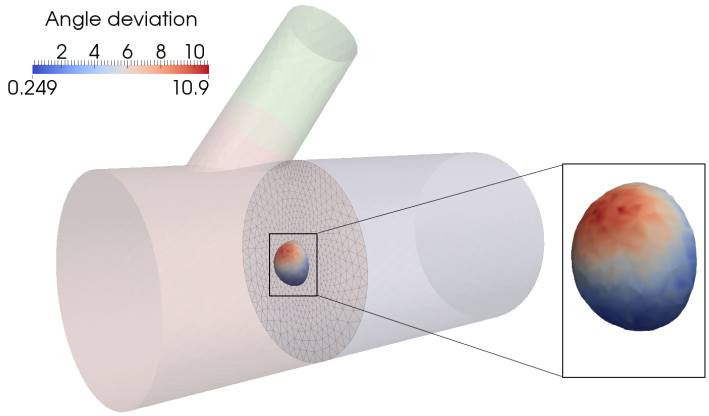


Figure 4: Angle deviation between the flow velocity and the PISA assumption (convergence toward the center of the hole). A deviation of the flow on the side of the aortic output can be observed.

tion, the estimation could have been under/overestimated by 25% to 50% respectively using the traditional PISA method. The variation of estimations of the isovelocity surface area is a result of the difficulty to choose the two parameters of the PISA estimation method. The patient-specific defect, for example, does not have a clear "center" definition. Moreover, choosing a distance between the *so-called* center and the isovelocity surface is not straightforward when the isovelocity surface is not hemispheric. These results corroborate previous observations [11].

In this work, it has been chosen to include a case where the mitral defect was extracted from clinical echocardiographic images. Using a more complex patient-specific model could be beneficial in the future as it could allow a more in-depth exploration of the mitral regurgitation pathology. This is important as a better understanding of the pathology and of haemodynamics can allow us to propose tools that provide more reliable quantitative assessment of the MR severity.

We found that the aortic outflow had an influence on the PISA estimation. Additional investigation needs to be conducted to understand its real influence. For example, it could be possible that the influence of aortic outflow varies depending on the aliasing velocity chosen as it means that the blood is closer or farther to the aortic root. This has not, to our knowledge, been highlighted in previous studies. As this haemodynamic behavior influence the evaluation of the mitral regurgitation, it is of interest to continue research further in this direction.

5 Conclusion

A one-way FSI model has been used to simulate three different mitral valve defects. It was exposed that differences in mitral holes geometries can have a great impact on the estimation of clinically relevant quantities (e.g. isovelocity surfaces) using traditional estimation methods. In particular, the hypothesis of hemispheric isovelocity surfaces close to the holes, which is the basis for the PISA methods, was shown to not hold in every cases. Moreover, thanks to the inclusion of an aortic output in the simulation, it was found that the aortic flow can also influence the estimations. The presented work is an example of the use of numerical techniques to improve the current clinical evaluation tools by improving our understanding of the haemodynamics and of the pathology.

6 Acknowledgements

The authors would like to acknowledge Medisys laboratory of Philips Healthcare which made this study possible in the context of an internship.

REFERENCES

[1] N. Amanifard and M. H. Behnam Rahbar. Numerical simulation of the mitral valve opening using smoothed particle hydrodynamics. In *Proceedings of the World Congress on Engineering*, volume 3, 2011.

[2] M. A. Fogel. *Ventricular function and blood flow in congenital heart disease*. John Wiley & Sons, 2008.

[3] P. A. Grayburn. How to measure severity of mitral regurgitation. *Heart*, 94(3):376–383, 2008.

[4] A. S. Lambert. Proximal isovelocity surface area should be routinely measured in evaluating mitral regurgitation: a core review. *Anesthesia & Analgesia*, 105(4):940–943, 2007.

[5] G. Lowe, A. Lee, A. Rumley, J. Price, and F. Fowkes. Blood viscosity and risk of cardiovascular events: the edinburgh artery study. *British journal of haematology*, 96(1):168–173, 1997.

[6] X. Ma, H. Gao, B. E. Griffith, C. Berry, and X. Luo. Image-based fluid–structure interaction model of the human mitral valve. *Computers & Fluids*, 71:417–425, 2013.

[7] R. Martin, O. Lairez, N. Boudou, S. Mejean, T. Lhermusier, N. Dumonteil, M. Berry, T. Cognet, P. Massabuau, M. Elbaz, H. Rousseau, M. Galinier, and D. Carrie. {R}elation between left ventricular outflow tract obstruction and left ventricular shape in patients with hypertrophic cardiomyopathy: a cardiac magnetic resonance imaging study. *Arch Cardiovasc Dis*, 106(8-9):440–447, 2013.

[8] R. A. Nishimura, C. M. Otto, R. O. Bonow, B. A. Carabello, J. Erwin 3rd, R. A. Guyton, P. T. O'Gara, C. E. Ruiz, N. J. Skubas, P. Sorajja, et al. 2014 aha/acc guideline for the management of patients with valvular heart disease: executive summary: a report of the american college of cardiology/american heart association task force on practice guidelines. *Circulation*, 129(23):2440, 2014.

[9] K. Peterson, J. Tsuji, A. Johnson, J. DiDonna, and M. LeWinter. Diastolic left ventricular pressure-volume and stress-strain relations in patients with valvular aortic stenosis and left ventricular hypertrophy. *Circulation*, 58(1):77–89, 1978.

[10] A. Quaini, S. Canic, R. Glowinski, S. Igo, C. J. Hartley, W. Zoghbi, and S. Little. Validation of a 3d computational fluid–structure interaction model simulating flow through an elastic aperture. *Journal of biomechanics*, 45(2):310–318, 2012.

[11] A. Quaini, S. Canic, G. Guidoboni, R. Glowinski, S. R. Igo, C. J. Hartley, W. A. Zoghbi, and S. H. Little. A three-dimensional computational fluid dynamics model of regurgitant mitral valve flow: validation against in vitro standards and 3d color doppler methods. *Cardiovascular engineering and technology*, 2(2):77–89, 2011.

[12] L. Rodriguez, J. Anconina, F. Flachskampf, A. Weyman, R. Levine, and J. Thomas. Impact of finite orifice size on proximal flow convergence. implications for doppler quantification of valvular regurgitation. *Circulation research*, 70(5):923–930, 1992.

[13] M. E. Silverman and J. W. Hurst. The mitral complex: interaction of the anatomy, physiology, and pathology of the mitral annulus, mitral valve leaflets, chordae tendineae, and papillary muscles. *American heart journal*, 76(3):399–418, 1968.

[14] S. J. Sonntag, W. Li, M. Becker, W. Kaestner, M. R. Büsen, N. Marx, D. Merhof, and U. Steinseifer. Combined computational and experimental approach to improve the assessment of mitral regurgitation by echocardiography. *Annals of biomedical engineering*, 42(5):971–985, 2014.

[15] P. Thavendiranathan, D. Phelan, P. Collier, J. D. Thomas, S. D. Flamm, and T. H. Marwick. Quantitative assessment of mitral regurgitation: How best to do it. *JACC: Cardiovascular Imaging*, 5(11):1161 – 1175, 2012.

[16] J. Udelson, S. Bacharach, R. Cannon, and R. Bonow. Minimum left ventricular pressure during beta-adrenergic stimulation in human subjects. evidence for elastic recoil and diastolic" suction" in the normal heart. *Circulation*, 82(4):1174–1182, 1990.

[17] C.-H. Yap, K. Thiele, Q. Wei, A. Santhanakrishnan, R. Khiabani, M. Cardinale, I. S. Salgo, and A. P. Yoganathan. Novel method of measuring valvular regurgitation using three-dimensional nonlinear curve fitting of doppler signals within the flow convergence zone. *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, 60(7):1295–1311, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

COMPARISON OF TETRAHEDRAL AND HEXAHEDRAL MESHES FOR FINITE ELEMENT SIMULATION OF CARDIAC ELECTRO-MECHANICS

Bernardo L. de Oliveira[1,2] **and Joakim Sundnes**[1,3]

[1]Simula Research Laboratory, Fornebu, Norway
[2]King's College London, London, United Kingdom
[3]University of Oslo, Oslo, Norway,
e-mail: bernardo.de_oliveira@kcl.ac.uk, sundnes@simula.no

Keywords: Cardiac Electro-mechanics, Finite Element Method.

Abstract. *The finite element method is an attractive method for solving the equations describing heart electrophysiology and mechanics. The finite element method with unstructured tetrahedral meshes is particularly suitable for representing complex geometrical shapes. On the other hand, hexahedral elements are generally thought to behave better for solid mechanics problems, and in particular for nearly incompressible, large-deformation hyperelasticity, which is normally encountered in the heart models. In this paper we compare the performance of tetrahedral and hexahedral elements for coupled electro-mechanics simulations, by running two different test cases chosen to represent different features of the coupled problem. A cuboid geometry was used to investigate the convergence of different elements for electrophysiology, while a truncated ellipsoid was used to assess the performance for active and passive mechanics. Both tests were run with linear and quadratic tetrahedral elements and with trilinear and triquadratic hexahedra. The results indicate that all four elements show very similar performance for the electrophysiology part of the problem, but linear tetrahedra stand out as the best choice. The differences are more substantial for the mechanics problem, where the linear tetrahedra showed severe mesh locking problems, and the trilinear hexahedra also performed poorly. The two quadratic elements behaved very similarly, with triquadratic hexahedra only slightly better. We may conclude based on these studies that quadratic tetrahedra are a suitable choice of element for coupled electro-mechanics simulations, which is convenient for image based modeling because of the large availability of automated meshing software.*

1 Introduction

Mathematical models of cardiac electro-mechanics are complex multiscale and multiphysics models consisting of coupled systems of ordinary and partial differential equations (ODEs and PDEs). The model components are typically systems of ODEs describing cellular dynamics and excitation-contraction coupling, a reaction-diffusion PDE describing tissue level electrophysiology, and a non-linear hyperelastic model describing the deformation of the tissue. The complete system is challenging to solve numerically, partly due to the overall complexity of the coupled system, and partly due to the properties of the model components. Specifically, the cell models ODEs are characterized by being stiff and strongly nonlinear, and more recent models consist of 50-100 equations. The reaction-diffusion equations based on bidomain or monodomain approximations require high resolution in space and time, because of rapid dynamics and steep gradients in the solution. Finally, the elasticity problem describing the mechanical deformations is characterized by strong nonlinearities, large deformations, strong anisotropy, and nearly incompressible material behavior. See, for instance, [1, 2, 3] for examples of solution approaches for the coupled system.

A variety of techniques have been used for cardiac electrophysiology, including the finite element (FE), finite difference (FD) and finite volume (FV) methods. For representing realistic heart geometries, FE and FV methods have an advantage over FD based schemes, as unstructured meshes based on tetrahedral elements are generally more convenient than structured meshes of hexahedra. The latter is due to the large availability of efficient automatic meshing tools for tetrahedral meshes [4], which facilitate constructing models based on medical images. A systematic comparison of electrophysiology solvers was presented by Niederer et al. [5], and further discussed and explored in subsequent studies [6, 7]. These studies discuss the different characteristics of FE and FD methods, and the effect of solver choices such as operator splitting schemes and mass lumping, but do not discuss the numerical properties of different element types. Although a number of different element types have been used and reported in the literature, their performance for solving the bidomain or monodomain equations has not been systematically compared.

For solid mechanics problems, and for coupled cardiac electro-mechanics in particular, FE methods are the dominating scheme for spatial discretization. Furthermore, there is a general agreement that for solid mechanics and structural engineering problems, hexahedral elements present better accuracy and efficiency than tetrahedral elements, in particular for low order elements [8]. Due to the higher inherent stiffness of tetrahedral elements, they require higher strain energy states to deform than hexahedral elements [9], and are associated with problems of mesh locking [10]. This phenomenon causes exaggerated errors in the solutions due to the inability of the finite element interpolation functions to correctly approximate volume preserving strain fields [11].

Benzley et al. [9] compared how different elements perform for various elastic and elasto-plastic problems. For elastic problems, linear tetrahedra (LT) produced unacceptably high errors. Trilinear hexahedra (LH) provided acceptable results but required significant refinement in certain cases and both quadratic tetrahedra (QT) and triquadratic hexahedra (QH) provided adequate results in all problems. On the elasto-plastic calculations, again LT produced unacceptable errors and QH produced the best results. However, LH produced results superior to QT.

Different studies have examined how tetrahedra and hexahedra perform when modeling biological tissues. Ramos and Simes [12] studied how the different elements perform to model

the proximal femur in a linear elasticity framework. For a simplified test case they found that LT better approximated the theoretical solution. On a more complex and realistic test case no significant difference were observed between the different elements solutions. A similar comparison, more closely related to the heart mechanics problem, was made by Tadepalli et al. [13] in a problem of foot biomechanics. This problem presents material nonlinearities, incompressibility and large deformations. They concluded that hexahedral elements consistently produced reasonable pressures and stresses, while linear tetrahedral elements only produced reasonable results when the incompressibility could be relaxed.

The purpose of this work is to study the convergence characteristics and performance of different types of elements for cardiac mechanics and electrophysiology. More specifically, we want to investigate the suitability of tetrahedral meshes for cardiac mechanics, and study the performance of linear and quadratic elements for both electrophysiology and mechanics. The paper is organized as follows. In Section 2 we give a detailed presentation of the mathematical problem, and of the test cases used to compare element performance. Results of the numerical experiments are then presented in Section 3, and their implications and limitations are discussed in Section 4. Finally, in Section 5 we present conclusions and recommendations regarding suitable element types for electrophysiology, mechanics, and coupled electro-mechanics problems.

2 Models and methods

2.1 Mathematical Models

Models for cardiac electromechanics can be formulated as coupled systems of ODEs and PDEs. In this work the monodomain model was adopted for the electrophysiology, while the tissue mechanics was described with a hyperelastic model. The main equations of the model read

$$\frac{\partial \mathbf{s}}{\partial t} = f(v, \mathbf{s}, \lambda), \tag{1}$$

$$\chi \left(C_m \frac{\partial v}{\partial t} + I_{ion}(v, \mathbf{s}, \lambda) \right) = \nabla \cdot (\mathbf{D} \nabla v), \tag{2}$$

$$\nabla \cdot (\mathbf{F}\mathbf{S}) = 0, \tag{3}$$

with the constitutive relations

$$\mathbf{S} = \mathbf{S}^p + \mathbf{S}^a \tag{4}$$

$$\mathbf{S}^a = J\mathbf{F}^{-1}\sigma^{\mathbf{a}}(\mathbf{s}, \lambda, \dot{\lambda})\mathbf{F}^{-T} \tag{5}$$

$$\mathbf{S}^p = \frac{\partial \Psi}{\partial \mathbf{E}} \tag{6}$$

$$\Psi = \frac{1}{2} C(e^W - 1) + C_{compr}(J \ln J - J + 1) \tag{7}$$

$$W = b_{ff}E_{ff}^2 + b_{xx}(E_{ss}^2 + E_{nn}^2 + E_{sn}^2 + E_{ns}^2) + b_{fx}(E_{fn}^2 + E_{nf}^2 + E_{fs}^2 + E_{sf}^2). \tag{8}$$

and the boundary conditions

$$\mathbf{n} \cdot (\mathbf{M}\nabla v) = 0, \qquad \mathbf{x} \in \partial H(t), \tag{9}$$

$$\mathbf{u} = 0, \qquad \mathbf{x} \in \partial H_1(0), \tag{10}$$

$$\mathbf{T} = J\bar{\sigma}_1 \mathbf{F}^{-T} \cdot \eta, \ \mathbf{x} \in \partial H_2(0), \tag{11}$$

$$\mathbf{T} = J\bar{\sigma}_2(t) \mathbf{F}^{-T} \cdot \eta, \ \mathbf{x} \in \partial H_3(0). \tag{12}$$

Here, (1) is a system of ODEs describing cell dynamics, including cell electrophysiology and excitation-contraction coupling. The cell state is characterized by the vector $\mathbf{s}$, v is the membrane potential, and λ is fiber stretch. The fiber stretch is calculated from the Green-Lagrange strain tensor $\mathbf{E}$, as $\lambda = \sqrt{2E_{ff} + 1}$, with f indicating the local fiber direction. In this work two different electrophysiology cell models were used, one by ten Tusscher et al [14] and and one from Winslow et al [15]. These models were coupled to the myofilament model by Rice et al [16]. More details on the coupling can be found in [3, 17]. Eq. (2) is the monodomain model, which describes electrical signal conduction through the tissue. Here χ is the ratio of cell membrane area to tissue volume, C_m is cell membrane capacitance, D is tissue conductivity and I_{ion} is the transmembrane current. A no-flux boundary condition is employed, see (9), where $\mathbf{n}$ is the unit surface normal of the deformed heart surface.

The mechanics of the heart muscle is described by the equilibrium equation (3), combined with the constitutive relations (4)-(8). Here, $\mathbf{F}$ is the deformation gradient, $\mathbf{S}$ is the second Piola-Kirchhoff stress tensor, and $E_{ij}, i, j = f, s, n$ are components of the Green-Lagrange strain tensor $\mathbf{E} = 1/2(\mathbf{F}^T\mathbf{F} - 1)$. Eq. (4) describes an additive decomposition of the stress tensor into a passive part $\mathbf{S}^p$ and an active part $\mathbf{S}^a$. The passive component $\mathbf{S}^p$ is derived from the strain-energy density (7)-(8) [18], where C, b_{ff}, b_{xx} and b_{fx} are material parameters. The material is considered almost incompressible, with C_{compr} being a penalty factor used to control volume changes, and J the determinant of $\mathbf{F}$. The active stress component is calculated by mapping the active Cauchy stress tensor, $\sigma^{\mathbf{a}}$, to the reference configuration, as in (5). In local fiber coordinates we have $\sigma^{\mathbf{a}} = \text{diag}(T_a, \eta T_a, \eta T_a)$, where η is a constant and $T_a(\mathbf{s})$ is the dynamic tension computed from the cell model state vector. The subscripts f, s and n correspond to the fiber, sheet and sheet-normal directions, respectively. The mechanics boundary conditions are given by (10)-(12), where $\mathbf{u}$ is displacement, $\bar{\sigma}_1, \bar{\sigma}_2$ are prescribed stresses on the heart surface, and η is the unit surface normal of the undeformed reference geometry.

2.2 Numerical Methods

A common approach to solve (1)-(8) is to apply an operator splitting technique to divide the system into smaller sub-problems. Here, we have used the splitting method from [3], where the mechanics problem is separated from the electrophysiology part by means of a Gauss-Seidel approach, and a Godunov splitting is applied to (2) to separate the nonlinear reaction term from the diffusion part. Three separate problems must be solved for each time step; a nonlinear system of ordinary differential equations describing cell electrophysiology and force generation, a linear parabolic PDE that represents the electro-diffusion in the cardiac tissue, and a nonlinear elasticity equation describing the mechanical deformations. Assuming that the variables v_n, $\mathbf{s}_n$ and $\mathbf{u}_n$ are known at time t_n, the solution is advanced by the following procedure:

1. Solve the nonlinear ODE system

$$\frac{dv}{dt} = -I_{ion}(v, s, \lambda_n), \qquad t_n < t < t_{n+1}, v(t_n) = v_n, \tag{13}$$

$$\frac{ds}{dt} = f(v, s, \lambda_n), \qquad t_n < t < t_{n+1}, s(t_n) = s_n. \tag{14}$$

An updated vector s_{n+1} is obtained, together with an intermediate updated transmembrane potential v^*. Note that the fiber stretch is held constant over this integration step, $\lambda = \lambda_n$ [19]. This system of ODEs was solve using a third-order Singly Diagonally Implicit Runge-Kutta method with adaptive time stepping, see [20].

2. Solve the parabolic linear PDE using v^* as the initial condition,

$$\nabla \cdot (\mathbf{D}\nabla v) = \frac{\partial v}{\partial t} \tag{15}$$

Our strategy to solve this equation is to discretize it in time using a backward Euler scheme:

$$\nabla \cdot (\mathbf{D}\nabla v^{n+1}) = \frac{v^{n+1} - v^*}{\Delta t} \tag{16}$$

For spatial discretization of (16), we apply a standard Galerkin FE formulation. In short, for a suitable function space $\Phi(\Omega)$, we multiply both sides of (16) with a test function $\phi \in \Phi(\Omega)$, and integrate over the domain Ω. Integration by parts and application of the homogeneous Neumann boundary condition leads to the following weak form:
Find $v^{n+1} \in \Phi(\Omega)$ such that:

$$\int_\Omega v^{n+1}\phi d\Omega + \Delta t \int_\Omega \mathbf{D}\nabla v^{n+1}\nabla\phi d\Omega = \int_\Omega v^*\phi d\Omega, \quad \forall \phi \in \Phi(\Omega) \tag{17}$$

This variational problem is discretized by defining a finite-dimensional subspace $\Phi_h(\Omega) \subset \Phi(\Omega)$, and approximating the solution as a linear combination of the basis functions of Φ_h; $v = \sum_i v_i, \phi_i$. This standard FEM procedure leads to a linear system that is solved using a multigrid preconditioned conjugate gradient method, see e.g. [19] for details.

3. Solve the nonlinear mechanics problem using the most recently calculated value of the state vector , s^{n+1}, to calculate the active stress tensor $\mathbf{S}^a$

$$\nabla \cdot \Big(\mathbf{F}(\mathbf{S}^p + \mathbf{S}^a(s^{n+1}, \lambda, \dot{\lambda}))\Big) = 0 \tag{18}$$

The mechanical problem is solved by a standard, displacement-based Galerkin FEM. Applying similar steps as for the diffusion problem above leads to a system of non-linear algebraic equations to be solved for the unknown displacement field. We solve this system iteratively by the Newton-Raphson method, which involves a linearization of all relevant quantities. For each iteration of the Newton-Raphson method, we solve a linear system with the stiffness matrix

$$A_{ij} = \int_\Omega \nabla\phi_i : \mathcal{A} : \nabla\phi_j d\Omega, \tag{19}$$

where $\mathcal{A}$ is the first elasticity tensor, which characterizes the tangential stiffness of the tissue. See for instance [21, 22] for details on the displacement-based FE formulation for large elastic deformations. Following the additive split in (4), the fourth-order tensor $\mathcal{A}$ has contributions from the active and passive strain components,

$$\mathcal{A}_{ijkl} = \frac{\partial\{\mathbf{F}\mathbf{S}^{\mathbf{p}}\}_{ij}}{\partial\mathbf{F}_{kl}} + \frac{\partial\{\mathbf{F}\mathbf{S}^{\mathbf{a}}\}_{ij}}{\partial\mathbf{F}_{kl}}. \tag{20}$$

Both terms are computed by numerical differentiation of the stress-strain relation. The passive part the stress-strain relation is obtained by analytical differentiation of the strain energy function in (6), while the active part is obtained by a pull-back of the active Cauchy stress, as defined in (5).

2.3 Test Cases

A key challenge in defining the test cases to compare the different element types is to decide the level of complexity that should be adopted. On one hand, it is interesting to keep the tests simple so they can be clearly expressed, take a limited amount of time and computational resources to be performed and can be more easily communicated and reproduced. On the other hand, the test cases must be complex enough to be physiologically relevant and unravel the performance characteristics of the methods when applied to more realistic and complex simulations. Two different test cases are considered in this study.

The first test case only considers electrophysiology, and is inspired by Niederer et al. [5], where different electrophysiology simulators were tested, and the results were compared in terms of the conduction velocity. We adopted the same simplified geometry and stimulation protocol. A rectangular box of dimensions 3 x 7 x 20 mm^3 was used, with fibers aligned in the longer axis. Regular grids were generated for this geometry using different element types and node spacing. The fiber conductivity was set to 1.33 mS cm^{-1}, the transversal conductivity was 0.18 mS cm^{-1} and the surface to volume ratio was 1400 cm^{-1}. A stimulation current was applied in one corner, point P1 in Figure 1. These tests were performed using two different cell models. A zero flux boundary condition was assumed implicitly in the whole boundary. The quality of the solutions was measured similarly to Niederer et al. [5], by comparing the activation times of the last node in the mesh, point P2 on Figure 1. This activation time can be used to get a good estimative of the conduction velocities of the electrical signal, which is known to be highly sensitive to the mesh resolution. As disturbances in the conduction velocity are often associated with arrhythmias and other pathologies it is of great importance to have converged conduction velocities in order to support any physiological or clinical conclusions based on electrophysiological simulations.

The second test case is focused on active and passive mechanics. The geometry is a truncated ellipsoid as shown in Figure 2, which is a reasonable approximation of a left ventricle. The fiber field was set as aligned in the tangential direction of the geometry, with a rotation from -30^o to 30^o from epicardium to endocardium. Although we ran the full coupled model, (1)-(8), electrical activation was applied homogeneously and simultaneously to the entire domain. Effects of electrical diffusion are therefore minimized, and the main focus is on the mechanics. The material parameters C, b_{ff}, b_{xx} and b_{fx} were chosen to be 0.25 kPa, 50, 25 and 25 respectively. The incompressibility penalty factor C_{compr} was chosen to be 200 kPa. This choice gave the realistic nearly incompressible behavior of the myocardium, with volume changes ¡5%, which is known to lead to mesh locking that will highlight the difference between element types.

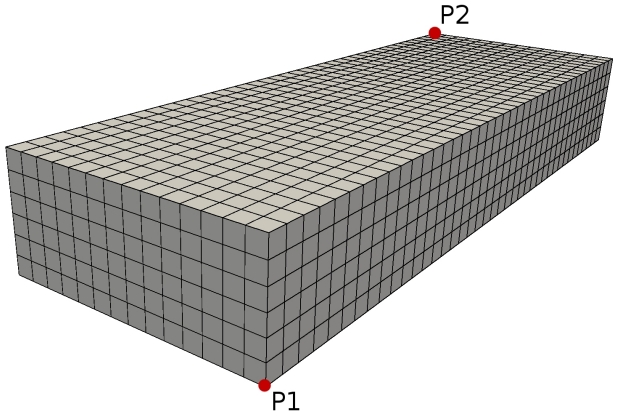


Figure 1: A rectangular box geometry taken from [5] was used for the pure electrophysiology comparison.

For boundary conditions, the base was fixed in the longitudinal direction, and two additional points on the base were constrained to avoid rigid body movement. The outer (epicardial) surface was assumed stress-free, while a time-varying pressure was applied to the endocardium, to capture the four phases of a generic pressure-volume (PV) loop. Each test would start with the passive filling phase, where the endocardial pressure was increased from 0 to 2 kPa over a time interval of 200 ms. At $t = 180$ms, stimulus was applied to trigger contraction, and a simplified circulation model was used for the remaining phases of the heart cycle. The following equations were solved to determine the endocardial pressure:

$$\frac{dV_{cav}}{dt} = -q_{ao}, \tag{21}$$

$$q_{ao} = \max(C_{art}\frac{dp}{dt} + R_{per}(p - p_{per}), 0). \tag{22}$$

Here V_{cav} is the cavity volume computed from the deformed FEM mesh, q_{ao} is the aortic flow, p is the cavity pressure applied to the endocardial boundary, C_{art} is the compliance of the systemic arterial tree, R_{per} is peripheral resistance and p_{per} is the peripheral pressure. By inserting (22) into (21) and applying a finite difference approximation for the time derivative, we get a non-linear equation which can solved for the cavity pressure p by an iterative Newton algorithm, see e.g. [3] for details.

3 Results

3.1 Electrophysiology

Several simulations were performed using meshes with different resolutions and element types. The coarsest mesh had 4305 degrees of freedom (DOFs), which corresponds to a node spacing of 0.5 mm. Four levels of refinement were used, with each level reducing the node spacing by a factor of two, which corresponds to a factor eight increase in the number of DOFs. Due to limitations in our software the last level of refinement could not be simulated with QTs. Since the PDE is discretized in both time and space, results from two different time stepping values, 50μs and 10μs, are presented to investigate how the error introduced by the time discretization affect the results.

The activation times obtained in the simulations can be seen on Figure 3. The top row shows the results from the model by ten Tusscher et al., and the bottom row shows results of

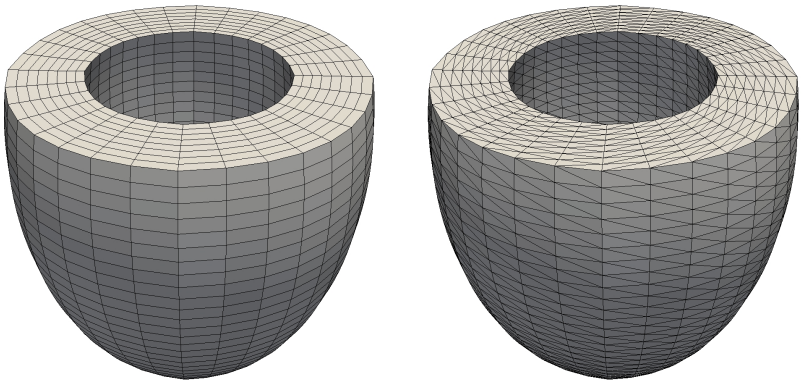


Figure 2: An ellipsoidal geometry was used for the mechanical test cases with hexahedra and tetrahedra

the Winslow et al. model. We observed signal conduction in all levels of refinement with the ten Tusscher et al model, while the model by Winslow et al did not conduct on the coarsest meshes. The results displayed in the bottom row therefore starts at one level of refinement. The conduction block can be explained by the lower upstroke velocity of the Winslow et al model, which leads to a lower conduction velocity and increased risk of conduction block. We see that the LT elements consistently give lower activation time (higher conduction velocity) than the other elements, but there are no substantial differences between the element types. All elements produce unacceptable errors on the coarsest meshes, while on the fourth level of refinement all solutions are visually very similar. Note also that the difference between the two time step values is larger for the model by ten Tusscher et al. The larger time discretization errors seen for this model can be explained by the higher upstroke velocity.

The relative errors in the activation times are shown in Figure 4. Since no analytical solution is available, we calculated the reference activation time as the average of the computations on the finest grid, with a time step of 10μs and a node spacing of 62.5μm. While this choice of reference will clearly not give accurate errors for the finer meshes, it does provide a useful measure for comparing the coarsest meshes. The error plots show a similar behavior as the previous plots, with relatively small differences between the element types. The slightly less consistent results can be explained by the shape of the curves in Figure 3, which show that all the elements overestimate the activation time (i.e. underestimate the conduction velocity) on the coarse meshes, but as the mesh is refined further most elements underestimate the time and approach the correct value from below. This behavior, which can be shown analytically for simplified models, means that a fairly coarse mesh may give a very accurate conduction velocity, and that the error may increase as the grid is refined. The error from time discretization is also very visible in the upper left corner, which shows all the curves converging to a value significantly different from the reference.

3.2 Mechanics

For this test case, again, the solutions using meshes with different levels of refinement and elements were compared. The coarsest mesh had 219 degrees of freedom (DOFs). Three levels of refinement were used, with each level increasing the number of DOFs by a factor of eight. As for the electrophysiology problem, no analytical solution is available. Instead, we compute a reference solution by running one simulation with an additional level of refinement.

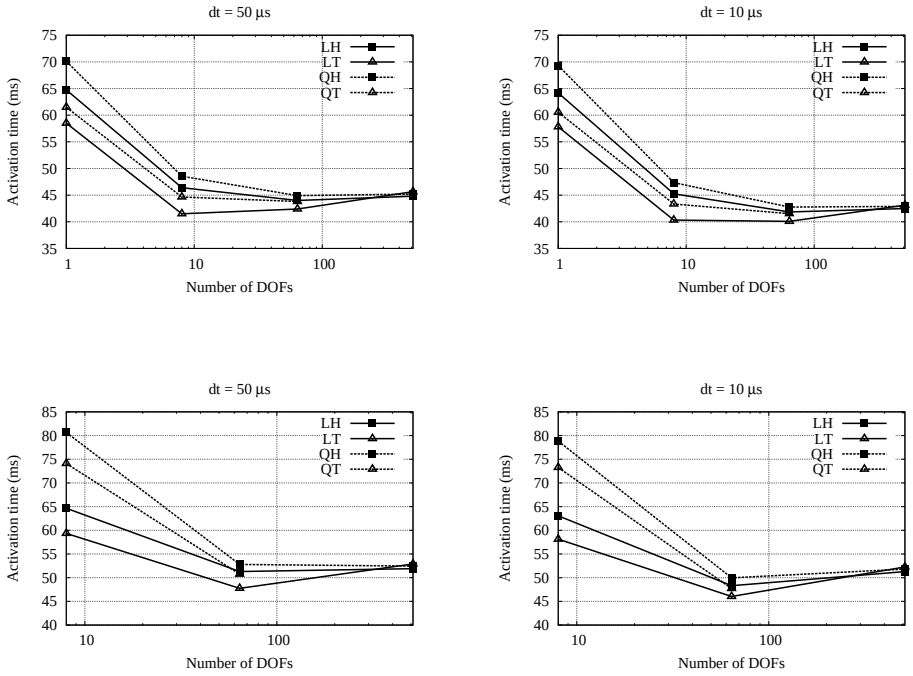


Figure 3: Activation times with the ten Tusscher (top) and Winslow (bottom) models. The number of DOFs is normalized by the coarsest mesh.

We first compared the elements for the passive inflation phase, by computing the relative error in displacement and stress. The left panel of Figure 5 shows the L2 error norm of the displacement fields. We see that for this problem the quadratic elements outperform the linear ones, and the LT elements are particularly bad, producing unacceptable errors for all refinement levels. This result is in line with previous results and with the general view of the solid mechanics community, that LT are not suited for nearly incompressible elasticity problems. However, QH and QT elements show very similar behavior, indicating that the poor performance of tetrahedra is limited to linear elements. The right panel of Figure 5 shows the L2 error norm of the stress fields. These results show an even clearer advantage of quadratic elements, as both the linear elements produce large errors. However, the difference between QH and QT is only significant on the coarser meshes.

To further investigate the performance of the different element types, we computed the full PV loops, shown in Figure 3.2. The PV loops were computed as specified in Section 2, by first ramping up the pressure to a specified end diastolic pressure, and then solve (21)-(22) for the isovolumic and ejection phases. The top left panel shows results for the coarsest mesh, the top right panel is for one level of refinement, while the bottom panel shows the finest mesh. We can see that for the coarsest mesh, both the linear elements show a much higher stiffness than the quadratic ones, and inflate to lower end diastolic volumes. The two quadratic elements show very similar behavior even for this coarse mesh. On the second level of refinement the curves for the two quadratic elements visually overlap. The linear elements also show much better results on this level, although the end diastolic volume is still significantly lower than the reference. On the third level of refinement all curves are visually very similar.

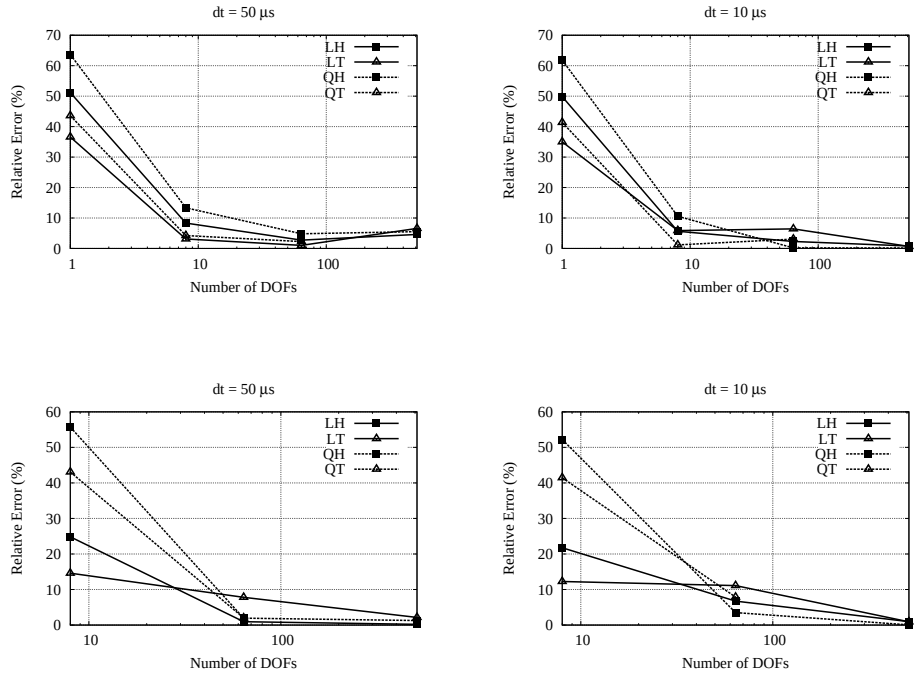


Figure 4: Relative error of the activation times with the ten Tusscher(top) and Winslow (bottom) models. The number of DOFs is normalized by the coarsest mesh.

4 Discussion

We have compared different element types for simulations of cardiac electrophysiology and mechanics, with the aim of testing the common conception that hexahedral elements are superior to tetrahedra for solid mechanics applications, and to compare the elements' performance for coupled simulations.

For the test case involving pure electrophysiology, no firm conclusions could be made regarding the suitability of different element types. All element types produced unacceptable errors for the coarsest mesh, and also for one level of refinement. On the second level of refinement, with node spacing of 250 μm, most of the results were within 5 to 10% from the reference activation time. These errors are normally considered too large for practical applications. For a node spacing of 62.5 μm, the variation between solutions is less than 1%. However, the com-

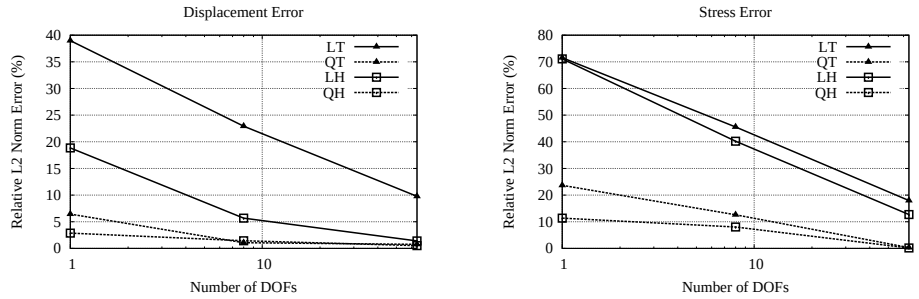


Figure 5: L2 Error of the displacement (left) and stress (right).

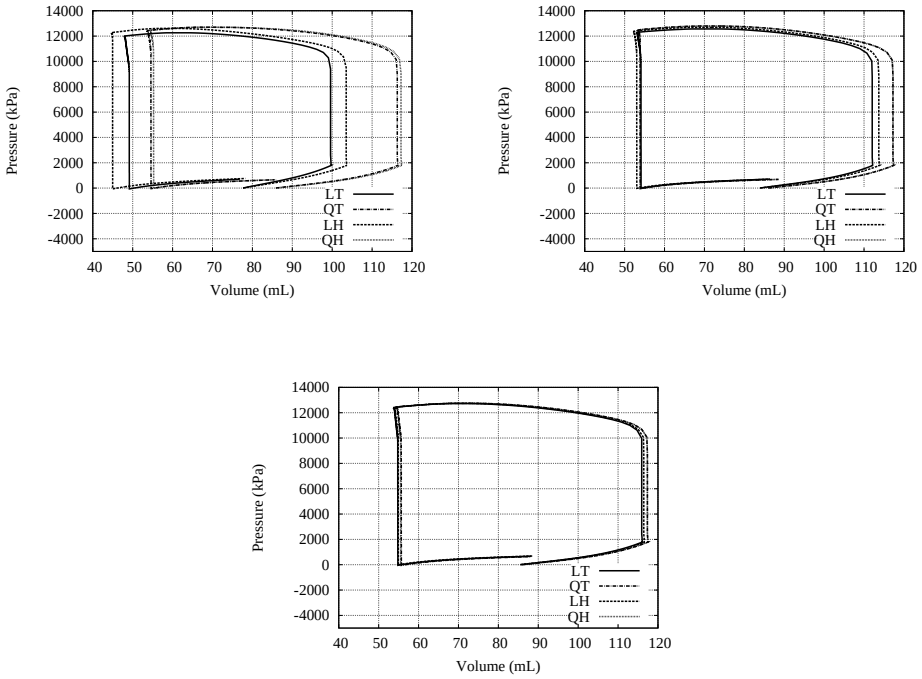


Figure 6: PV loops for the four different element types and three levels of refinement. The top left is the coarsest mesh, top right is for one level of refinement, and the bottom panel is the finest mesh.

parison made here is entirely based on comparing the error for a given number of degrees of freedom, not looking at the actual CPU time for each element type. The CPU time per degree of freedom is higher for quadratic than for linear elements, and one may argue that our results indicate that linear elements perform better. For some cases, the CPU time for QH and QT was up to 30% higher than for LH and LT, resulting from increased cost of assembling and solving the linear systems.

The mechanics cases show a clear advantage of quadratic elements, which is in line with previous results in the literature, see, e.g. [9], and what seems to be the general opinion in the field. However, the difference between hexahedra and tetrahedra was somewhat smaller than expected. For passive inflation, LH clearly outperformed LT, but they both showed large errors, in particular when comparing the stress fields. The QH elements proved superior to QT, but the differences are small. The simulations of full PV loops confirmed the results from the passive inflation case, in that the linear elements substantially overestimated the overall stiffness of the ventricle. However, when comparing the plots in Figure 3.2 to the displacements in Figure 5, the PV loops seem to be more converged than the displacements for the second grid level, which has an average node spacing of 0.75cm. A more careful examination reveals that this apparent discrepancy is due to linear elements better capturing the radial displacement than the shear strains and torsion of the ellipsoid. This inability of the linear elements, particularly the tetrahedra, to capture the shear strains observed in the reference solution are a strong indication of the phenomenon of mesh locking [10].

There are a number of limitations to the present study, some of which are natural candidates for future research. Our study has focused on a limited selection of element types, and considering two relatively simple model problems. However, the elements considered herein are widely used in the cardiac modeling community, and are among the most widely available ele-

ment types in automatic meshing software. For an even more representative element selection it would be useful to include elements based on cubic Hermite interpolation, as these are common for cardiac modeling, see e.g. [18]. However, Hermite elements were not available in the FE frameworks available for this study, and adding these elements to the frameworks is a considerable task which has been left for future investigations. Other limitations include the simplicity of the two test cases. These were chosen to be representative of more advanced simulation models, but still be sufficiently simple to enable a rudimentary convergence study. Furthermore, although the final test case is a fully coupled simulation, the effects of electro-diffusion are negligible because of the simultaneous activation pattern. A fully coupled simulation with a propagating signal may give additional results, but will also considerably complicate the analysis, since the meshes considered here are too coarse for conduction of the electrical signal.

5 Conclusions

Any conclusions on element choice are naturally affected by the available software and choice of solution method. The operator splitting technique employed in this study is representative for a number of solution methods applied in the field, and enables a decoupled solution of the mechanics and electrophysiology sub-problems. However, even with this uncoupling, it is convenient to work with nested meshes, where the electrophysiology mesh is a refinement of the mechanics mesh. This consideration motivates the use of identical or similar elements for the two problems. It is possible to implement a more general operator splitting scheme, where the meshes and elements for electrophysiology and mechanics are completely independent, as well as more tightly coupled schemes that employ identical meshes for the two sub-problems. These solver choices will obviously affect the choice of elements, but results presented here are representative for a widely used class of solution scheme.

The main conclusion of this work is that quadratic and triquadratic elements are best suited for coupled electro-mechanics. These elements show superior performance for mechanics, and comparable performance as the linear elements for the electrophysiology part. For pure electrophysiology simulations, our results indicate that LT elements should be chosen, but this should always be avoided for mechanics computations. No significant difference is seen between QH and QT elements, which indicates that the added geometric flexibility of tetrahedral meshes make them highly attractive for image based and patient specific modeling.

REFERENCES

[1] S. A. Niederer and N. P. Smith. An improved numerical method for strong coupling of excitation and contraction models in the heart. *Prog Biophys Mol Biol*, 96:90–111, 2008.

[2] Serdar Göktepe and Ellen Kuhl. Electromechanics of the heart: a unified approach to the strongly coupled excitation–contraction problem. *Computational Mechanics*, 45(2-3):227–243, November 2009.

[3] J. Sundnes, S. Wall, H. Osnes, T. Thorvaldsen, and A.D. McCulloch. Improved discretisation and linearisation of active tension in strongly coupled cardiac electro-mechanics simulations. *Computer Methods in Biomechanics and Biomedical Engineering*, 17:604–615, 2014.

[4] Hang Si. Tetgen: A quality tetrahedral mesh generator and a 3d delaunay triangulator. url: http://wias-berlin.de/software/tetgen/, June 2014.

[5] Steven A. Niederer, Eric Kerfoot, Alan P. Benson, Miguel O. Bernabeu, Olivier Bernus, Chris Bradley, Elizabeth M. Cherry, Richard Clayton, Flavio H. Fenton, Alan Garny, Elvio Heidenreich, Sander Land, Mary Maleckar, Pras Pathmanathan, Gernot Plank, Jos F. Rodrguez, Ishani Roy, Frank B. Sachse, Gunnar Seemann, Ola Skavhaug, and Nic P. Smith. Verification of cardiac tissue electrophysiology simulators using an n-version benchmark. *Philosophical Transactions of the Royal Society A*, 369:43314351, 2011.

[6] P. Pathmanathan, M.O.Bernabeu, S.A.Niederer, D.J.Gavaghan, and D. Kay. Computational modelling of cardiac electrophysiology: explanation of the variability of results from different numerical solvers. *International Journal for Numerical Methods in Biomedical Engineering*, 28:890–930, 2012.

[7] Shankarjee Krishnamoorthi, Mainak Sarkar, and William S. Klug. Numerical quadrature and operator splitting in finite element methods for cardiac electrophysiology. *International Journal for Numerical Methods in Biomedical Engineering*, 29:1243–1266, 2013.

[8] John R. Brauer. *What Every Engineer Should Know About Finite Element Analysis*. Marcel Dekker Inc, 1993.

[9] Steven E. Benzley, Ernest Perry, Karl Merkley, Brett Clark, and Greg Sjaardema. A comparison of all hexagonal and all tetrahedral finite element meshes for elastic and elastoplastic analysis. In *In Proceedings, 4th International Meshing Roundtable*, 1995.

[10] Thomas J.R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Englewood Cliffs, 1987.

[11] Allan F. Bower. *Applied Mechanics of Solids*. CRC Press, 2009.

[12] A. Ramos and J.A. Simes. Tetrahedral versus hexahedral finite elements in numerical modelling of the proximal femur. *Medical Engineering and Physics*, 28:916924, 2006.

[13] Srinivas C. Tadepalli, Ahmet Erdemir, and Peter R. Cavanagh. Comparison of hexahedral and tetrahedral elements in finite element analysis of the foot and footwear. *Journal of Biomechanics*, 44:23372343, 2011.

[14] K. H. ten Tusscher, D. Noble, P. J. Noble, and A. V. Panfilov. A model for human ventricular tissue. *Am J Physiol Heart Circ Physiol*, 286:1573–1589, 2004.

[15] Raimond L. Winslow, Jeremy Rice, Saleet Jafri, Eduardo Marbn, and Brian ORourke. Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure, ii model studies. *Circulation Research*, 84:571–586, 1999.

[16] J. Rice, F. Wang, D. Bers, and P. Tombe. Aproximate model of cooperative activation and crossbridge cycling in cardiac muscle using ordinary differential equations. *Biophysical Journal*, 95:2368–2390, 2008.

[17] B. L. de Oliveira, B. M. Rocha, L. P. S. Barra, E. M. Toledo J. Sundnes, and R. Weber dos Santos. Effects of deformation on transmural dispersion of repolarization using in silico models of human left ventricular wedge. *International Journal for Numerical Methods in Biomedical Engineering*, 29:13231337, 2013.

[18] Taras P. Usyk, Ian J. LeGrice, and Andrew D. McCulloch. Computational model of three-dimensional cardiac electromechanics. *Computing and Visualization in Science*, 4:249–257, 2002.

[19] J. Sundnes, G. T. Lines, and A. Tveito. An operator splitting method for solving the bidomain equations coupled to a volume conductor model for the torso. *Mathematical Biosciences*, 194:233–248, 2005.

[20] J. Sundnes, G. Lines, and A. Tveito. Efficient solution of ordinary differential equations modeling electrical activity in cardiac cells. *Mathematical Biosciences*, 172:55–72, 2001.

[21] G A Holzapfel. *Nonlinear solid mechanics: a continuum approach for engineering*. Wiley, 2000.

[22] J. Bonet and R.D. Wood. *Nonlinear continuum mechanics for finite element analysis*. Cambridge University Press, 1997.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A PARAMETRIC STUDY OF A MULTIPHASE POROUS MEDIA MODEL FOR TUMOR SPHEROIDS AND ENVIRONMENT INTERACTIONS

P. Mascheroni[1], D.P. Boso[1], C. Stigliano[2], M. Carfagna[3], L. Preziosi[3], P. Decuzzi[4], B.A. Schrefler[1,2]

[1] Dipartimento di Ingegneria Civile Edile ed Ambientale, Università di Padova
Via Marzolo 9, 35131 Padova, Italy
e-mail: pietro.mascheroni@dicea.unipd.it, daniela.boso@unipd.it, bernhard.schrefler@dicea.unipd.it

[2] Nanomedicine Department, Houston Methodist Research Institute
6670 Bertner Street, 77030 Houston, TX
cstigliano@houstonmethodist.org

[3] Dipartimento di Scienze Matematiche, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
melania.carfagna@polito.it, luigi.preziosi@polito.it

[4] Laboratory of Nanotechnology for Precision Medicine, Italian Institute of Technology
Via Morego 30, 16163 Genova, Italy
e-mail: paolo.decuzzi@iit.it

Keywords: Mathematical Modeling, Parametric analysis, Cancer Growth, U-87MG Spheroids, Mechanical Compression of Spheroids, Growth Inhibition.

Abstract. *Computational models for tumor growth provide an effective in silico tool to investigate the different stages of cancer growth. Recently, a series of computational models based on porous media theory has been proposed to predict tumor evolution and its interactions with the host tissue. In addition, a specialization of the original models, adapted for tumor spheroids, has been proposed and validated experimentally. However, due to the complexity of the modeling framework, a systematic understanding of the role of the parameters governing the equations is still lacking. In this work, we perform a parametric analysis on a set of fundamental parameters appearing in the model equations. We investigate the effects of a variation of these coefficients on the spheroid growth curves and, in particular, on the final radii reached by the cell aggregates in the growth saturation stage. Finally, we provide a discussion of the results and give a brief summary of our findings.*

1 INTRODUCTION

Cancer is one of the leading causes of the death in the world [1], involving primarily altered cell proliferation and migration of cancer cells to form metastasis in different regions of the body [2]. Recently, it has become clear that a collective effort from all the physical sciences is required to improve our understanding of this illness and design new strategies for therapeutical treatments [3–6]. The development of effective computational models to investigate the growth of a tumor mass is certainly a valid contribution to this field. Mathematical models can analyze the tumor evolution in detail, with a strict control of the parameters governing the equations. Moreover, they can help the experimental investigators in dissecting the dynamics of the systems under study and guide the design of new experiments [7].

In [8], the authors present a computational model for avascular tumor growth based on porous media theory. They apply their modeling framework to the growth of a tumor spheroid *in vitro* and confined in a healthy tissue, where the relative adhesion of tumor cells and host cells to the extracellular matrix is analyzed. Moreover, they study a tumor cord in a three dimensional geometry, where tumor cells grow around microvessels carrying nutrients. They further develop the existing model in [9], where different adhesion interactions between host cells, tumor cells and interstitial fluid are taken into account. Five different cases are dealt with, where interfacial tensions between the cellular constituents and the interstitial fluid are varied, together with their dynamic viscosities. In [10], the authors relax the hypothesis of a rigid extracellular matrix in the tumor tissue, and investigate its deformability making use of rate-dependent plasticity. They apply this framework to analyze the growth of a tumor spheroid in a decellularized extracellular matrix and then in the presence of host cells. Further, they apply the model to study the growth of a melanoma and investigate its temporal and spatial evolution. Another development of the model is given in [11], where the authors introduce remodeling of the extracellular matrix during tumor growth and cell lysis. They study the effect of matrix remodeling on spheroid growth inside a decellularized matrix by comparing their results to the previous implementation. They investigate also the impact of lysis in the dynamics of the system. In [12], the authors specialize the previous modeling framework for tumor spheroids. They carry out a set of experiments on U-87 spheroids to validate the equations and test new constitutive relations. Comparison with experiments is performed both with spheroids freely suspended in a culture medium and subjected to different mechanical loads.

Due to the complexity of these models, an understanding of the role of the different parameters is difficult. For example, in [12] it is not clear which model parameters have a significant effect on the growth of the spheroids. In this paper, we perform a parametric analysis on a set of governing coefficients appearing in the model equations. We test the effect of parameter variation on the spheroid growth curve and in particular on the final radius reached by the cell aggregate. Finally, we provide a discussion of the results and summarize our findings in the conclusions.

2 THE MATHEMATICAL MODEL

The governing equations of the model are derived in the context of porous media theory. Starting from microscopic relations between the constituents, the theory makes use of suitable spatial and temporal averaging theorems to provide balance laws at the scale of the tissue (also termed, “the macroscale”). In this way, the complexity arising from the high spatial variability and several interactions characterizing the microscale is overcome. Then, the introduction of a suitable set of constitutive relations into the macroscale equations provides

the closed form of the problem. Detailed information about the model derivation can be found in previous works of the authors [8–12].

We model the tumor tissue as a biphasic porous medium, composed of the following constituents, or phases: (i) the tumor cells (TCs), which are divided into living (LTCs) and necrotic (NTCs) cells, and (ii) the interstitial fluid (IF), represented in Figure 1. In the language of porous media theory, the union of TCs and extracellular matrix (ECM) constitutes the solid skeleton of the system, whereas the IF represents the fluid phase permeating the pores. The IF carries nutrients, growth factors and waste products; for the sake of simplicity, we consider only one nutrient in our model, namely oxygen (ox), which can diffuse and be consumed by LTCs. Adequate levels of nutrient are necessary for cell proliferation, otherwise they start necrosis and lysis. Finally, we assume cell duplication to be influenced by the local level of mechanical stress, with cells proliferating poorly when subjected to compression. In the following, *t* (tumor) and *l* (liquid) will denote quantities related to the solid and fluid part of the biphasic system, respectively.

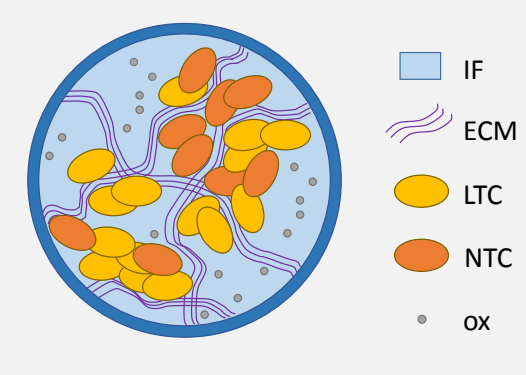


Figure 1: The constituents of the biphasic system.

2.1 The governing equations for tumor spheroids

We denote the volume fractions of the solid and the fluid by ε^t and ε^l, respectively. We assume that the fluid permeates completely the voids left by the solid skeleton, and apply the saturation constraint:

$$\varepsilon^t + \varepsilon^l = 1 \tag{1}$$

Then we write the governing equations for the tumor volume fraction, the interstitial fluid pressure (p^l), the oxygen mass fraction (ω^{ox}) and the necrotic cell mass fraction (ω^{Nt}). In the following, we report the balance laws as appear for the tumor spheroid growth case. We refer the interested reader to [12] for a full derivation. We solve the system of equations given by:

$$\frac{\partial \varepsilon^t}{\partial t} - \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \varepsilon^t \frac{k}{\mu^l} \Sigma' \frac{\partial \varepsilon^t}{\partial r} \right) - \frac{1}{\rho}\left(\overset{l\to t}{M}_{growth} - \overset{t\to l}{M}_{lysis} \right) = 0 \tag{2}$$

$$\frac{\partial \left(\omega^{Nt} \varepsilon^t \right)}{\partial t} - \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \varepsilon^t \omega^{Nt} \frac{k}{\mu^l} \Sigma' \frac{\partial \varepsilon^t}{\partial r} \right) - \frac{1}{\rho}\left(\varepsilon^t r^{Nt} - \overset{t\to l}{M}_{lysis} \right) = 0 \tag{3}$$

$$\frac{\partial \left[\left(1-\varepsilon^t \right) \omega^{ox} \right]}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \varepsilon^t \omega^{ox} \frac{k}{\mu^l} \Sigma' \frac{\partial \varepsilon^t}{\partial r} \right) - \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 \left(1-\varepsilon^t \right) D^{ox} \frac{\partial \omega^{ox}}{\partial r} \right] + \frac{1}{\rho} \overset{ox\to t}{M}_{oxygen} = 0 \tag{4}$$

where r is the radial coordinate over the spheroid radius; k is the intrinsic permeability of the solid matrix; μ^l is the dynamic viscosity of the IF; D^{ox} is the diffusion coefficient of oxygen; and ρ is the density of the phases. As described in detail in [12], Σ is a quantity relating the mechanical stress in the tumor to the solid volume fraction. In the equations above we make use of Σ', the derivative of Σ respect to ε^t, which can be computed analytically from the expression reported in [12].

The mass exchange terms appearing in equations (2)-(4) have the form:

$$\underset{growth}{\overset{l\to t}{M}} = \gamma_g^t \left\langle \frac{\omega^{ox} - \omega_{crit}^{ox}}{\omega_{env}^{ox} - \omega_{crit}^{ox}} \right\rangle_+ H\left(\mathrm{t}_{\mathrm{eff}}^t\right) \omega^{Lt} \varepsilon^t \tag{5}$$

$$\underset{lysis}{\overset{t\to l}{M}} = \lambda_l^t \omega^{Nt} \varepsilon^t \tag{6}$$

$$\varepsilon^t r^{Nt} = \gamma_n^t \left\langle \frac{\omega_{crit}^{ox} - \omega^{ox}}{\omega_{env}^{ox} - \omega_{crit}^{ox}} \right\rangle_+ \omega^{Lt} \varepsilon^t \tag{7}$$

$$\underset{oxygen}{\overset{ox\to t}{M}} = \gamma_0^t \frac{\omega^{ox}}{\omega^{ox} + c^{ox}} \omega^{Lt} \varepsilon^t \tag{8}$$

Here γ_g^t, λ_l^t, γ_n^t and γ_0^t are coefficients that account for the nutrient and IF mass that becomes tumor due to cell growth; the degradation of cellular membranes and the following mass conversion into IF; the rate of cell death; and the oxygen uptake rate in the tumor, respectively. The quantity $\omega^{Lt} = 1 - \omega^{Nt}$ represents the mass fraction of LTCs, and guarantees that growth, death and oxygen uptake are active only on the living portion of the spheroid. The Macaulay brackets $\langle \cdot \rangle_+$ appearing in equations (5) and (7) return the positive value of their argument. Since the oxygen mass fraction inside the spheroid is equal or smaller than its environmental level in the culture medium ω_{env}^{ox}, the brackets in equation (5) will return a value between unity (for $\omega^{ox} = \omega_{env}^{ox}$) and zero (for $\omega^{ox} \le \omega_{crit}^{ox}$). Note that here ω_{crit}^{ox} is the oxygen threshold level below which cell proliferation is inhibited. Equation (6) describes cell lysis occurring in the NTCs, whereas a consideration similar to the one for equation (5) holds true for equation (7), which describes TC death due to the lack of nutrient. Finally, equation (8) describes the uptake of oxygen by LTCs and accounts for the dependence of nutrient consumption on its local level. Here c^{ox} is the oxygen mass fraction at which oxygen consumption is reduced by half. Note that the function H in equation (5) describes the inhibition of cell proliferation due to the mechanical stress exerted on the TCs. Even though several alternatives are given in the literature, in [12] we provide a mathematical expression for this quantity that is able to describe accurately our experimental observations on spheroid growth under a controlled external compression.

We model the growth of the spheroid as a free-boundary problem, where the interface constituted by the TCs moves with velocity v^t, given by:

$$\frac{dR}{dt} = v^t = -\frac{k}{\mu^l} \Sigma' \frac{\partial \varepsilon^t}{\partial r} \tag{9}$$

with R the external radius of the spheroid. The closed form of the differential problem is obtained by defining a set of boundary and initial conditions. In particular, symmetry requires no-flow boundary conditions at the spheroid center, while we assume Dirichlet boundary conditions on the tumor external surface:

$$\frac{\partial \varepsilon^t}{\partial r} = \frac{\partial \omega^{Nt}}{\partial r} = \frac{\partial \omega^{ox}}{\partial r} = 0, \quad \text{in } r = 0, \tag{10}$$

$$\varepsilon^t = \varepsilon^t_{ext}, \quad \omega^{Nt} = 0, \quad \omega^{ox} = \omega^{ox}_{env}, \quad \text{in } r = R \tag{11}$$

Finally, we assume the following initial conditions over the spheroid radius:

$$\varepsilon^t = \varepsilon^t_{ext}, \quad \omega^{Nt} = 0, \quad \omega^{ox} = \omega^{ox}_{env}, \quad \text{on } 0 < r < R \text{ at } t = 0 \tag{12}$$

3 RESULTS AND DISCUSSION

In [12], we have performed numerical simulations of the equations in (2)-(4) and we have recorded the resulting growth curves, namely the evolution of the spheroid radius over time. We have analyzed both the case of spheroids freely growing in their culture medium, and the case where an external compression is applied. The values of the governing parameters are obtained from the literature, when they are available, and from the fit of the experimental curves. In this work, we investigate the dependence of the growth curves on a set of these parameters, summarized in Table 1.

Parameter	Description	Reference value	Unit
R_0	Initial radius of the spheroid	145	μm
ω^{ox}_{crit}	Critical mass fraction of oxygen	2.0×10^{-6}	(−)
γ^t_g	Coefficient related to growth	5.4×10^{-3}	$\mathrm{kg/(m^3\cdot s)}$
γ^t_n	Coefficient related to necrosis	1.5×10^{-1}	$\mathrm{kg/(m^3\cdot s)}$
λ^t_l	Coefficient related to lysis	1.15×10^{-2}	$\mathrm{kg/(m^3\cdot s)}$
α	Coefficient in the definition of Σ	10^5	Pa

Table 1: Parameters considered in this work. The reference value is the one used in [12].

We start our analysis from R_0, the initial radius of the spheroid. In Figure 2 we report the behavior of the growth curve for different initial spheroid radii. Note that in all the following figures the curve in red is the one corresponding to the reference value in Table 1. We consider spheroids with initial radii of 90, 117.5, 145, 172.5 and 200 μm. For all the different conditions, in particular for the small initial radii, it is possible to visualize the three stages characterizing the growth of the spheroids [13]: the exponential phase in the first days of growth, where the cells proliferate in a nutrient-rich environment; the linear phase, where the the tumor mass becomes larger and the nutrient starts to run low; and the growth saturation phase, where a significant portion of the spheroid is necrotic and only a small rim of cell proliferates at the tumor border. Notably, although the spheroids with the larger initial radii are constitued by more tumor cells than the others, they reach a similar final radius, of about 475 μm. This behavior is consistent with our assumption of growth as limited by nutrient deprivation. For a fixed level of external oxygen (ω^{ox}_{env}), only a fixed number of TCs is allowed to coexist in the spheroid mass. This condition is met sooner for the spheroids with larger initial radii and later for the others, as shown in the curves and observed experimentally in [12].

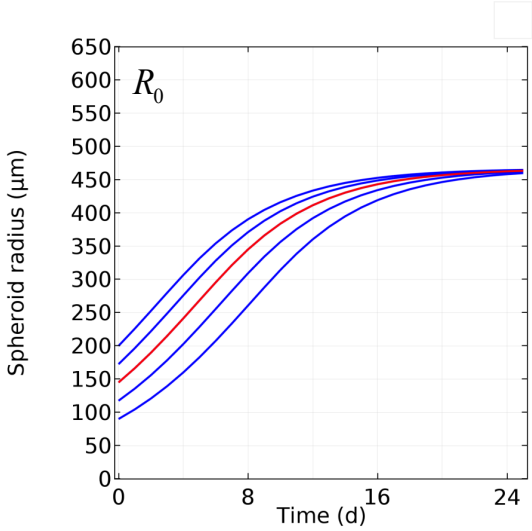


Figure 2: Spheroid growth curves for different initial radii.

Next, we study the effect of a variation in the critical level of oxygen ω_{crit}^{ox}. The resulting growth curves are shown in Figure 3. We consider values for ω_{crit}^{ox} of 1.0×10^{-6}, 2.0×10^{-6}, 3.0×10^{-6}, 4.0×10^{-6}, 5.0×10^{-6} and 6.0×10^{-6} (the black arrow points in the direction of increasing values of ω_{crit}^{ox}; this holds true also for the following figures). The choice of this parameter affects significantly the final radius reached by the spheroids. In particular, higher values of the critical level of oxygen provide smaller final radii. This follows from the modeling choice in equation (5), where cell proliferation is a linear function of the oxygen critical level. If this threshold value is high when compared to the external mass fraction of oxygen, only a small fraction of the spheroid is able to proliferate and the final radius is reduced.

The next parameter that we consider is the growth coefficient γ_g^t. In Figure 4, we consider values of this coefficient that are ±25, ±50 and ±75% of the reference value. Also for this parameter, the final radius reached by the spheroid strongly depends on its value. Interestingly, γ_g^t seems to regulate the time scale of the phenomenon. Actually, for small values of the growth coefficient, at the end of the simulations the spheroid is still in the first stages of growth. On the contrary, for higher values of γ_g^t the spheroid reaches faster its final radius. We also observe that the steady radius increases for increasing values of the growth coefficient. In fact, growth saturation is established when the net production of new tumor mass is zero. This means that, in a given time interval, the number of new cells produced by growth has to be equivalent to the number of cells undergoing lysis. This number is approximately given by the number of necrotic cells times the lysis rate. If the lysis rate is fixed, as in the present case, and the growth coefficient is increased, there is a net production of TCs and the radius grows further. To reach a new steady condition, the necrotic core of the spheroid has to increase in size, so that the number of cells undergoing lysis is again equal to the number of generated TCs. This idea can be tested through a variation of the parameters that regulate cell death, namely γ_n^t and λ_l^t, together with a consistent variation of the growth coefficient γ_g^t. If we double each of these constants at the same time, we expect the steady radius of the spheroid to be unaltered, since we have maintained the original ratio between cell production and removal. This condition is shown in Figure 4, by the gray dashed line. Note that the final radius is the same of the reference value, but the time needed to reach the steady state is significantly reduced.

Then, we consider the case of a variation in the coefficient γ_n^t, regulating the necrosis of the LTCs. The influence of this parameter is shown in Figure 5. For this coefficient, we ana-

lyze values that are ±25, ±50 and +75% of the reference value. As shown in the figure, the variation of γ_n^t has a very little effect on the resulting growth curves. The shape of the curve is not significantly altered and the final radius has a variation of less than 5%. This is probably due to the fact that this term is active on a population of TCs that is still alive, but it is located in a region of the spheroid where the nutrient level is below the critical threshold.

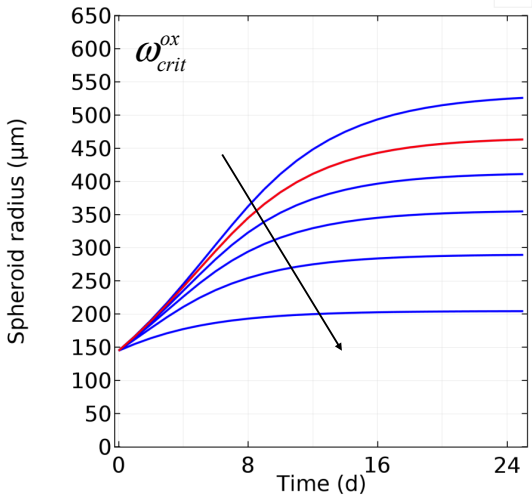


Figure 3: Spheroid growth curves for different critical levels of oxygen.

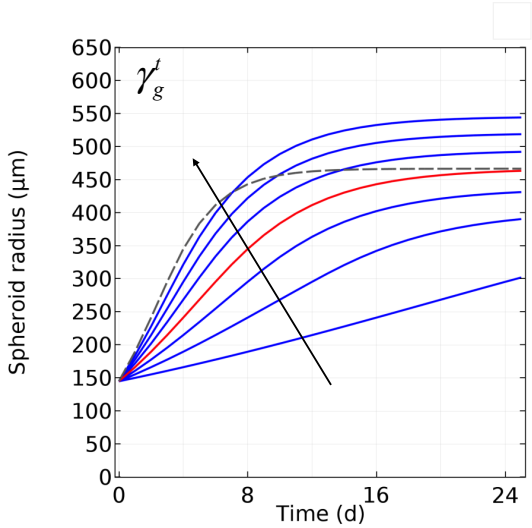


Figure 4: Spheroid growth curves for different values of the growth coefficient.

This condition is poorly encountered in the spheroids at this stage of their growth [14], resulting in the small influence of this parameter.

The fifth parameter that we study is the lysis coefficient λ_l^t. Figure 6 shows the results of considering values for this parameter that are ±25, ±50 and ±75% of its reference value. We first observe that the effects of varying the parameter only appear after day 8 in the simulation. This is consistent with the onset of a necrotic population inside the spheroid, which occurs after the first days of the simulation. These results show that the value of the lysis coefficient has a significant impact on the spheroid final radius. Notably, there is a saturation effect for high values of λ_l^t. This may be due to the limited amount of NTCs that exists in the necrotic core at fixed γ_n^t, and that can therefore undergo lysis.

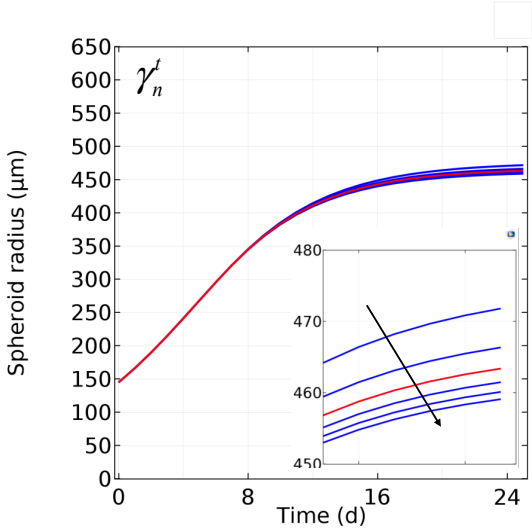


Figure 5: Spheroid growth curves for different values of the necrosis coefficient.

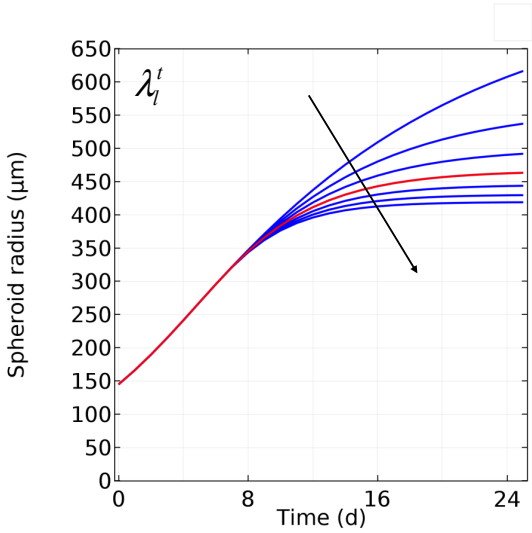


Figure 6: Spheroid growth curves for different values of the lysis coefficient.

Finally, we study the effects of varying α in the mathematical expression for Σ. The value of the derivative Σ' is directly proportional to this parameter, as shown in [12]. We vary the value of α for the ±25, ±50 and ±75% respect to its reference value. Even though we apply a significant variation, there is no apparent effect in the growth curves, which appear superimposed (results not shown). This result may point to the fact that the dynamics of the system, at least for the set of parameters considered, is mainly governed by the constitutive relations for the mass exchange between the phases.

4 CONCLUSIONS

In this work, we have performed a parametric study on a recent mathematical model for tumor spheroid growth [12]. The influence of a set of parameters on the growth curves of the spheroids has been evaluated, and we have provided a discussion of the results. In summary, some of the parameters show a little effect on the growth dynamics, such as the mechanical

coefficient α and the coefficient related to necrosis. On the other hand, other parameters have a significant impact on the final radius reached by the spheroids, namely the critical oxygen level, the coefficient related to growth and the coefficient related to lysis.

This work is certainly open to a number of improvements. In particular, we considered only one nutrient species, namely oxygen, limiting the growth of the tumor mass. Even though the influence of other chemicals is implicitly incorporated in the mass exchange term in (5), the inclusion of different nutrients, growth and necrosis factors could provide additional insights into the evolution of the tumor system [15,16]. Moreover, as it is common in the literature, most of the laws defining the constitutive relations for the mass exchange terms are derived from phenomenological arguments. It is desirable that they could be inferred from experimental measurements and linked to their biochemical and biomechanical background. Finally, here we consider a simple mechanical picture of the tumor tissue, explicitly dependent on the volume fraction of the TCs. Although this assumption describes conveniently the experimental data, it does not take into account several phenomena related to the mechanics of cell interactions at the macroscale, such as their rearrangement after the breakage of the cellular bonds [17].

In the future, we aim to design a new set of experiments that will provide better estimates of the model parameters and help the derivation of the constitutive relations. A better characterization of the interactions between the cancer cells and their microenvironment should offer important insights into the understanding of the disease, and for designing new treatments.

ACKNOWLEDGEMENTS

The partial support by the Italian National Research Project 60A09-0225 "Modellazione fisico-matematica e numerica di materiali porosi multifase basata su modelli termodinamicamente coerenti" is gratefully acknowledged.

REFERENCES

[1] Stewart, B. and Wild, C.P., editors. (2014) World Cancer Report 2014.

[2] Longo, D., Fauci, A., Kasper, D. and Hauser, S. (2011) Harrison's principles of internal medicine. McGraw-Hill Professional.

[3] Michor, F., Liphardt, J., Ferrari, M. and Widom, J. (2011) What does physics have to do with cancer? *Nature Reviews Cancer*, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. **11**, 657–70. http://dx.doi.org/10.1038/nrc3092

[4] Giverso, C. and Preziosi, L. (2012) Modelling the compression and reorganization of cell aggregates. *Mathematical Medicine and Biology*, **29**, 181–204. http://dx.doi.org/10.1093/imammb/dqr008

[5] Pascal, J., Bearer, E.L., Wang, Z., Koay, E.J., Curley, S. a and Cristini, V. (2013) Mechanistic patient-specific predictive correlation of tumor drug response with microenvironment and perfusion measurements. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14266–71. http://dx.doi.org/10.1073/pnas.1300619110

[6] Leder, K., Pitter, K., Laplant, Q., Hambardzumyan, D., Ross, B.D., Chan, T.A. et al. (2014) Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell*, **156**, 603–16. http://dx.doi.org/10.1016/j.cell.2013.12.029

[7] Altrock, P.M., Liu, L.L. and Michor, F. (2015) The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, Nature Publishing Group. **15**, 730–45. http://dx.doi.org/10.1038/nrc4029

[8] Sciumè, G., Shelton, S., Gray, W.G., Miller, C.T., Hussain, F., Ferrari, M. et al. (2013) A multiphase model for three-dimensional tumor growth. *New Journal of Physics*, **15**, 015005. http://dx.doi.org/10.1088/1367-2630/15/1/015005

[9] Sciumè, G., Gray, W.G., Hussain, F., Ferrari, M., Decuzzi, P. and Schrefler, B. A. (2013) Three phase flow dynamics in tumor growth. *Computational Mechanics*, **53**, 465–84. http://dx.doi.org/10.1007/s00466-013-0956-2

[10] Sciumè, G., Santagiuliana, R., Ferrari, M., Decuzzi, P. and Schrefler, B.A. (2014) A tumor growth model with deformable ECM. *Physical Biology*, **11**, 65004. http://dx.doi.org/10.1088/1478-3975/11/6/065004

[11] Santagiuliana, R., Stigliano, C., Mascheroni, P., Ferrari, M., Decuzzi, P. and Schrefler, B. a. (2015) The role of cell lysis and matrix deposition in tumor growth modeling. *Advanced Modeling and Simulation in Engineering Sciences*, Springer International Publishing. **2**, 19. http://dx.doi.org/10.1186/s40323-015-0040-x

[12] Mascheroni, P., Stigliano, C., Carfagna, M., Boso, D.P., Preziosi, L., Decuzzi, P. et al. (2016) Predicting the growth of glioblastoma multiforme spheroids using a multiphase porous media model. *Biomechanics and Modeling in Mechanobiology*, Springer Berlin Heidelberg. http://dx.doi.org/10.1007/s10237-015-0755-0

[13] Sutherland, R.M., McCredie, J. a and Inch, W.R. (1971) Growth of multicell spheroids in tissue culture as a model of nodular carcinomas. *Journal of the National Cancer Institute*, **46**, 113–20. http://dx.doi.org/10.1093/jnci/46.1.113

[14] Mikhail, A.S., Eetezadi, S. and Allen, C. (2013) Multicellular tumor spheroids for evaluation of cytotoxicity and tumor growth inhibitory effects of nanomedicines in vitro: A comparison of Docetaxel-loaded block copolymer micelles and Taxotere®. He X, editor. *PLoS ONE*, **8**, e62630. http://dx.doi.org/10.1371/journal.pone.0062630

[15] Chauhan, V.P. and Jain, R.K. (2013) Strategies for advancing cancer nanomedicine. *Nature Materials*, Nature Publishing Group. **12**, 958–62. http://dx.doi.org/10.1038/nmat3792

[16] Jain, R.K., Martin, J.D. and Stylianopoulos, T. (2014) The role of mechanical forces in tumor growth and therapy. *Annual Review of Biomedical Engineering*, **16**, 321–46. http://dx.doi.org/10.1146/annurev-bioeng-071813-105259

[17] Preziosi, L. and Vitale, G. (2011) A multiphase model of tumor and tissue growth including cell adhesion and plastic reorganization. *Mathematical Models and Methods in Applied Sciences*, **21**, 1901–32. http://dx.doi.org/10.1142/S0218202511005593

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

STABILITY OF ASYMMETRIC GRAIN BOUNDARIES IN GRAPHENE

J. P. Mendez[1], F. Arca[2] and M. P. Ariza[2]

[1]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: jmendezg@caltech.edu

[2] Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, Camino de los descubrimientos, s.n., 41092 Sevilla, Spain
e-mail: {farca,mpariza}@us.es

Keywords: Graphene, Asymmetric grain boundaries, Thermal stability.

Abstract. *In the field of electronics, due to its excellent mechanical and electrical properties, graphene has become the most promising material for the production of next generation thin and flexible graphene-based electronic components. In this work, we present an assessment of the thermal stability and dynamics of asymmetric grain boundaries in graphene for different misorientation angles at finite temperature and up to extremely high temperatures. In particular, we have focused on configurations with misorientation angle of 16.1°, 30° and 38.2°. In contrast to pristine defect-free graphene, which has no band-gap and therefore is of limited use for semiconductor-based electronics, it has been shown theoretically that line defects in graphene might insert transport gaps, opening up the possibility of device applications based on the structural engineering of graphene boundaries.*

1 INTRODUCTION

Graphene, owing to its extraordinary properties [1, 2], can be used to a wide range of applications: aerospace, where carbon fibers that are used to reinforce composite materials can be replaced by graphene and, as a result, the weight of aircraft might be reduced considerably as well as the electrical conductivity will be enhanced for lightning strikes during flights; military industry, where graphene can be used as substitution of Kevlar in bulletproof vest [3]; bioengineering, where graphene can be used to fabricate bioelectric sensors [4, 5]; or in the field of electronics, where graphene has become the most promising material for the production of thin and flexible electronic components of next generation [6, 7].

However, the use of graphene in semiconductor-based devices is limited due to the fact that free-defect graphene lacks band gap, which is crucial for the correct operation of these devices. In order to achieve an electrical function and due to the presence of band gap, semiconductors exhibit the ability to switch currents on and off as desired. Many attempts have been carried out in order to introduce this desired band gap, e.g., by doping graphene with boron and nitrogen [8] or by engineering defects such as vacancies, Stone-Wales and dislocations [9], and grain boundaries (GBs) [10]. It has been proved, either experimentally or theoretically, that the presence of point defects modify locally the electronic properties [11, 12], but do not endow the layer with a band gap wide enough for such applications. Nevertheless, it has been seen that the presence of grain boundaries alters the electronic properties, and more specifically, some configurations might introduce a transport gap across them that would block the charge carriers in a range of energy. As a consequence, due to the current technology advances, the design of suited GB configurations opens a new potential via to tailor a tunable transport gap in graphene and, therefore, to use graphene as the base material for the manufacturing of semiconductor-based devices, e.g., nanotransistors [13].

With the main target of tuning the carrier transport in graphene, we will carry out an appraisal of the stability and dynamics of asymmetric GBs that could appear in graphene lattice. Additionally, owing to the fact that temperature is always present during the performance of electronic devices, the thermal stability of these GBs will play an important role when designing electronic nanocomponents. Therefore, in this work we will also conduct an analysis of the thermal stability of the predicted GB configurations during a long period of time and up to extremely high temperatures. To this end, we will proceed by applying the same methodology followed by the authors for the thermal characterization of dislocations [14] and symmetric grain boundaries [10] in graphene.

2 ASYMMETRIC GRAIN BOUNDARIES IN GRAPHENE

The simulation model used for studying the stability of GBs is depicted in Fig. 1. This model consists of two grains (A and B) with different crystallographic orientations (θ_1 and θ_2) and periodic boundary conditions applied in both directions (X and Y axis). The boundary effects are an obstacle that one has to tackle in this kind of simulations and usually emerge when the cell size is small. Owing to avoid these effects, the simulation cell has to be as large as possible, therefore, the cell size is of the order of thousands of atoms ($\sim 10^3$) in this study.

The misorientation angle, θ, is defined as $\theta = \theta_1 + \theta_2$, where θ_1 and θ_2 are the orientation angles of grain A and B with respect to the horizontal axis X, respectively. Another interesting parameter to be considered here is the inclination angle of the boundary line, ϕ, that is defined as $\phi = |\theta_1 - \theta_2|$. On the other hand, inasmuch as graphene crystallographic structure possesses two high symmetry directions, i.e., the zig-zag and armchair directions (see Fig. 2), we fix the

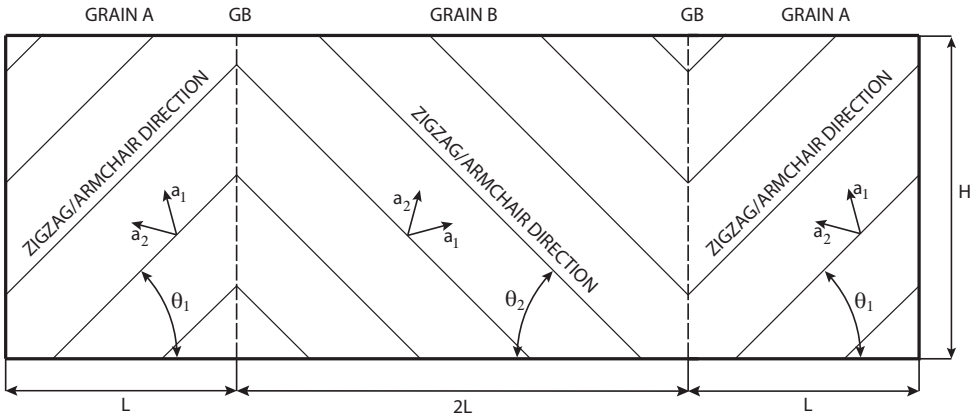
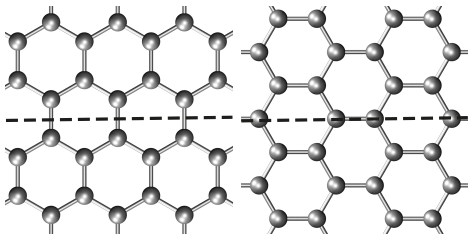


Figure 1: Grain boundary model.

zig-zag direction when the orientation angle of grain θ_i ($i = 1, 2$) is equal to 0° and therefore the armchair direction is reached when the orientation angle θ_i ($i = 1, 2$) is equal to 30°.



(a) (b)

Figure 2: High symmetry directions of graphene: (a) zig-zag direction and (b) armchair direction.

Regarding grain boundary configurations, we can distinguish two types: symmetric and asymmetric. When $\theta_1 = \theta_2$, GBs are symmetric and the inclination angle is $\phi = 0$, whereas GBs are asymmetric when $\theta_1 \neq \theta_2$ and, therefore, $\phi = \theta_1 - \theta_2 \neq 0$. For the purpose of evaluating all possible GB configurations, it is only necessary to cover the range of θ_1 and θ_2 that varies between $0°$ and $30°$. Therefore, symmetric configurations posses one degree of freedom θ (θ_1 or θ_2) and have to be defined by the misorientation angle, and asymmetric GBs have two degrees of freedom (θ_1 and θ_2) and can be denoted by ($\theta_1 \mid \theta_2$). Additionally, as GBs are traditionally modelled as a periodic arrangement of dislocations [15], they can be described by $(n_A, m_A) \mid (n_B, m_B)$, where (n_i, m_i) $i = A, B$ represents the periodic translations vector of grains A and B along GB and referred to their respective basis set (a_1,a_2). In particular, $n_A = n_B$ and $m_A = m_B$ for symmetric configurations. Finally, the misorientation angle can be calculated in a straightforward manner through

$$\theta = \arctan((n_A - m_A)/\sqrt{3}/(m_A + n_A)) + \arctan((m_B - n_B)/\sqrt{3}/(m_B + n_B)) \quad (1)$$

2.1 Structures and energies

With the main purpose of finding misorientation angles corresponding to stable configurations at 0K, we have simulated an endless number of configurations with different angles by means of molecular dynamics, covering a wide range of possibilities. Owing to the fact that $\theta_1 \neq \theta_2$ for asymmetric GBs, the characterization of these GBs is more challenging than for symmetric configurations [10]. To accomplish this, we have resorted the widely used molecular

Misorientation angle(θ) θ (θ_A|θ_B)	GB configuration (n_A,m_A)|(n_B,m_B)	Energy (eV/Å)
16.1° (16.1°|0°)	(3,1)|(2,2)	0.932
16.1° (0°|16.1°)	(2,2)|(1,3)	0.323
30° (30°|0°)	(5,0)|(3,3)	0.361
30° (0°|30°)	(3,3)|(0,5)	1.241
38.2° (8.21°|30°)	(5,3)|(0,7)	0.394

Table 1: Energies for different asymmetric grain boundaries in graphene.

dynamics code LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) and the AIREBO [16] and ReaxFF potentials [17] that are implemented in LAMMPS code. The strategy proceeded to find these configurations is as follows: first, employing the ReaxFF potential and NVE ensemble, we grow two grains with different misorientation angles and force them to interact and form an interface between both grains. It should be noted that sometimes it is necessary to remove atoms along the interface in order to achieve the desired grain boundary structure; and second, we stabilize the GB structure using the AIREBO potential and the Nose-Hoover thermostat (NVT ensemble) at very low constant temperature ($\sim$0K).

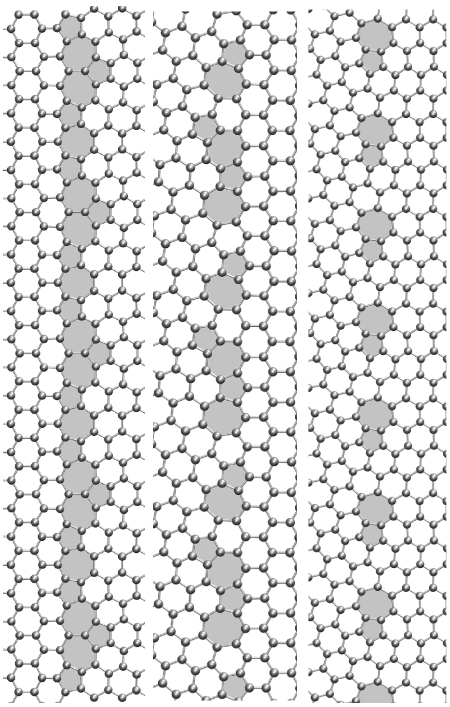
In our simulations, we have adopted that GBs are constituted by periodic arrays of dislocations. This periodicity can be defined by two periodic translation vectors, (n_A,m_A) and (n_B,m_B), belonging to grains A and B, respectively, and parallel to the GB (see Fig. 1). Unlike symmetric configurations, these translation vectors might differ slightly in magnitude for some asymmetric configurations, $d_A \neq d_B$ ($d_A = 2.46(n_A \cos(30 - \theta_1) + m_A \cos(30 + \theta_1))$ and $d_B = 2.46(n_B \cos(30 + \theta_2) + m_B \cos(30 - \theta_2))$). However, in order to form stable asymmetric GB, the mismatch between both domains at the GB must be quite small and its stored energy must be reasonably low. In order to evaluate the energies and structures of these asymmetric configurations, the height of the computational cell could be considered proportional to d_A or to d_B. In our model, we take the height of the computational cell H proportional to d_A.

In Fig. 3, we present the obtained asymmetric configurations in our calculations, and the corresponding energies for all possible cases and the misorientation angles of $16.1°$, $30°$ and $38.2°$ are tabulated in Table 2.1. First, we have found that asymmetric configurations do not follow the same pattern as symmetric configurations, and a large number of dislocation types and distribution along the grain boundary can be found. For instance, the GB structures with misorientation angles of $\theta = 30°$ and $\theta = 38.2°$ are constructed by (1,0)+(0,1) and (1,0) dislocations, whereas (1,0) dislocations are arranged along an interface of two grains with a misorientation angle of $16.1°$ (see Figs. 3 and 4). These results have also been corroborated by previous studies [18, 19]. Second, we have found that our low energy configurations correspond to those with $d_A < d_B$. Third, for these low energy configurations, our simulations show the existence of out-of-plane displacement, about 0.5 Å.

Finally, these results have been also corroborated experimentally, [20] observed GBs with misorientation angles of $\sim 16°$ and $\sim 29°$ which can be related to the angles of $16.1°$ and $30°$ in our calculations, respectively.

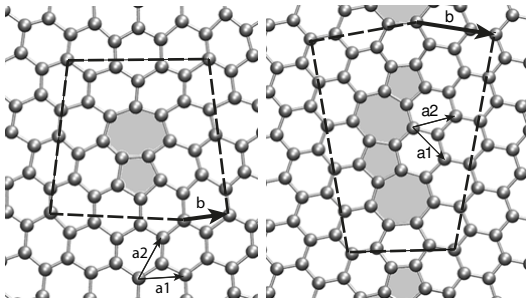
3 Thermal stability

Afterwards, we have analysed the mechanical behaviour and thermal stability at finite temperature of these GB structures up to extremely high temperatures. For this purpose, we have relaxed each configuration during a long period of time, up to 10.000ps, and at fixed constant



(a) (b) (c)

Figure 3: Observed structures of GB in graphene lattice (a) 30°|0° (θ=30°) or (5,0)|(3,3) (b) 8.21°|30° (θ=38.21°) or (5,3)|(0,7) and (c) 16.1°|0° (θ=16.1°) or (3,1)|(2,2).



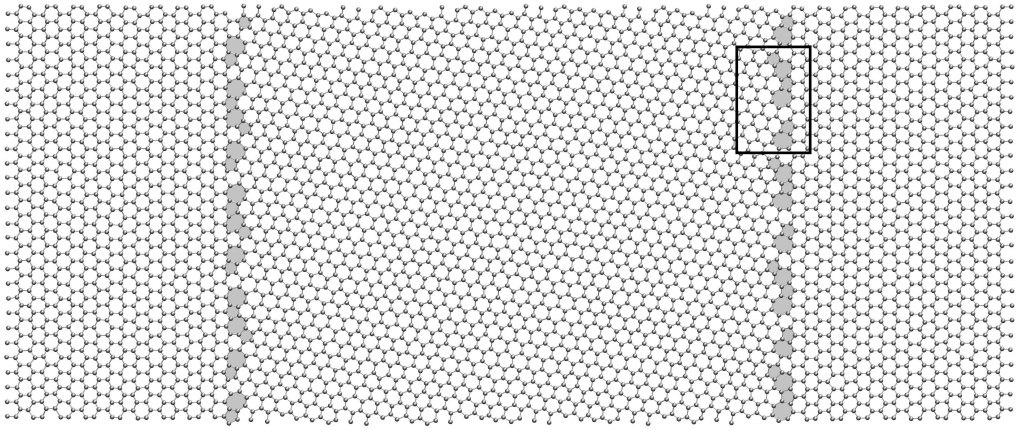
(a) (b)

Figure 4: Dislocation type (a) (1,0) and (b) (1,0)+(0,1).

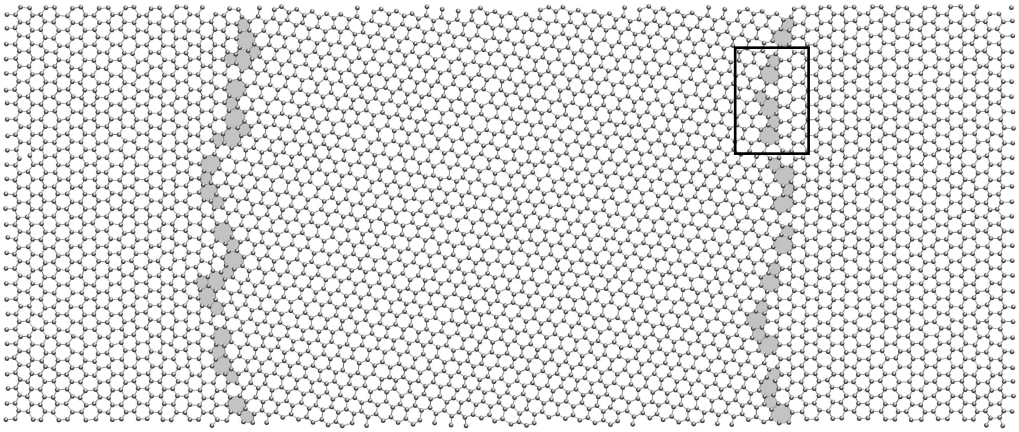
temperature, covering a wide range of temperatures from 0K to temperatures close to the melting point of graphene [21, 22, 23]. Similarly, to carry out the thermal stability analysis, we have employed the LAMMPS code, the full AIREBO potential and the NVT ensemble (Δt = 0.1 fs). As main outcomes of our simulations, we state the following conclusions:

1. Asymmetric configurations remain stable up to very high temperature, around 3000K. As we can observe from Fig. 5a for a misorientation angle of $38.2°$, dislocations type (1,0) and (1,0)+(0,1) which form the grain boundary are unaffected by the temperature.

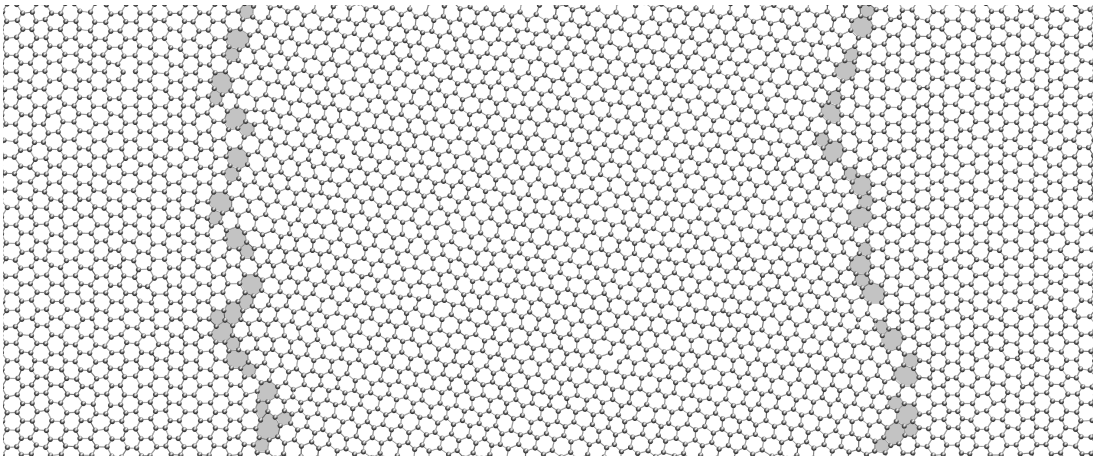
2. At temperature exceeding 3000K, it is observed transformations, motions and slip-up of dislocations. For instance, as we can see in the rectangles marked in Figs. 5b and c, a (1,0)+(0,1) dislocation segregate into two dislocations, and the resulting dislocations move and one of them joins to a (1,0) dislocation to form new dislocation. Additionally, we show in Fig. 6 as an example the glide motion of a (1,0) dislocation.



(a)

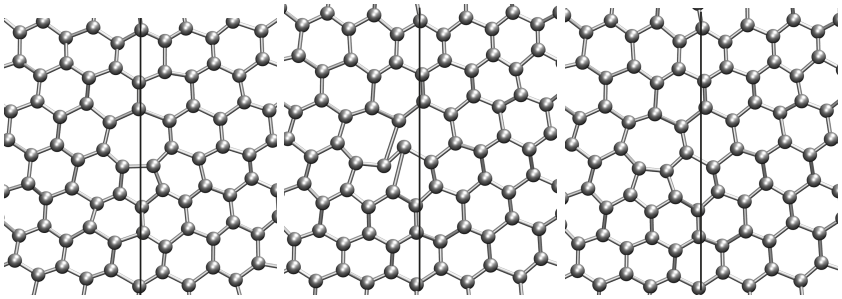


(b)



(c)

Figure 5: Observed asymmetric grain boundary structures for a misorientation angle of 38.2° at different temperatures (a) T$\leq$3000K, (b) T=3500K, and (c) T=4000K.



(a) (b) (c)

Figure 6: Details of the glide motion of (1,0) dislocations showing at temperatures above 2000K.

3. Above 4000K, asymmetric GBs collapse. Similar to symmetric configurations, the presence of GBs accelerates the melting process and decreases the melting temperature. Therefore, these defects act as a trigger for melting.

4 CONCLUSIONS

In the present work, we have analysed the dynamics and the thermal stability of asymmetric GBs for different misorientation angles. First, we have obtained atomic configurations corresponding to stable GBs using the LAMMPS code and the AIREBO and ReaxFF potentials. In particular, we have obtained that the misorientation angles $16.1°$, $30°$ and $38.2°$ are stable in graphene. Afterwards, we have relaxed these configurations at finite temperature up to very high temperature, close to the melting temperature for graphene. As results, we have reported a high stability against annihilation up to extreme temperatures, about 3000K, high mobility and transformation of dislocations at higher temperatures and the melting of grain boundaries at temperarutes above 4000K.

REFERENCES

[1] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. The electronic properties of graphene. *Rev. Mod. Phys.*, 81:109–162, 2009.

[2] I. W. Frank, D. M. Tanenbaum, A. M. Van der Zande, and P. L. McEuen. Mechanical properties of suspended graphene sheets. *Journal of Vacuum Science & Technology B*, 25(6):2558–2561, 2007.

[3] Jae-Hwang Lee, Phillip E. Loya, Jun Lou, and Edwin L. Thomas. Dynamic mechanical behavior of multilayer graphene via supersonic projectile penetration. *Science*, 346(6213):1092–1096, 2014.

[4] Y. Shao, J. Wang, H. Wu, J. Liu, I.A. Aksay, and Y. Lin. Graphene based electrochemical sensors and biosensors: A review. *Electroanalysis*, 22(10):1027–1036, 2010.

[5] Z. Tang, H. Wu, J.R. Cort, G.W. Buchko, Y. Zhang, Y. Shao, I.A. Aksay, J. Liu, and Y. Lin. Constraint of dna on functionalized graphene improves its biostability and specificity. *Small*, 6(11):1205–1209, 2010.

[6] S. Bae, H. Kim, Y. Lee, X. Xu, J.S. Park, Y. Zheng, J. Balakrishnan, T. Lei, H. R. Kim, Y. I. Song, Y.J. Kim, K.S. Kim, B. zyilmaz, J.H. Ahn, B. H. Hong, and S. Iijima. Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnology*, 5(8):574–578, 2010.

[7] Y.U. Jung, K.W. Park, S.T. Hur, S.W. Choi, and S.J. Kang. High-transmittance liquid-crystal displays using graphene conducting layers. *Liquid Crystals*, 0(0):1–5, 2013.

[8] Pooja Rani and VK Jindal. Designing band gap of graphene by b and n dopant atoms. *RSC Advances*, 3(3):802–812, 2013.

[9] M.P. Ariza and J.P. Mendez. Stability of discrete topological defects in graphene. *Journal of Mechanics of Materials and Structures*, 10(2):239–253, 2015.

[10] J.P. Mendez, F. Macias, and M.P. Ariza. Thermal stability of tilt grain boundaries in graphene. *Applied Mechanics and Materials*, 481:129–132, 2014.

[11] Sharmila N. Shirodkar and Umesh V. Waghmare. Electronic and vibrational signatures of stone-wales defects in graphene: First-principles analysis. *Phys. Rev. B*, 86:165401, Oct 2012.

[12] A. Lherbier, S.M.M. Dubois, X. Declerck, Y.M. Niquet, S. Roche, and J.C. Charlier. Transport properties of graphene containing structural defects. *Physical Review B*, 86:075402, Aug 2012.

[13] B. Standley, W. Bao, J. Zhang, H.and Bruck, C. N. Lau, and M. Bockrath. Graphene-based atomic-scale switches. *Nano Letters*, 8(10):3345–3349, 2008.

[14] M.P. Ariza, M. Ortiz, and R. Serrano. Long-term dynamic stability of discrete dislocations in graphene at finite temperature. *International Journal of Fracture*, 166(1-2):215–223, 2010.

[15] W. T. Read and W. Shockley. Dislocation models of crystal grain boundaries. *Phys. Rev.*, 78:275–289, May 1950.

[16] S. J. Stuart, A. B. Tutein, and J. A. Harrison. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.*, 112(14):6472–6486, 2000.

[17] Adri C. T. van Duin, Siddharth Dasgupta, Francois Lorant, and William A. Goddard. Reaxff: A reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[18] Jihoon Han, Seunghwa Ryu, Dongwoo Sohn, and Seyoung Im. Mechanical strength characteristics of asymmetric tilt grain boundaries in graphene. *Carbon*, 68(0):250 – 257, 2014.

[19] Xiuyun Zhang, Ziwei Xu, Qinghong Yuan, John Xin, and Feng Ding. The favourable large misorientation angle grain boundaries in graphene. *Nanoscale*, 7(47):20082–20088, 2015.

[20] Jinho An, Edgar Voelkl, Ji Won Suk, Xuesong Li, Carl W. Magnuson, Lianfeng Fu, Peter Tiemeijer, Maarten Bischoff, Bert Freitag, Elmira Popova, and Rodney S. Ruoff. Domain (grain) boundaries and evidence of twinlike structures in chemically vapor deposited grown graphene. *ACS Nano*, 5(4):2433–2439, 2011. PMID: 21361332.

[21] K V Zakharchenko, Annalisa Fasolino, J H Los, and M I Katsnelson. Melting of graphene: from two to one dimension. *Journal of Physics: Condensed Matter*, 23(20):202202, 2011.

[22] Y Kowaki, A Harada, F Shimojo, and K Hoshino. Radius dependence of the melting temperature of single-walled carbon nanotubes: molecular-dynamics simulations. *Journal of Physics: Condensed Matter*, 19(43):436224, 2007.

[23] Seong Gon Kim and David Tománek. Melting the fullerenes: A molecular dynamics study. *Phys. Rev. Lett.*, 72:2418–2421, Apr 1994.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DEFORMATION-DIFFUSION COUPLED ANALYSIS OF LONG-TERM HYDROGEN DIFFUSION IN NANOFILMS

X. Sun[1], **M. P. Ariza**[2] **and K. G. Wang**[1]

[1] Department of Aerospace and Ocean Engineering, Virginia Polytechnic Institute and State University
Blacksburg, VA 24061, USA
e-mail: {xssun, kevinwgy}@vt.edu

[2] Escuela Técnica Superior de Ingeniería, Universidad de Sevilla
Sevilla 41092, Spain
mpariza@us.es

Keywords: Deformation-Diffusion Coupling, Non-Equilibrium Statistical Mechanics, Mean-field Theory, Discrete Kinetic Law, Long-Term Processes, Hydrogen Diffusion.

Abstract. *The absorption and desorption of hydrogen in nanomaterials can be characterized by an atomic, deformation-diffusion coupled process with a time scale of the order of seconds to hours. This time scale is beyond the time windows of conventional atomistic computational models such as molecular dynamics (MD) and transition state theory based accelerated MD. In this paper, we present a novel, deformation-diffusion coupled computational model basing on non-equilibrium statistical mechanics, which allows long-term simulation of hydrogen absorption and desorption at atomic scale. Specifically, we propose a carefully designed trial Hamiltonian in order to construct our meanfield based approximation, then apply it to investigate the palladium-hydrogen (Pd-H) system. Specifically, here we combine the meanfield model with a discrete kinetic law for hydrogen diffusion in palladium nanofilms. This combination in practice defines the evolution of hydrogen atomic fractions and lattice constants, which facilitates the characterization of the deformation-diffusion process of hydrogen over both space and time. Using the embedded atom model (EAM) potential, we investigate the deformation-diffusion problem of hydrogen desorption and absorption in palladium nanofilms and compare our results with experiments both in equilibrium and non-equilibrium cases.*

1 INTRODUCTION

In a large number of areas, the behavior of material systems depends sensitively on properties that pertain to the atomistic scale. Thus, nanotechnology has made an extensive contribution to opening up novel frontiers in material science and engineering, particularly in the design and manufacture of more efficient and reliable materials for energy storage applications. Due to its high-energy, easy availability, and non-toxicity, hydrogen is a promising energy carrier as a potential substitute for fossil fuels in future transport applications. In particular, the storage of hydrogen in metals is one of the technologies that has been the focus of extensive research. However, there exist some failure phenomena in hydrogen-metal systems that constitute the main barriers against their use in the large scale. Hydrogen embrittlement is one of the most common failure processes, which reduces the mechanical properties of metals due to the inclusion of hydrogen atoms [1, 2]. In order to prevent most of the failures associated to the degradation of metals, a natural starting point is to characterize and understand the hydrogen transport process in metals. This gives rise to the development of predictive design of nanostructured materials for hydrogen-storage.

The absorption and desorption of hydrogen in nanomaterials is characterized by an atomic deformation-diffusion process on a length scale of Angstroms and a time scale of femtoseconds. Nonetheless, the properties and behaviors of interest in real-world engineering applications (e.g. fuel cell vehicles) are macroscopic and take place on the scales of centimeters to meters, and are characterized by a relatively slow evolution on the scale of seconds to years. This vast disparity of length and time scales poses extraordinary challenges in theoretical and computational material science, as well as material design. Most computational efforts up to date have relied on traditional first-principles-based Molecular Dynamics (MD) or Monte Carlo (MC) methods [3, 4] as their chief representational and computational paradigm. However, these two computational methods are limited to relatively small material samples and to time windows of microseconds at best. Considerable efforts have been devoted to accelerating MD and MC methods, such as accelerated MD [5, 6] and diffusive MD [7]. However, no computationally-tractable atomistically-based models appear to be available to study deformation-diffusion coupled problems with the vast disparity of length and time scales described above.

Our objective in this work is to enable long-term (e.g., up to seconds), predictive simulation of hydrogen absorption and desorption in metal-based nanomaterials, while maintaining a strictly atomistic description of the material. To achieve this objective, we apply the atomistic-continuum coupling approach proposed in Ref. [8]. In Ref. [8], a non-equilibrium statistical thermodynamics theory is proposed for multi-species particulate solids based on Jayne's maximum entropy principle and the meanfield approximation approach that allow the statistical treatment of systems away from equilibrium. Unlike in MD and MC, the individual hops of atoms are not explicitly tracked, but instead accounted for in the theory in a statistical sense. Most recently, this theoretical model has been applied to simulate hydrogen diffusion in Pd nanofilms [9]; however, the Pd sub-system was assumed to be rigid, and the phase-dependent deformation of the sample during hydrogen absorption/desorption was not accounted for.

This work is an extension of Ref. [9] to long-term, atomistic, deformation-diffusion coupled analysis. The metal (e.g. Pd) sub-system is no longer assumed rigid. Instead, using the non-equilibrium statistical thermodynamics model proposed in Ref. [8], the deformation of metal hydride will be characterized by the time-dependent mean atomic positions of both metal and hydrogen atoms. The empirical discrete kinetic model will be developed by extending the Fick's laws for continuum, and calibrated against experimental data and results obtained by first-principle simulations. The organization of the paper is as follows. Section 2 generalizes the non-equilibrium statistical model with the carefully designed trial Hamiltonians. In Section 3, we

will validate the proposed model of H diffusion in Pd nanofilms, accounting for both equilibrium and non-equilibrium cases. A summary of this work is finally presented in Section 4.

2 NON-EQUILIBRIUM STATISTICAL MODEL

We combine the atomistic-continuum coupling approach proposed in Ref. [8] with a novel meanfield trial model. This non-equilibrium statistical thermodynamics theory is proposed for multi-species particulate solids based on Jayne's maximum entropy principle and the meanfield approximation approach that can deal with a non-equilibrium problem.

2.1 Non-equilibrium statistical mechanics

We consider a discrete system consisting of N particle sites, each of which can be occupied by one of M species. At each particle site $i = 1, 2, \ldots, N$, and for each species $k = 1, 2, \ldots, M$, we introduce an *occupancy function* defined as

$$n_{ik} = \begin{cases} 1 & \text{if the site } i \text{ is occupied by the species } k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Thus, the microscopic states of the system can be defined by the instantaneous positions $\{\boldsymbol{q}\} = (\boldsymbol{q}_i)_{i=1}^N$, momenta $\{\boldsymbol{p}\} = (\boldsymbol{p}_i)_{i=1}^N$, and occupancy arrays $\{\boldsymbol{n}\} = (\boldsymbol{n}_i)_{i=1}^N$. It is assume that the statistics of the system obeys Jaynes' principle of maximum entropy [10, 11], namely maximizing the information-theoretical entropy

$$S[\rho] = -k_{\mathrm{B}} \langle \log \rho \rangle \tag{2}$$

among all probability measures of $\rho(\{\boldsymbol{q}\}, \{\boldsymbol{p}\}, \{\boldsymbol{n}\})$. k_{B} indicates Boltzmann's constant and $\langle A \rangle$ denotes the expected value of the observation A. We specifically consider the system consisting of distinguishable particles whose Hamiltonians have the additive structure of local Hamiltonian h_i. Suppose that the expected particle energies and the expected particle atomic fractions on each atomic site are known, namely

$$\langle h_i \rangle = e_i, \ \langle n_{ik} \rangle = x_{ik} \tag{3}$$

for $i = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, M$. It is worth noting that different from the classical equilibrium framework where only global constraints are enforced, here constraints are now local. Maximizing Eq. (2) with local constraints among probability measures results, after a straightforward calculation, in

$$\rho = \frac{1}{\Xi} \mathrm{e}^{-\{\beta\}^T \{h\} + \{\gamma\}^T \{\boldsymbol{n}\}} \tag{4}$$

where

$$\Xi = \sum_{\{\boldsymbol{n}\} \in O_{NM}} \frac{1}{h^{3N}} \int_{\Gamma} \mathrm{e}^{-\{\beta\}^T \{h\} + \{\gamma\}^T \{\boldsymbol{n}\}} \mathrm{d}\boldsymbol{q} \mathrm{d}\boldsymbol{p} \tag{5}$$

and $(\{\beta\}, \{\boldsymbol{\gamma}\})$ are Lagrange multipliers. h is Planck's constant. By comparison with equilibrium statistical mechanics, we may interpret Eq. (4) and Eq. (5) as non-equilibrium generalizations of the Gibbs grand-canonical probability density function and the grand-canonical partition function, respectively. Also, we may regard $T_i = 1/(k_{\mathrm{B}}\beta_i)$ and $\boldsymbol{\mu}_i = k_{\mathrm{B}} T_i \boldsymbol{\gamma}_i$ as the particle absolute temperature and the chemical potential array of particle i, respectively. It should be noted that unlike equilibrium statistical mechanics, the temperature and chemical

potential fields need not be uniform anymore and may vary from particle to particle when the system is away from equilibrium.

2.2 Variational meanfield theory

Despite the formal simplicity of the non-equilibrium statistical model just mentioned above, it is generally intractable to calculate the thermodynamic potentials in a closed form, which leads to the need for approximation theory. Venturini et al. [8] have extended the classical variational meanfield theory to systems described by the proposed non-equilibrium statistical framework. The resulting variational framework provides a convenient basis for the formulation of computationally tractable models. Specifically, the maximum-entropy principle in Eq. (2) with local constraints is equivalent to the minimality of the functional of free entropy

$$F\left(\{h_0\},\{\beta\},\{\gamma\}\right)=k_{\mathrm{B}}\{\beta\}^{\mathrm{T}}\left\{\left\langle h-h_0\right\rangle_0\right\}-\Phi_0 \tag{6}$$

over some class H_0 of the local trial Hamiltonians $\{h_0\}$ with $\Phi_0 = k_{\mathrm{B}}\log\Xi_0$.

We are considering a binary system consisting of metal atoms occupying a perfect lattice whose interstitial sites are either empty or occupied by H atoms. For ease of reference, we designate the metal atoms in the system by means of an index set I_{M} and the H atoms by an index set I_{H}. For the system under consideration, the occupancy field $\{\boldsymbol{n}\}$ can be restricted to the interstitial sites, I_{H} since the metal sites are always occupied. Hence, M = 1, $\{n\} = \{n_i, i \in I_{\mathrm{H}}\}$ and $n_i = 0$ if the i-th interstitial site is empty and $n_i = 1$ if it is occupied by H atom. We restrict our attention to quasistatic process, so the Hamiltonian is fully determined by the interatomic potential. In order to obtain explicit formulations of thermodynamic potentials suitable for computations, we combine the effects of positions and fractions and apply the meanfield approximation theory with carefully designed trial Hamiltonians

$$V_0\left(\{\boldsymbol{q}\},\{n\}\right)=\sum_{i\in I_{\mathrm{M}}\cup I_{\mathrm{H}}}k_{\mathrm{B}}T_i\alpha_{0i}\left|\boldsymbol{q}_i-\bar{\boldsymbol{q}}_i\right|^2-\sum_{i\in I_{\mathrm{H}}}k_{\mathrm{B}}T_i\gamma_{0i}n_i \tag{7}$$

where $\{\bar{\boldsymbol{q}}\}$ are the meanfield values for atomic instantaneous positions. $\{\alpha_0\}$ and $\{\gamma_0\}$ are coefficients which need to be determined. We consider throughout isothermal conditions at fixed temperature T, so that $\beta_i = 1/(k_{\mathrm{B}}T_i)$ is uniform at all sites of the system. By calculating the meanfield value for H occupancy function $\{n\}$**,** we have

$$\gamma_{0i}=\log\frac{x_i}{1-x_i}-\gamma_i \tag{8}$$

where we have used the relation $x_i = \langle n_i\rangle_0, \forall i \in I_{\mathrm{H}}$, as the meanfield H atomic fraction.

By using Eq. (8), the free entropy function which needs to be minimized can be expressed as

$$\begin{aligned}F\left(\{h_0\},\{\gamma\}\right)=&\frac{1}{T}\langle V\rangle_0+\sum_{i\in I_{\mathrm{H}}}\left(k_{\mathrm{B}}\log\frac{x_i}{1-x_i}-k_{\mathrm{B}}\gamma_i\right)x_i+k_{\mathrm{B}}\sum_{i\in I_{\mathrm{H}}}\log\left(1-x_i\right)\\&+\frac{3}{2}k_{\mathrm{B}}\sum_{i\in I_{\mathrm{M}}\cup I_{\mathrm{H}}}\log\alpha_{0i}-\frac{3}{2}k_{\mathrm{B}}\sum_{i\in I_{\mathrm{M}}\cup I_{\mathrm{H}}}\left(1+\log\pi\right)\end{aligned} \tag{9}$$

where $V\{\{\boldsymbol{q}\},\{n\}\}$ is the interatomic potential.

2.3 Local equilibrium relations

The optimal values of $\{\bar{\boldsymbol{q}}\}$, $\{x\}$, and $\{\alpha_0\}$ can be obtained from the minimum grand-canonical free-entropy principle of Eq. (9). Adopting Euler-Lagrange equations, the optimality condition for meanfield atomic fraction $\{x\}$ can give the relation for nondimensional chemical potential

$$\gamma_i = \log\frac{x_i}{1-x_i} + \frac{1}{k_B T}\frac{\partial\langle V\rangle_0}{\partial x_i} \tag{10}$$

This form of chemical potential characterizes the instantaneous state of a system and provides driving forces for its evolution. When calculating the chemical potential of one H site, unlike the classical entropy-of-mixing relation that is only based on the fraction of its own atomic site, this novel model not only considers the effective neighboring H sites, but also includes the deformation of the lattice structure.

2.4 Discrete kinetic laws

This thermodynamics model just described is then coupled with discrete kinetic laws of Onsager type which governs mass transport. Actually, this defines the evolution of H fraction, whose characteristic computational time step is much greater than femtosecond. In this work, we consider linear models of the form

$$\dot{x}_i = -\sum_{j\in I_{\mathrm{H}},\, j\neq i} B_{ij} x_{ij} k_{\mathrm{B}} \left(\gamma_i - \gamma_j\right), \quad i \in I_{\mathrm{H}} \tag{11}$$

where $x_{ij} = \left(x_i + x_j\right)/2$ and B_{ij} is the bondwise diffusivity coefficient between the i and j atomic pair.

To sum up, the computational framework of the deformation-diffusion coupled analysis of H diffusion in nanomaterials is shown in Fig. 1, where t_n denotes the n-th time step.

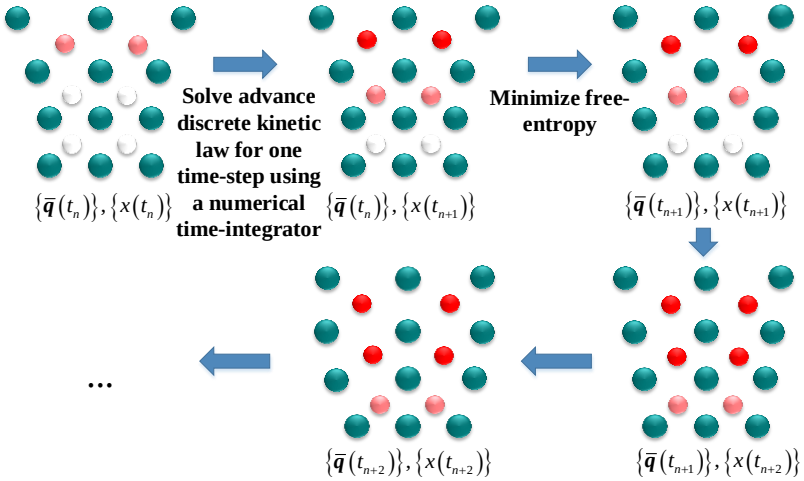


Figure 1: The computational framework of the deformation-diffusion coupled analysis.

3 APPLICATION TO PALLADIUM NANOFILMS

Since the feasibility of the described method depends critically on the formulation of computationally tractable average of the interatomic potential, we proceed to assess the fidelity of such models by direct comparison with experimental observation. To this end, we specifically investigate the palladium-hydrogen (Pd-H) system. It has been known that the Pd-H system

exists in two different phases: the α-phase at a low H concentration (up to $PdH_{0.015}$), and the β-phase at a high H concentration ($PdH_{0.607}$ and above). In the process of the α/β-phase transition, there is a lattice expansion leading to about 10.4% increase in volume [12]. It is noted that the Pd sub-lattice remains face-centered cubic (FCC) in both α and β phases, while the H atoms occupy interstitial octahedral sites, which themselves also define a second FCC sub-lattice.

3.1 Computational model

We focus on the EAM potential for Pd-H system proposed in Ref. [13]. In order to simplify calculations, we use the quadratic polynomial to approximate the EAM potential. By using Jensen's inequality, the meanfield average of the simplified EAM is of the form

$$\begin{aligned}\langle V\rangle_0 = A+\sum_{i\in I_{\mathrm{H}}} B_i x_i + \sum_{i,j\in I_{\mathrm{Pd}},i\neq j} C_{ij}\langle r_{ij}\rangle_0 + \sum_{i,j\in I_{\mathrm{H}},i\neq j} D_{ij}x_i x_j + \sum_{i\in I_{\mathrm{H}},j\in I_{\mathrm{Pd}}} E_{ij}x_i\langle r_{ij}\rangle_0 \\ + \sum_{i,j\in I_{\mathrm{Pd}},i\neq j} F_{ij}\langle r_{ij}\rangle_0^2 + \sum_{i,j\in I_{\mathrm{H}},i\neq j} G_{ij}x_i x_j\langle r_{ij}\rangle_0 + \sum_{i\in I_{\mathrm{H}},j\in I_{\mathrm{Pd}}} H_{ij}x_i\langle r_{ij}\rangle_0^2 + \sum_{i,j\in I_{\mathrm{H}},i\neq j} I_{ij}x_i x_j\langle r_{ij}\rangle_0^2\end{aligned} \tag{12}$$

Inserting the parameterization and calibration provided in Ref. [13], for T = 300 K, we have $A = -12.9356$ eV and $B_i = 3.2448$ eV. The other coefficients in Eq. (12) are listed in Table 1, where $N_i^{(k)}$ denotes the k-th shell of neighbors of site i.

	C_{ij} (eV/Å)	D_{ij} (eV)	E_{ij} (eV/Å)	F_{ij} (eV/Å^2)	G_{ij} (eV/Å)	H_{ij} (eV/Å^2)	I_{ij} (eV/Å^2)
$j \in N_i^{(1)}$	−1.1529	0.5956	−0.1496	0.2016	−0.3971	0.0574	0.0658
$j \in N_i^{(2)}$	0.2684	−0.0459	0.1940	−0.0225	0.0203	−0.0296	−0.0022
$j \in N_i^{(3)}$	0.3642	−0.0500	−0.1307	−0.0315	0.0207	0.0120	−0.0021

Table 1: Coefficients in the simplified EAM potential.

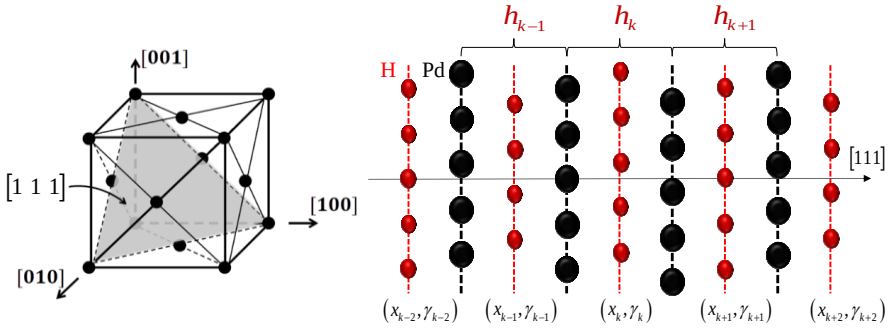


Figure 2: 1D diffusion and 3D deformation problem.

Specifically, we investigate the H diffusion problem in Pd nanofilms. Since the size of the surface area is much greater than the thickness, the H diffusion process is essentially one-dimensional in the thickness direction, that is, the [111] direction as shown in Fig. 2. But the lattice deformation is three-dimensional. As a result, the H atomic fractions $\{x\}$ and the non-dimensionalized chemical potential $\{\gamma\}$ can be constant on each [111] H plane of the FCC sub-lattice. Let N be the number of [111] planes in a specific film. We denote through x_k and γ_k, k = 1, 2, …, N, the H atomic fraction and nondimensional chemical potential on the k-th [111] plane. At the same time, it is assumed that the average instantaneous position of each atomic site can be fully determined by the average Pd plane distances $\{h\}$. That is, the average distance between any atoms is the function of average Pd plane distances along the diffusion direction. As the result, the local equilibrium relation about deformation is equal to minimizing meanfield

interatomic potential with respect to average Pd plane distances. The nominal lattice constant at the *k*-th plane is defined as $a_k = \sqrt{3}h_k$ for Pd [1 1 1] nanofilms in this work.

3.2 Model validation in equilibrium properties

The equilibrium relation of nondimensional chemical potential in Eq. (10) is plotted in Fig. 3 for a uniform H fraction, compared with the classical relation $\gamma = \log(x)$ derived from the entropy of mixing. Obviously, the proposed model exhibits an up-down-up equilibrium relation characteristic of phase transitions. In detail, the two local extreme points can be regarded as the α and β phases of the Pd-H system, respectively. The ability of the proposed model to predict the phase transition of Pd-H in the process of H diffusion will be further investigated.

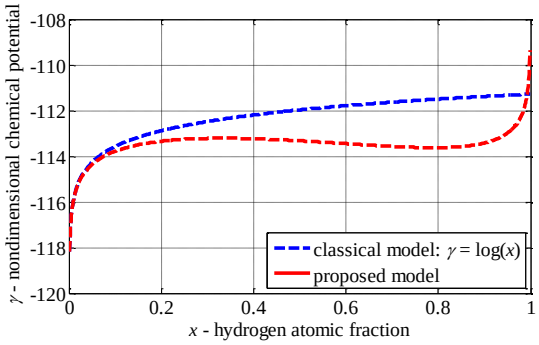


Figure 3: Comparison of equilibrium chemical potential between the proposed model and the classical relation $\gamma = \log(x)$.

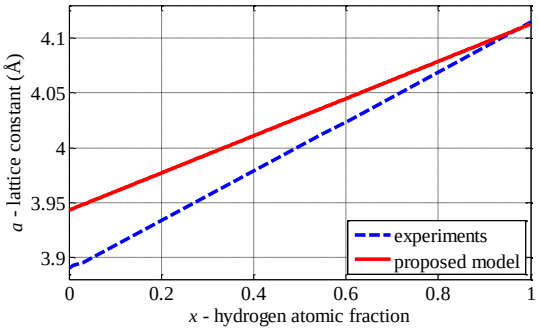


Figure 4: Comparison of equilibrium lattice constant between the proposed model and the experiments.

The equilibrium relation of lattice constant in the proposed model is compared with the experimental one from Ref. [14] in Fig. 4. The experimental points are connected by dashed lines in order to guide eye. It can be found that the trend of an increasing equilibrium lattice constant with the increasing of H fraction is predicted clearly by our proposed model. The difference between simulation and experiment results is a little greater during the low concentration and decreases gradually with the increase of H fraction, nevertheless.

3.3 Specific discrete kinetic equation

In order to describe the time-dependent H diffusion in Pd nanofilms, the general discrete kinetic law should be specified in one-dimensional direction. Restricting mass transport to the nearest shell of neighbors and using Taylor expansion, we express x_j and γ_j about x_i and γ_i. Then considering FCC Pd lattice crystal and [1 1 1] H diffusion direction, the discrete kinetic equation is of the form

$$\dot{x}_k = \frac{3}{2} k_{\mathrm{B}} \left[\left(x_k + x_{k-1} \right) \left(\gamma_{k-1} - \gamma_k \right) B_{k,k-1} + \left(x_k + x_{k+1} \right) \left(\gamma_{k+1} - \gamma_k \right) B_{k,k+1} \right] \tag{13}$$

for k = 2, 3, …, N−1. $B_{k,k-1}$ denotes the bondwise diffusivity coefficient between k-th and (k−1)-th H planes, which can be computed through solving the matrix equation by comparison with Fick's second law

$$\begin{bmatrix} 3\left(h_{k-2}+2h_{k-1}+h_k\right)^2 & 18\left(h_{k-1}+2h_k+h_{k+1}\right)^2 & 3\left(h_k+2h_{k+1}+h_{k+2}\right)^2 \\ 3\left(h_{k-1}+h_k\right)^2 & 0 & 3\left(h_k+h_{k+1}\right)^2 \\ -\left(h_{k-1}+h_k\right) & 0 & h_k+h_{k+1} \end{bmatrix} \begin{bmatrix} B_{k,k-1} \\ B_{k,k} \\ B_{k,k+1} \end{bmatrix} = \frac{8D_{\mathrm{H}}}{k_{\mathrm{B}}} \begin{bmatrix} 16 \\ 1 \\ 0 \end{bmatrix} \tag{14}$$

where D_{H} is the isotropic macroscopic diffusivity constant for H in Pd. The equilibrium relation of Eq. (14) is plotted in Fig. 5. It can be found that the decreasing trend of bondwise diffusivity coefficient with increasing plane distance is predicted. This trend is intense during mall plane distances and then becomes moderate gradually.

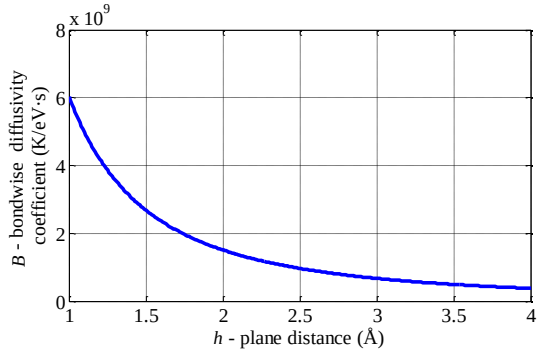


Figure 5: Equilibrium relation of bondwise diffusivity coefficient.

3.4 Model validation for hydrogen desorption

We simulate the nanofilms of thickness $L = 460$ Å and 1350 Å. Eq. (13) is calculated by the following initial and boundary conditions based on the experimental set-up

$$\begin{cases} x_k(0) = x_0, & k = 2,3,\ldots,N \\ x_1(t) = x^*, & 0 \le t \le t_{\max} \\ x_{N-1}(t) = x_N(t), & 0 \le t \le t_{\max} \end{cases} \tag{15}$$

where x_0 denotes the initial uniform H atomic fraction in the film interior, and x^* denotes the fixed H atomic fraction at the film surface in contact with the electrolyte solution. No-flux boundary condition is imposed at the other side of the nanofilms, since H diffusion between the Pd film and the gold-coated nickel substrate is negligible. x^* can be determined according to Ref. [9], and x_0 is set to 0.009 in the computation based on the measurement [15] in both cases. All the case-dependent coefficients used in the simulations are reported in Table 2.

	L (Å)	N (Å)	x^*	D_{H}(Å²/s)
Case I	460	204	0.0019	2.1×10^5
Case II	1350	595	0.0047	1.56×10^6

Table 2: Case-dependent coefficients used in H desorption simulations.

Eq. (13) is discretized in time by the mid-point rule which has second-order accuracy. The computational time steps are set to $\Delta t = 5.0 \times 10^{-6}$ s and 1.0×10^{-6} s in Case I and II, respectively. The maximum simulation time for both cases is $t_{\max} = 1.0$ s.

The spatial variation of H atomic fraction on the surface of nanofilms is converted into electric current using the relation in Ref. [16]. The results are plotted in Fig. 6 in comparison with the experimental data reported in Ref. [17] at T = 300 K. It is obvious that in both cases, the predicted electric current agrees well with the experimental measurements with regard to both the initial transient and the long-term behavior.

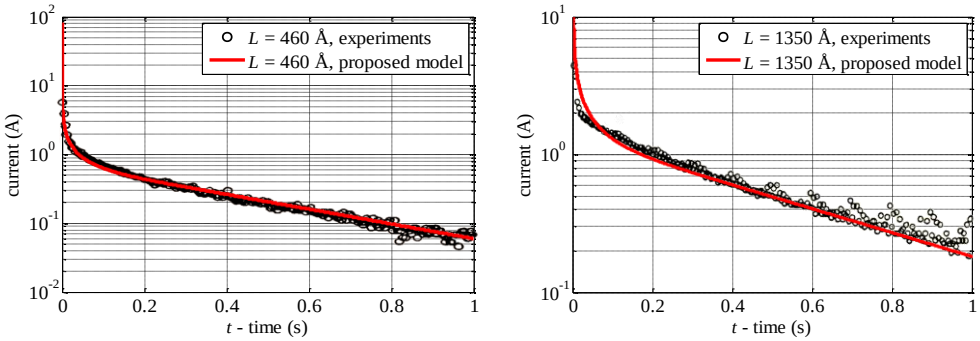


Figure 6: Comparison of experimental and predicted results for current on the surface. (a) 460 Å Pd nanofilms. (b) 1350 Å Pd nanofilms.

3.5 Prediction for phase transition

As discussed in Fig. 3, the equilibrium condition of Eq. (10) is able to capture the phase characteristic of one Pd-H system. Here, we further investigate this ability by taking the Pd-H system into its β-phase, and highlight the capability of the proposed computational model as to capturing complicated phase transition in a very slow diffusion process which has time scale of seconds. We simulate H absorption in the Pd nanofilms with the thickness $L = 460$ Å characterized by a high H fraction on the surface $x^* = 0.99$. Using the deformation-diffusion coupled model proposed in this paper, the H atomic fractions and lattice constants at six time points are plotted in Fig. 7, respectively. Also, the similar figures through the classical mixing-entropy relation $\gamma = \log(x)$ are shown in Fig. 8 for comparison. Clearly, the H diffusion process in Pd lattice structure is characterized by the propagation of an α/β-phase boundary: the α-phase about $PdH_{0.165}$ and $a = 3.968$ Å, and the β-phase about $PdH_{0.898}$ and $a = 4.098$ Å. This phase propagation cannot be observed by using the classical relation $\gamma = \log(x)$. Moreover, the lattice expansion in the α/β-phase transition is about 3.27%, which gives rise to about 9.83% increase in volume. This value is close to that observed in previous experiments. Nevertheless, the H atomic fractions at α/β-phase are greater than those obtained from experiments, which is probably due to the EAM potential we adopt.

Fig. 9 shows the deformation-diffusion coupled process of H absorption at several time steps in the part of the deformed Pd lattice structure, where the big spheres denote Pd sites and the small ones are H sites. It can be seen that the lattice structure will expand in three dimensions as H atoms diffuse in [1 1 1] direction. Obviously there exists a phase boundary between the domains with high and low H atomic fractions. As a result, the capability of the proposed model in predicting the phase transition in a long-term process is remarkable.

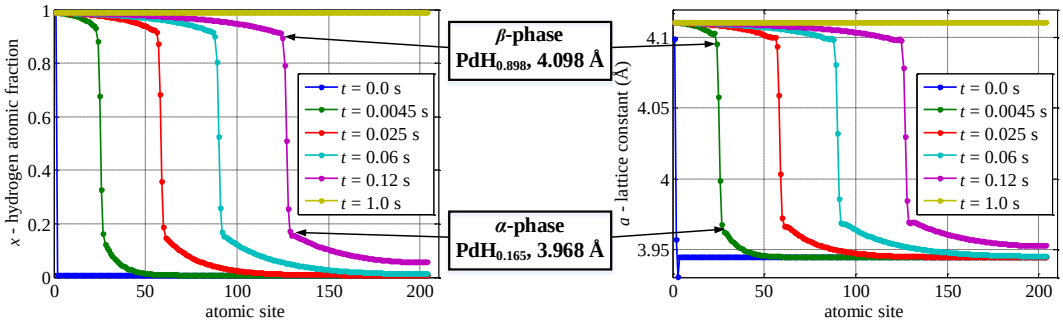


Figure 7: H absorption in 460 Å Pd nanofilms by the proposed model. (a) H atomic fraction. (b) Lattice constant.

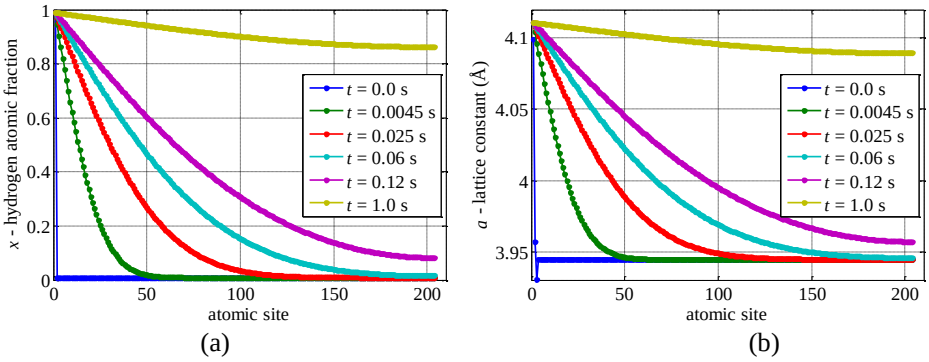


Figure 8: H absorption in 460 Å Pd nanofilms by the classical relation $\gamma = \log(x)$. (a) H atomic fraction. (b) Lattice constant.

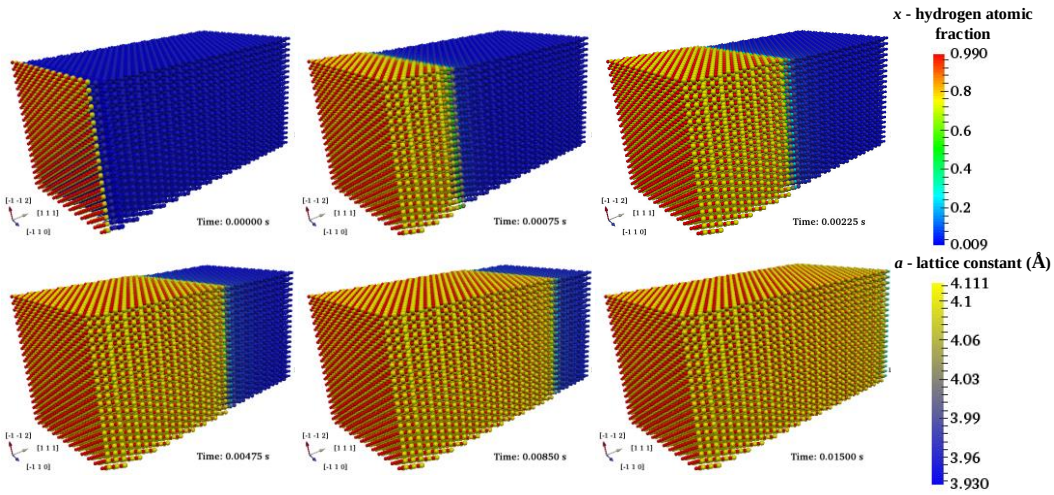


Figure 9: Deformation-diffusion coupled process of H absorption in Pd nanofilms.

4 SUMMARY AND CONCLUDING REMARKS

We have applied the theory of non-equilibrium statistical thermodynamics to the study of hydrogen absorption and desorption in metal-based nanomaterials, while maintaining a strictly atomistic description of the metal material. We have proposed a new form of trial Hamiltonians and combined it with the specific discrete kinetic law for hydrogen diffusion in palladium nano-films. The coupled thermodynamic and kinetic models are solved numerically in a partitioned

procedure using explicit time-integrators. The proposed model has been validated in both equilibrium and non-equilibrium cases, and the computational results are compared with those by classical chemical potential relation derived from entropy-of-mixing. Our model can have a good agreement with the experimental measurements in hydrogen desorption process and predict the phase boundary transition of hydrogen absorption with the well-known volume expansion. The results demonstrate that the presented computational framework is capable of capturing phase-change and phase-dependent diffusion of hydrogen in a deformed metal structure as part of a long-term analysis.

REFERENCES

[1] W. Li, C. Li, H. Ma, J. Chen, Magnesium nanowires: enhanced kinetics for hydrogen absorption and desorption. *Journal of the American Chemical Society*, **129**, 6710-6711, 2007.

[2] D. Delafosse, T. Magnin, Hydrogen induced plasticity in stress corrosion cracking of engineering systems. *Engineering Fracture Mechanics*, **68**, 693-729, 2001.

[3] K.A. Williams, P.C. Eklund, Monte Carlo simulations of H-2 physisorption in finite-diameter carbon nanotube ropes. *Chemical Physics Letters*, **320**, 352-358, 2000.

[4] V.J. Surya, K. Iyakutti, H. Mizuseki, Y. Kawazoe, First principles study on desorption of chemisorbed hydrogen atoms from single-walled carbon nanotubes under external electric field. *International Journal of Hydrogen Energy*, **36**, 13645-13656, 2011.

[5] R.A. Miron, K.A. Fichthorn, Accelerated molecular dynamics with the bond-boost method. *The Journal of Chemical Physics*, **119**, 6210-6216, 2003.

[6] I. Brihuega, O. Custance, J.M. Gómez-Rodríguez, Surface diffusion of single vacancies on Ge (111) - c (2×8) studied by variable temperature scanning tunneling microscopy. *Physical Review B*, **70**, 165410, 2004.

[7] J. Li, S. Sarkar, W.T. Cox, T.J. Lenosky, E. Bitzek, Y. Wang, Diffusive molecular dynamics and its application to nanoindentation and sintering. *Physical Review B*, **84**, 054103, 2011.

[8] G. Venturini, K.G. Wang, I. Romero, M.P. Ariza, M. Ortiz, Long-term simulation of heat and mass transport in alloys at finite temperature. *Journal of the Mechanics and Physics of Solids*, **73**, 242-268, 2014.

[9] K.G. Wang, M. Ortiz, M.P. Ariza, Long-term atomistic simulation of hydrogen diffusion in metals. *International Journal of Hydrogen Energy*, **40**, 5353-5358, 2015.

[10] E.T. Jaynes, Information theory and statistical mechanics I. *Physical Review*, **106**, 620-630, 1957.

[11] E.T. Jaynes, Information theory and statistical mechanics II. *Physical Review*, **108**, 171-190, 1957.

[12] R.J. Wolf, M.W. Lee, R.C. Davis, P.J. Fay, J.R. Ray, Pressure-composition isotherms for palladium hydride. *Physical Review B*, **48**, 12415, 1993.

[13] X.W. Zhou, J.A. Zimmerman, B.M. Wong, J.J. Hoyt, An embedded-atom method interatomic potential for Pd-H alloys. *Journal of Materials Research*, **23**, 704-718, 2008.

[14] Y. Sakamoto, K. Yuwasa, K. Hirayama, X-ray investigation of the absorption of hydrogen by several palladium and nickel solid solution alloys. *Journal of the Less Common Metals*, **88**, 115-124, 1982.

[15] H. Hagi, Diffusion coefficient of hydrogen in palladium films prepared by RF sputtering. *Materials Transactions, JIM*, **31**, 954-958, 1990.

[16] N. Boes, H. Züchner, Electrochemical methods for studying diffusion, permeation and solubility of hydrogen in metals. *Journal of the Less Common Metals*, **49**, 223-240, 1976.

[17] Y. Li, Y.T. Cheng, Hydrogen diffusion and solubility in palladium thin films. *International Journal of Hydrogen Energy*, **21**, 281-291, 1996.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MESH OBJECTIVE DAMAGE MODELING OF DUCTILE FRACTURE AT VISCO-PLASTIC CONTINUUM RESPONSE

S. Razanica[1], R. Larsson[1], and B. Lennart Josefson[2]

[1]Division of Material and Computational Mechanics
Department of Applied Mechanics
Chalmers University of Technology, SE-412 96 Göteborg, Sweden
e-mail: {razanica,ragnar.larsson}@chalmers.se

[2] Department of Shipping and Marine Technology
Chalmers University of Technology, SE-412 96 Göteborg, Sweden
e-mail: lennart.josefson@chalmers.se

Keywords: Ductile damage, fracture modeling, mesh objective damage models , rate dependency, Johnson-Cook model.

Abstract. *Local- continuum damage models, such as the JC-dynamic failure criterion are often combined with the constitutive JC-model to represent the material behavior during machining. The major drawback is that the failure criterion exhibit a pathological mesh dependence. In literature it has been argued that a viscous regularization of the continuum material model coupled to damage via visco-plasticity may remove the mesh dependence. We present results, based on the objective enhanced element removal model derived from previous contribution, form a mesh dependence study using a extended formulation of the constitutive JC- model covering visco-plasticity. The results show that, excluding the mesh objective enhancement the pathological mesh dependence still exist if the visco-plastic formulation is used. If the mesh objective enhancement is used a relatively good convergence is obtained for a set of mesh sizes for the damage model.*

1 INTRODUCTION

In order to model ductile fracture combining the constitutive JC-model with the JC- dynamic failure criteria as in [1] is a possible approach. The JC-model describes the damage evolution in a gauss-point of an element as a function of the failure strain in relation to the accumulated plastic strain. When the JC- dynamic failure criterion is fulfilled, the stress response in the gauss-point is instantaneously relaxed and the element is removed from the simulation. As for the failure strain, it is composed by three dependencies, the stress triaxiality, plastic strain rate and temperature similar to the JC- model. This approach refers to the standard JC-dynamic failure criterion. Despite its simplicity it is not possible to disregard that the JC-model, due to its softening effects, might result in a loss of solution uniqueness. This results in in so called pathological mesh dependence which is a problem for local damage models.

In previous contribution [1] a remedy was proposed for the observed inherent mesh dependence in ductile fracture models. To remove the dependence, a set of mesh objective damage models were developed based on a local-continuum damage formulation combined with concepts from a scalar damage phase-field [3], [4]. Independent of damage model, the JC- dynamic failure criterion was used to initialize a "fracture state" which signal for damage initialization and damage evolution. One of the objective damage models relates to the traditional "element removal" which is nothing else than the coupling between the JC- models, described by an instantaneous damage evolution.

In [5], [6] it is argued that a viscous regularization of the continuum material model coupled to damage via visco-plasticity may remove the mesh dependence. In this contribution, the rate dependent response was incorporated in the JC- model extending the mesh objective damage enhancement. The model was implemented in ABAQUS/Explicit as user subroutine based on the hypo-elastic inelastic framework which is computationally efficient [7]. In order to investigate the assertion regarding viscous regularization postulated, a mesh dependency study was conducted for highly refined mesh sizes of a 2D plane strain plate. The results show that, excluding the mesh objective enhancement the pathological mesh dependence still exist if the visco-plastic formulation is used. If the mesh objective enhancement is used a relatively good convergence is obtained for a set of mesh sizes.

2 CONSTITUTIVE MODELING BASED ON VISCO-PLASTIC CONTINUUM RESPONSE

In this section mesh objective element removal model formulated in [1] with the potential to describe ductile fracture in materials, for rate independent response, is evaluated in a visco-plastic/rate-dependent context. As for the considered model, the damage evolution takes place in one single step when the the damage criterion is fulfilled, representing a instantaneous stress relaxation. In order to represent the effective stress response the Johnson-Cook (JC) constitutive material model, extended to include visco-plastic response, is considered.

2.1 Visco-plastic continuum response and ductile damage models derived from the Johnson Cook failure criterionl

In the present context it is of significant interest to enhance the effective material in terms of a "scalar damage" measure α acting on the effective material so that

$$\psi = (1-\alpha)\hat{\psi} \text{ with } \hat{\psi} = \hat{\psi}\left[\bar{\mathbf{b}}, k\right] \tag{1}$$

where $\hat{\psi}$ is the free energy of the effective (undamaged) material, denoted by a superimposed hat. It is assumed that $\hat{\psi}$ depends on the elastic Finger deformation tensor $\bar{\mathbf{b}}$ and an isotropic hardening variable k. It turns out that the total dissipation rate $\mathcal{D}$ in the dissipation inequality can be formulated in terms of the effective dissipation rate $\hat{\mathcal{D}}$ as

$$\mathcal{D} = (1-\alpha)\hat{\mathcal{D}} + \hat{\psi}\dot{\alpha} \geq 0 \text{ with } \hat{\mathcal{D}} = \hat{\tau} : \mathbf{l}_p + \kappa\dot{k} \geq 0 \tag{2}$$

which corresponds to the constitutive state equations

$$\hat{\tau} = 2\frac{\partial\hat{\psi}}{\partial\bar{\mathbf{b}}} \cdot \bar{\mathbf{b}} \, , \, \hat{\kappa} = -\frac{\partial\hat{\psi}}{\partial k} \tag{3}$$

In (3), we introduced the effective Kirchhoff stress $\hat{\tau}$ and the hardening stress $\hat{\kappa}$ pertinent to the effective material. To ensure positive total dissipation $\mathcal{D} \geq 0$ it suffices to consider $\hat{\mathcal{D}} \geq 0$ as in (2) and $\dot{\alpha} \geq 0$ (since $\hat{\psi} > 0$ always). This corresponds to the constitutive state equations

$$\tau = (1-\alpha)\hat{\tau}, \kappa = (1-\alpha)\hat{\kappa}, A = -\frac{\partial\psi}{\partial\alpha} = \hat{\psi} \tag{4}$$

where A is the damage driving force. Upon assuming neo-Hookean isotropic elasticity, we assume for the stored elastic behavior

$$\begin{aligned} A = \hat{\psi} &= \hat{\psi}^{\text{iso}} + \hat{\psi}^{\text{vol}} + \hat{\psi}^{\text{hard}}[k] \text{ with} \\ \hat{\psi}^{\text{iso}} = \frac{1}{2}GJ^{-\frac{2}{3}}\left(1 : \hat{\mathbf{b}} - 3\right) &\, , \, \hat{\psi}^{\text{vol}} = \frac{1}{2}K(J-1)^2 \, , \, \hat{\psi}^{\text{hard}} = \frac{B}{1+n}(-k)^{1+n} \end{aligned} \tag{5}$$

where G is the shear modulus, K is the bulk modulus, B is the isotropic hardening parameter and n is the hardening exponent.

Moreover, we consider the visco-plastic Johnson Cook model for the inelastic effective continuum response. We thereby introduce a static yield function $\phi = \phi\left[\hat{\tau}_e, \hat{\kappa}\right]$ in terms of the the effective von Mises stress $\hat{\tau}_e$, and the evolution of the internal dissipative variables $\left\{\mathbf{l}_p, \dot{k}\right\}$

$$\mathbf{l}_p = \lambda\frac{\partial\phi}{\partial\hat{\tau}} = \frac{3}{2}\frac{\hat{\tau}_{\text{dev}}}{\hat{\tau}_e} \, , \, \dot{k} = \lambda\frac{\partial\phi}{\partial\hat{\kappa}} \tag{6}$$

where $\lambda \geq 0$ is the plastic multiplier determined by the *Johnson-Cook* (JC) overstress

$$\lambda = \dot{\epsilon}_0 \exp\left[\frac{\phi}{C\left(A+\hat{\kappa}\right)}\right] \text{ for } \lambda > \dot{\epsilon}_0 \tag{7}$$

where $\hat{\kappa} = B(-k)^n$ is the isotropic micro-hardening stress and ϕ the yield function explicitly formulated as $\phi = \hat{\tau}_e - (A+\hat{\kappa})$. We note that A, C and $\dot{\epsilon}_0$ are material parameters representing initial yield and rate sensitivity, respectively. The response becomes visco-plastic (or rate dependent) whenever $\lambda \geq \dot{\epsilon}_0$. To handle the rate independent case $\lambda < \dot{\epsilon}_0$, the plastic multiplier is controlled by the Kuhn-Tucker (KT) loading conditions

$$\phi \leq 0, \quad \lambda \geq 0, \quad \lambda\phi = 0 \text{ for } \lambda \leq \dot{\epsilon}_0 \tag{8}$$

Following the developments in [1], we outline here the mesh objective element removal model associated with the JC-material and failure models representing ductile fracture. To this

end, we consider the JC- dynamic failure criterion, where it is postulated that ductile fracture occurs when the equivalent plastic strain approaches the fracture strain ϵ_f^p defined as

$$\epsilon_f^p = (d_1 + \exp[-d_2 r]) \left(d_3 + \exp\left[d_4 \frac{\lambda}{\dot{\epsilon}_0} \right] \right) \text{ with } r = -\frac{\hat{p}}{\hat{\tau}_e} \text{ and } \hat{p} = -\frac{\mathbf{1} : \hat{\tau}}{3} \tag{9}$$

Hence, a fracture state is obtained when $k + \epsilon_f^p = 0$ at $t = t_c$, where t_c is time when local failure occurs. Please note that this critical time depends on the rate of loading $\lambda/\dot{\epsilon}_0$, the triaxialty r and initial failure strain. These inter-dependences are described by (9) along with the failure parameters d_1–d_4.

The total damage force is introduced as

$$A_T = \hat{\mathcal{D}}_T^f[t] + A \, , \hat{\mathcal{D}}_T^f[t] = \int_{t_c}^{t} \hat{\mathcal{D}} \, dt \tag{10}$$

We consider next the *damage-element removal model* based on the instantaneous damage evolution rule, as formulated by the Dirac-delta function $\delta_S[\bullet]$ for the damage evolution, i.e.

$$\dot{\alpha} = \mu \delta_S \left[t - t_c \right] \tag{11}$$

where $\dot{\alpha}$ is controlled by

$$\phi_\alpha = A_T - \frac{L_c}{L_e} A_c \leq 0 \, , \quad \mu \geq 0 \, , \quad \phi_\alpha \mu = 0 \tag{12}$$

In this case, the critical time is defined by when $t = t_c$ has been reached in (9) and $A_T - A_c = 0$.

We thus conclude that a "fracture state", $t = t_c$, is arrived at when the JC- dynamic failure criterion (9) is met. For the damage model, the stored free energy A_c in (12) to be released is evaluated at the Gauss-point level for each element size. For finer meshes with $L_e < L_c$ the continuous dissipation $\hat{\mathcal{D}}_T^f$ is integrated so that A_T can be computed from (10). For the considered damage model, the stress state is "removed" when the condition (12) is approached.

3 NUMERICAL EXAMPLES

In [5], [6] it is argued that a viscous regularization of the continuum material model coupled to damage via visco-plasticity may remove the mesh dependence. In order to investigate the assertion the rate dependent/visco-plastic material model described in previous sections will be considered together with the proposed mesh objective element removal damage model in a mesh dependency investigation for highly refined meshes.

3.1 Finite element modeling

As a point of departure for the mesh dependency investigation the same 2-D shear loaded plane strain plate as in [1] and [7] is considered with the material parameters for the JC material and failure models in Tables 1. The 50x50 mm square shaped plate is divided into three parts. On the right side of the top edge of the plate a prescribed vertical displacement is applied on the nodes. The right boundary of the plate is constrained in horizontal direction allowing vertical motion. Therefore a displacement of 25 mm is applied on the upper right edge for the displacement rate 500 mm/s. The left section is fully constrained in both directions as shown in Figure 1. Since the mid section is subjected to severe shear deformation it suffices to use a fine FE-representation here in order to model the damage with a reasonable accuracy, while using

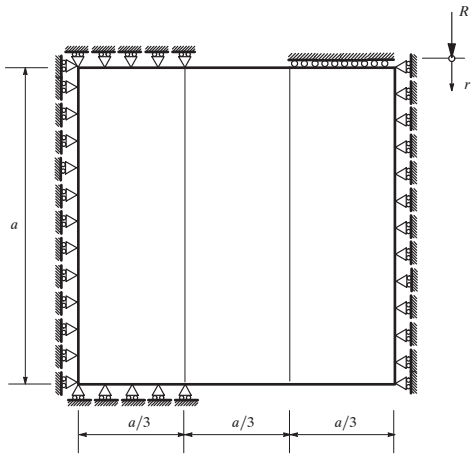


Figure 1: Considered geometry used for the analyses with corresponding boundary conditions

coarser elements for the remaining parts of the geometry. In this way the computational cost is reduced. The investigation is performed for the element removal model using the hypo-elastic inelastic framework implemented in the commercial software ABAQUS/Explicit using the user subroutine VUMAT.

Table 1: Johnson-Cook constitutive material and failure parameters

A [MPa]	B [MPa]	C	n	$\dot{\epsilon}_0$	d_1	d_2	d_3	d_4
550	500	0.0804	0.4	0.001	0.25994	0.61368	2.5569	-0.027652

To perform the mesh dependency investigation a set of six different FE discretizations of the plate are created. The 4-node plane strain bilinear elements described in previous section are used with different element sizes: 1, 0.75, 0.5, 0.25, 0.125 and 0.1 mm. Therefore, a wise discretization of the plate, regarding the finer mesh sizes, is beneficial from the computational point of view. Due to the severe shear deformation that occur in the mid-section of the geometry, we choose to enrich this region with the specific element sizes described. Doing so, highly resolved damage modes are achieved. Noteworthy, only a structured mesh is used in the mid-section while a free distribution of elements is used for the remaining geometry.

4 RESULTS

In this section, the performed simulations based on the suggested examples of different FE discretizations are analyzed with respect to mesh dependency. The simulations are conducted for the enhanced objective element removal model proposed in [1] with the extension of the JC - material model to include rate dependent behavior. The results are presented in form of force-displacement curves, where the magnitude of the force is computed at the edge where the displacement boundary conditions are applied.

4.1 Element removal model - Excluding mesh objective enhancement

If we consider the response without any mesh objective enhancement, using standard JC dynamic failure model, it could be argue that the pathological mesh dependency should decrease with the introduction of vico-plasticity in the material model. However, our simulations show that there is no tendency of minimization of the mesh dependency for the deformation rate

chosen as illustrated in Figure 2. But as the mesh size is refined, the response becomes more and more brittle, eventually a convergence is obtained. It is also clear that as the FE discretization of the plate is successively refined, the kinematics allow for other possible damage modes which may be observed in Figure 2.

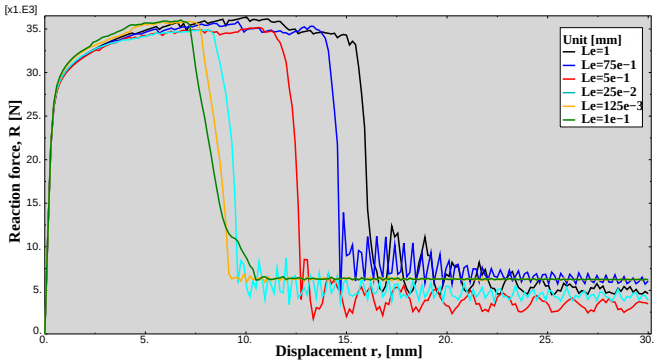


Figure 2: Force-displacement curves for different element sizes L_e and deformation rate v = 500 mm/s excluding objective enhancement in the element removal mode, representing standard JC dynamic failure model.

4.2 Element removal model - Including mesh objective enhancement

The results for the element removal model including the objective enhancement are presented for the chosen displacement rate where the reference element size of $L_c = 0.5$ mm was chosen. When the mesh objective enhancement is introduced a relatively good convergence is achieved for element sizes finer than the reference size. No indication of mesh size convergence is notable for $L_e > L_c$ as observed from Figure 3, which is similar to the rate independent case in [1].

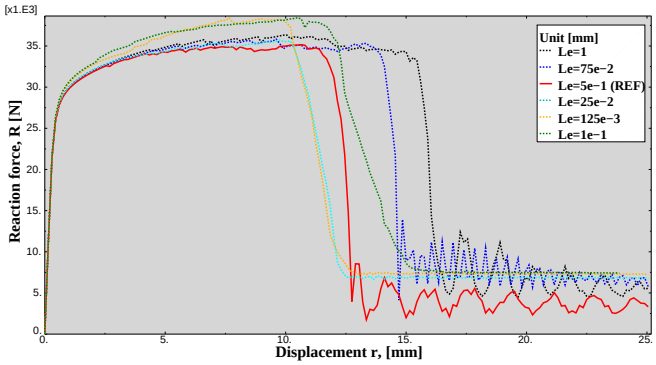


Figure 3: Force-displacement curves for different element sizes L_e and deformation rate v = 500 mm/s including objective enhancement in the element removal model.

5 CONCLUDING REMARKS

As a point of departure in the current contribution, the derived mesh objective element removal model, based on the local-continuum damage formulation combined with scalar damage phase-field concepts, in [1] is investigated. The constitutive JC- model was expanded to incorporate visco-plastic response, taking the rate dependence into consideration. As it was argued in literature that a that a viscous regularization of the continuum material model coupled to damage via visco-plasticity may remove the mesh dependence an investigation was conducted. The mesh objective damage models were implemented in ABAQUS/Explicit as user subroutines. In

order to conduct the mesh dependency investigation the same 2D plane strain plate as in previous contribution was used with different structured FE discretizations. Our results show that the the pathological mesh dependence is till present when the mesh objective enhancement is excluded although visco-plasticity is incorporated in the constitutive model. Furthermore, when the mesh objective enhancement is included a convergence in the force-displacement curves is evident for a set of mesh sizes, independent of damage model. Nevertheless, the models do predict reliable response regarding the force-displacement curves and the pathological mesh dependence is removed for a set of mesh sizes.

REFERENCES

[1] R. Larsson S. Razanica, B. Lennart Josefson. Mesh objective continuum damage models for ductile fracture *International Journal for Numerical Methods in Engineering*, (2015) doi: 10.1002/nme.5152..

[2] Goran Ljustina,Martin Fagerstrm, and Ragnar Larsson. Rate Sensitive Continuum Damage Models and Mesh Dependence in Finite Element Analyses. *The Scientific World Journal*, (2014) http://dx.doi.org/10.1155/2014/260571.

[3] C. Miehe, F. Welschinger and M. Hofacker, "Thermodynamically consistent phase-field models of fracture: Variational principles and multi-field FE implementations, *Int. J. Numer. Meth. Engng.*, **83**:1273-1311 (2010).

[4] C. Miehe, M. Hofacker, L. Schanzel, F. Aldakheel, Phase field modeling of fracture in multi-physics problems. Part II. Coupled brittle-to-ductile failure criteria and crack propagation in thermo-elasticplastic solids, *Comput. Methods Appl. Mech. Engrg.*, http://dx.doi.org/10.1016/j.cma.2014.11.017 (2014).

[5] M. S. Niazi, H. H. Wisselink, T. Viscoplastic regularization of local damage models: revisited. *Comput Mech*, **51**:203-216 (2013).

[6] V. Dias da Silva. A simple model for viscous regularization of elasto-plastic constitutive laws with softening. *Commun. Numer. Meth. Eng.*, **29**:547-568 (2004).

[7] Ljustina G, Fagerstrom M, Larsson R. Hypo and hyperinelasticity applied to modeling of compact graphite iron machining simulations. *European Journal of Mechanics - A/Solids*, **37**:57-68 (2012).

[8] Abaqus 6.12 - Abaqus/CAE User's manual. *Dassault Systemes*, (2012).

[9] M. Barge, H. Hamidi, J. Rech, J.-M. Bergheau.Numerical modelling of orthogonal cutting: influence of numerical parameters. *Journal of Materials Processing Technology*, **164-165**:1148-1153 (2005).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EQUIVALENT SDOF MODEL FOR ESTIMATING BLAST-INDUCED DYNAMIC REACTIONS OF EQUILATERAL TRIANGULAR HARDENED WALL ELEMENTS

Sebastian Mendes[1], **Liling Cao**[1], **Douglas Heinze**[1], **Elisabeth Malsch**[1]

[1] Thornton Tomasetti
40 Wall Street, 19th Floor, New York, NY 10005-1304
e-mail: {SMendes, LCao, DHeinze, EMalsch}@ThorntonTomasetti.com

Keywords: Blast protection, blast-resistant wall, triangular plate, dynamic reaction, equivalent single degree-of-freedom system, elasto-plastic material, plate theory

Abstract. *Tall buildings oftentimes employ internal blast-resistant walls for protecting vital building areas and components from the damaging effects of high energy explosions. Architectural constraints may require hardened walls aligned with steel braced frames to be built in a shape equal to or near to that of an equilateral triangle. The analysis and design of the hardened walls and their supporting members for blast-induced dynamic loading is an imperative aspect of the design process. Parameters are derived for transforming the analysis of a simply-supported equilateral triangular hardened wall element subjected to blast loading into an equivalent single degree-of-freedom (SDOF) system. Expressions are then derived for obtaining the corresponding distributions of peak blast-induced dynamic reactions along the perimeter supports of the triangular wall element. The equivalent SDOF model and the expressions for the peak dynamic reaction distributions are subsequently validated with dynamic finite element analyses (FEA), therein verifying their applicability for rapidly analyzing and designing numerous quantities of equilateral triangular hardened wall elements and their supporting members in lieu of more rigorous and time-consuming dynamic FEA.*

1 INTRODUCTION

Tall buildings designed to resist high energy explosions oftentimes employ internal hardened walls for protecting means of egress and critical mechanical/electrical/plumbing (MEP) systems from blast infill pressures. For steel buildings with braced frame lateral systems, architectural constraints may require hardened walls to be subdivided into non-rectangular elements and supported in-plane of a building's braced frames. The magnitude and distribution of blast loads upon the braced frame as caused by blast-induced dynamic reactions of hardened walls are typically required for checking the performance and connection capacity of the individual braced frame members.

In this study a proposed equivalent SDOF model is developed for an equilateral triangular hardened wall element simply-supported in-plane of an inverted V-braced frame and uniformly loaded by blast infill pressure described by a linearly decaying pressure-time history *q(t)* (see Figure 1). Typical V-braced frames are constructed in a shape closer to that of a wide isosceles triangle. However, to simplify the analysis an equilateral triangle is chosen. The SDOF model is used to analyze the distribution of peak blast-induced dynamic reactions along the edges of triangular wall elements, and the results are validated with dynamic FEA. This verification indicates that the equivalent SDOF model can be practically incorporated into standard SDOF blast analysis spreadsheets for highly repetitive analysis and design of braced frame members supporting equilateral triangular hardened wall elements in-plane for blast.

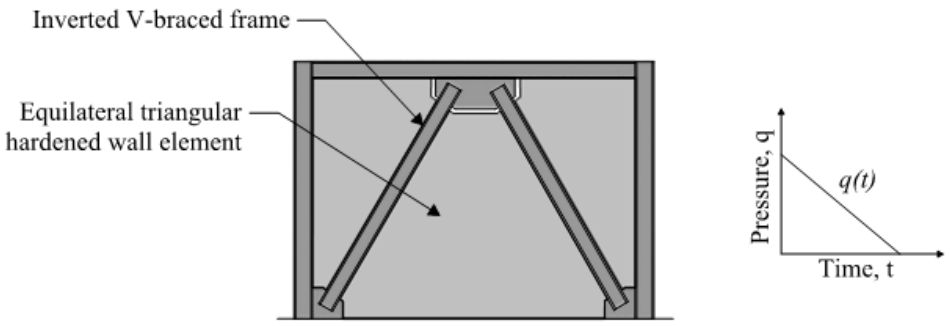


Figure 1: Equilateral triangular hardened wall element simply-supported in-plane of inverted V-braced frame and uniformly loaded by blast infill pressure.

2 EQUIVALENT SDOF MODEL FOR EQUILATERAL TRIANGULAR PLATE

2.1 SDOF Governing Equations of Motion

The equilateral triangular hardened wall element is assumed to behave as a homogeneous isotropic plate possessing elastic-perfectly plastic material properties (i.e. steel plate). The dynamic blast analysis of the real system, consisting of the equilateral triangular plate loaded by *q(t)*, is transformed into an equivalent undamped SDOF system subjected to an equivalent load-time history (see Figure 2). In the elastic phase the governing undamped equation of motion for the equivalent system is:

$$M_e \frac{d^2}{dt^2} z(t) + k_e z(t) = F_e(t) \tag{1}$$

where M_e and k_e are the equivalent mass and stiffness, respectively, in the elastic phase. Also, $z(t)$ is the real displacement and $F_e(t)$ is the equivalent load-time history in the elastic phase. In the plastic phase the equation of motion becomes:

$$M_p \frac{d^2}{dt^2} z(t) + R_m = F_p(t) \tag{2}$$

where R_m is the maximum resistance, and M_p and $F_p(t)$ are the equivalent mass and load-time history, respectively, in the plastic phase. The system converts to free vibration upon the termination of the load-time history.

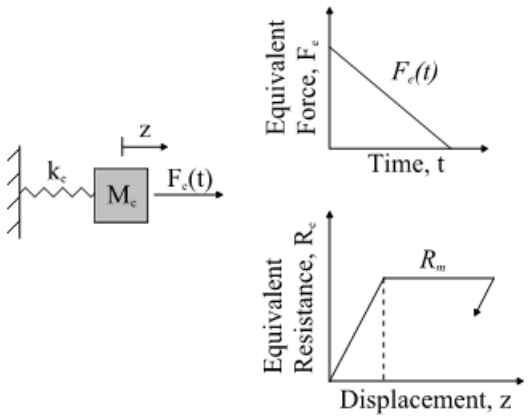


Figure 2: Equivalent SDOF system loaded by equivalent load-time history.

The real stiffness and maximum resistance of the equilateral triangular plate are required to develop the equivalent SDOF system. Also, the deflection surface, shape function, and transformation factors for the elastic and plastic phases are required to develop the equivalent SDOF system. Numerical analysis such as with the constant-velocity procedure allows for a solution of the response [1].

In the case of this study we are interested in the maximum displacement, z_{max}, of the system in the free vibration phase such that the peak dynamic reaction, Q_{max}, may be obtained from the constitutive relations:

$$Q_{\max} = \min \begin{cases} kz_{\max} \\ R_m \end{cases} \tag{3}$$

where k is the real stiffness. A third factor limiting the maximum dynamic reaction is the shear strength of the plate along its supported edges. In the case of this study the shear strength is assumed to exceed the dynamic shear demand. Using (3) the distribution of the peak dynamic reaction along the edges of the equilateral triangular plate may therein be determined using the plate deflection surface and constitutive equations. This distribution represents the peak blast-induced dynamic reaction of the equilateral triangular hardened wall element upon the supporting braced frame members.

2.2 Deflection Surface, Shape Function, and Stiffness in Elastic Phase

It is assumed that the deformed shape of the triangular plate during elastic response to blast loading conforms to the plate deflection surface obtained from a statically applied uniform surface load. Any deflection surface, *w(x,y)*, for a homogeneous plate must satisfy the governing plate equation:

$$\nabla^4 w(x,y) = \frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = \frac{q}{D} \tag{4}$$

where q is a uniform surface load and D is the plate flexural rigidity defined by $Et^3/[12(1-\nu^2)]$. The following deflection surface satisfies (4) for an equilateral triangular plate (see Figure 3a) simply-supported along the edges and uniformly loaded by q [2]:

$$w(x,y) = \frac{q}{64aD}\left[x^3 - 3y^2 x - a\left(x^2+y^2\right) + \frac{4}{27}a^3\right]\left(\frac{4}{9}a^2 - x^2 - y^2\right) \tag{5}$$

The deflection surface intrinsically includes the plate bending stiffness and applied surface load. The more general shape function, $\phi(x,y)$, describing the deflection surface independently of stiffness and load (see Figure 3b) is determined by extricating from (5) the plate bending stiffness explicit to the displacement at the center of the plate due to surface load q:

$$\phi(x,y) = \frac{243}{16a^5}\left[x^3 - 3y^2 x - a\left(x^2+y^2\right) + \frac{4}{27}a^3\right]\left(\frac{4}{9}a^2 - x^2 - y^2\right) \tag{6}$$

where the plate bending stiffness corresponding to the displacement at the center of the plate is solved to be:

$$k = \frac{324\sqrt{3}D}{a^2} \tag{7}$$

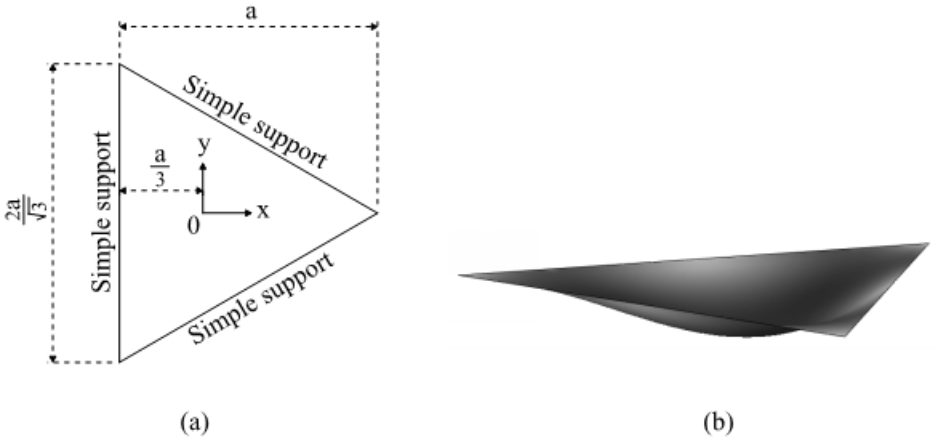


Figure 3: (a) Simply-supported equilateral triangular plate and (b) shape function in elastic phase.

2.3 Approximate Maximum Resistance

In accordance with yield line theory the maximum resistance of the equilateral triangular plate signifying the initiation of the plastic phase is defined by the formation of plastic hinges along the lines bisecting the angles of the triangle (see Figure 4). It is approximated that the formation of plastic hinges occurs when the von Mises stress, σ_v, along the bisecting lines achieves the yield stress, σ_Y. The elasto-plastic transition range leading to the true formation of plastic hinges is complex and has been ignored in this study. However, if the plate thickness is relatively thin compared to the area dimensions then the elasto-plastic range becomes exceedingly negligible [1].

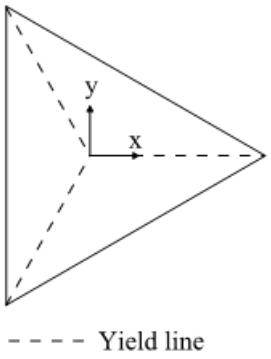


Figure 4: Yield lines on equilateral triangular plate.

An expression for the approximate maximum resistance requires a relation between the applied surface load and the von Mises stress along the yield lines. The plate constitutive equations relating bending moments to curvatures are written as follows [2]:

$$M_x(x,y) = D\left(\frac{\partial^2 w}{\partial x^2} + \nu\frac{\partial^2 w}{\partial y^2}\right) \qquad M_y(x,y) = D\left(\nu\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) \qquad M_{xy}(x,y) = D(1+\nu)\frac{\partial^2 w}{\partial x \partial y} \tag{8}$$

where ν is Poisson's ratio. The plane stress components in turn are written in terms of the plate bending moments as:

$$\sigma_x(x,y) = \frac{6M_x}{t^2} \qquad \sigma_y(x,y) = \frac{6M_y}{t^2} \qquad \tau_{xy}(x,y) = \frac{6M_{xy}}{t^2} \tag{9}$$

The principal plane stress components σ_1 and σ_2 are defined in terms of the plane stress components as [3]:

$$\sigma_{1,2} = \frac{\sigma_x + \sigma_y}{2} \pm \sqrt{\left(\frac{\sigma_x - \sigma_y}{2}\right)^2 + \tau_{xy}^2} \tag{10}$$

Finally, the von Mises stress is defined in terms of the principal plane stress components as:

$$\sigma_v(x,y) = \sqrt{\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2} \tag{11}$$

As a result, the von Mises stress at any location on the plate can be expressed in terms of the deflection surface given by (5) by sequentially substituting equations (8) through (10) into (11).

The onset of yielding along each yield line does not occur uniformly as demonstrated in Figure 5. The stress profile is approximately rectangular with the peak stress occurring at the center of the plate and sharply decreasing near the edge of the plate. It is therein approximated that uniform yielding along each yield line occurs when the peak von Mises stress at the center of the plate achieves the yield stress. The von Mises stress at the center of the plate is solved from (11) to be:

$$\sigma_v(0,0) = \frac{1}{9}\sqrt{\frac{(1+\nu)^2 a^4 q^2}{t^4}} \tag{12}$$

Setting $\sigma_v = \sigma_Y$ and solving for the maximum surface load q_m corresponding to yielding at the center of the plate results in:

$$q_m = \frac{9t^2\sigma_Y}{a^2(1+\nu)} \tag{13}$$

Furthermore, multiplying (13) by the area of the triangle results in an expression for the approximate maximum resistance:

$$R_m = \frac{3t^2\sigma_Y\sqrt{3}}{1+\nu} \tag{14}$$

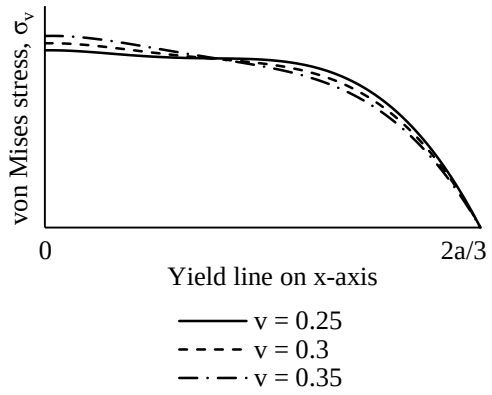


Figure 5: Plot of von Mises stress along yield line on x-axis.

2.4 Shape Function in Plastic Phase

In the plastic phase the plate resistance is R_m and the plate bending stiffness explicit to the displacement at the center of the plate vanishes. The corresponding shape function is assumed to take on the shape of a tetrahedron with its edges defined by the yield lines (see Figure 6). The presence of the edges disallows the use of a continuous shape function for describing the deflection surface over the entire triangular area. As an alternative, a one-dimensional linear shape function, $\Phi(x)$, describing any line in-plane of a tetrahedron face and oriented perpendicular to the perimeter edge can be defined and integrated over each face to obtain

corresponding transformation factors. The linear shape function projected parallel to the x-axis and originating along the perimeter edge at $x = -a/3$ is written as:

$$\Phi(x) = \frac{3x}{a} + 1 \tag{15}$$

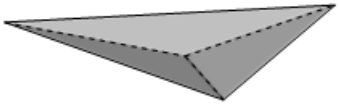


Figure 6: Tetrahedron shape function in plastic phase.

2.5 Transformation Factors

Transformation factors are derived to convert the mass, stiffness, resistance, and load-time history of the real system into equivalent values for use in the equivalent SDOF system, of which is governed by the equations of motion defined by (1) and (2). Transformation factors are obtained for the elastic and plastic phases by integrating the corresponding shape function over the surface of the plate [1]. In general, the mass factor is defined by:

$$K_M = \frac{\int\int^A m\phi^2(x,y)dxdy}{mA} \tag{16}$$

where m is the mass per unit area and A is the area of the triangle. Also, the load factor is defined by:

$$K_L = \frac{\int\int^A q\phi(x,y)dxdy}{qA} \tag{17}$$

Transformation factors for the elastic phase are obtained by integrating $\phi(x,y)$ over the surface of the plate. Integration over the triangular surface requires the limits of integration to vary as functions of the linear equations defining the perimeter of the triangle. Substituting (6) into (16) results in the mass factor for the elastic phase:

$$K_{M,e} = \frac{\int_{-\frac{a}{3}}^{\frac{2a}{3}} \int_{\frac{x\sqrt{3}}{3}-\frac{2a\sqrt{3}}{9}}^{-\frac{x\sqrt{3}}{3}+\frac{2a\sqrt{3}}{9}} m\phi^2(x,y)dydx}{\frac{ma^2\sqrt{3}}{3}} = 0.24 \tag{18}$$

Also, substituting (6) into (17) results in the load factor for the elastic phase:

$$K_{L,e} = \frac{\int_{-\frac{a}{3}}^{\frac{2a}{3}} \int_{\frac{x\sqrt{3}}{3}-\frac{2a\sqrt{3}}{9}}^{-\frac{x\sqrt{3}}{3}+\frac{2a\sqrt{3}}{9}} q\phi(x,y)\,dy\,dx}{\frac{qa^2\sqrt{3}}{3}} = 0.39 \tag{19}$$

Transformation factors for the plastic phase are obtained by integrating the one-dimensional linear shape function $\Phi(x)$ over one-sixth of the total plate area bounded by $x = -a/3$, $y = 0$, and the yield line along $y = x\sqrt{3}$. It is apparent that the considered area is representative of the remaining five divisions of plate area owing to the symmetry of an equilateral triangle. Integration over this area requires the limits of integration to vary as functions of the linear equations defining the yield line. By symmetry it is allowable to multiply the double integrals by six to obtain the final transformation factors. Substituting (15) into (16) results in the mass factor for the plastic phase:

$$K_{M,p} = \frac{6\int_{0}^{\frac{a\sqrt{3}}{3}} \int_{-\frac{a}{3}}^{-\frac{y\sqrt{3}}{3}} m\Phi^2(x)\,dx\,dy}{\frac{ma^2\sqrt{3}}{3}} = 0.17 \tag{20}$$

Also, substituting (15) into (17) results in the load factor for the plastic phase:

$$K_{L,p} = \frac{6\int_{0}^{\frac{a\sqrt{3}}{3}} \int_{-\frac{a}{3}}^{-\frac{y\sqrt{3}}{3}} q\Phi(x)\,dx\,dy}{\frac{qa^2\sqrt{3}}{3}} = 0.33 \tag{21}$$

The derived transformation factors, maximum resistance, and stiffness for the elastic and plastic phases are summarized in Table 1. The response of the equivalent SDOF system may therein be solved using the constant-velocity procedure [1] to obtain the maximum displacement z_{max} in the free vibration phase, from which the distribution of the peak dynamic edge reaction can be determined.

Strain range	Load factor, K_L	Mass factor, K_M	Real stiffness, k	Maximum resistance, R_m
Elastic	0.39	0.24	$\frac{324\sqrt{3}D}{a^2}$	$\frac{3t^2\sigma_Y\sqrt{3}}{1+\nu}$
Plastic	0.33	0.17	0	

Table 1: Parameters for equivalent SDOF model.

3 DISTRIBUTION OF PEAK DYNAMIC EDGE REACTION

Assuming the blast load duration is much shorter than the period of response for the plate element, the distribution of the peak dynamic reaction along the edges of the equilateral triangular plate is dependent upon the dynamic response of the system during the free vibration phase; if the system remains elastic then the reaction distribution is derived from the elastic deflection surface given by (5) and the peak response z_{max}. Conversely, if the system becomes plastic then the reaction distribution is derived from the tetrahedron shape function and R_m.

3.1 Edge Reaction Distribution in Elastic Phase

The distribution of the peak dynamic reaction for the elastic phase is first developed. The edge reaction distribution requires a relation between z_{max} in the free vibration phase and the corresponding plate shear force, $Q_i(x,y)$, at any location on the plate. It is apparent that the shear force distribution along the perimeter edges corresponding to z_{max} is identical to the peak dynamic edge reaction. The plate shear force is related to the plate bending moments by [2]:

$$Q_x(x,y) = \frac{\partial M_x}{\partial x} + \frac{\partial M_{xy}}{\partial y} \qquad Q_y(x,y) = \frac{\partial M_{xy}}{\partial x} + \frac{\partial M_y}{\partial y} \tag{22}$$

Sequentially substituting (5) and (8) into $(22)_1$ and setting $x = -a/3$ results in an expression for the edge reaction distribution along the perimeter edge parallel to the y-axis due to uniform surface load q:

$$Q_{x,eu}\left(-\frac{a}{3}, y\right) = -\frac{q}{8a}\left(2a^2 - 6y^2 + \nu a^2 - 9\nu y^2\right) \tag{23}$$

The force developed into the perimeter edge supports during the free vibration phase is derived from the inertia force. The distribution of the inertia force across the plate surface is non-uniform and is proportional to the deflection surface. As a result, the edge reaction distribution given by (23) must be scaled by a ratio, α, of the uniform surface load q_E corresponding to the peak displacement z_{max} at the center of the plate, and an equivalent surface load $r_E(x,y)$ proportional to $\phi(x,y)$. The equivalent surface load $r_E(x,y)$ represents the inertia force imposed upon the plate.

The uniform surface load q_E corresponding to z_{max} is solved from (5) to be:

$$q_E = \frac{972 D z_{\max}}{a^4} \tag{24}$$

Conversely, the equivalent surface load $r_E(x,y)$ is expressed in terms of a unit uniform surface load q_u and a constant C_e which controls the amplitude of $\phi(x,y)$ at the center of the plate:

$$r_E(x,y) = q_u C_e \phi(x,y) \tag{25}$$

The total force imposed by $r_E(x,y)$ upon the plate must be equivalent to the total force developed into the perimeter edge supports corresponding to z_{max}. The constant C_e is determined by integrating $r_E(x,y)$ over the surface of the triangular plate and equating the result with the total force derived from the constitutive relation corresponding to z_{max}:

$$q_u C_e \int_{-\frac{a}{3}}^{\frac{2a}{3}} \int_{\frac{x\sqrt{3}}{3}-\frac{2a\sqrt{3}}{9}}^{-\frac{x\sqrt{3}}{3}+\frac{2a\sqrt{3}}{9}} \phi(x,y)\,dy\,dx = kz_{\max} \tag{26}$$

where the limits of integration vary as functions of the linear equations defining the perimeter of the triangular plate. Substituting (6) and (7) into (26), resolving the integral, solving for C_e, and substituting the result into (25) gives an expression for the equivalent surface load:

$$r_E(x,y) = \frac{2520 D z_{\max}}{a^4}\phi(x,y) \tag{27}$$

The ratio α is developed by considering the forces per unit length, Q_E and R_E, developed into a perimeter edge as derived from surface loads q_E and $r_E(x,y)$, respectively, within the corresponding tributary width (see Figure 7). The boundaries delineating the tributary width of each perimeter edge are coincident with the assumed yield lines. Considering the tributary area bounded by $x = -a/3$, $y = 0$, and the yield line along $y = x\sqrt{3}$, the ratio α is defined as follows:

$$\alpha(y) = \frac{R_E(y)}{Q_E(y)} \tag{28}$$

where $Q_E(y)$ and $R_E(y)$ are the forces per unit length developed into the perimeter edge along $x = -a/3$. It is apparent that the magnitudes of $Q_E(y)$ and $R_E(y)$ are represented by the areas beneath q_E and $r_E(x,y)$ at any section parallel to the x-axis between $y = 0$ and $y = a\sqrt{3}/3$, and within the domain bounded by $x = -a/3$ and $x = -y\sqrt{3}/3$ (see Figure 8). The force per unit length as derived from q_E is solved to be:

$$Q_E(y) = q_E\left(\frac{a}{3} - \frac{y\sqrt{3}}{3}\right) \tag{29}$$

Also, the force per unit length as derived from $r_E(x,y)$ is defined by:

$$R_E(y) = \int_{-\frac{a}{3}}^{-\frac{y\sqrt{3}}{3}} r_E(x,y)\,dx \tag{30}$$

where the limits of integration vary as functions of the linear equations defining the yield lines.

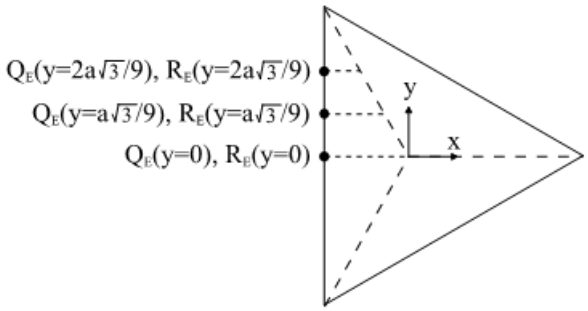


Figure 7: Forces per unit length $Q_E(y)$ and $R_E(y)$ developed into the perimeter edge along $x = -a/3$.

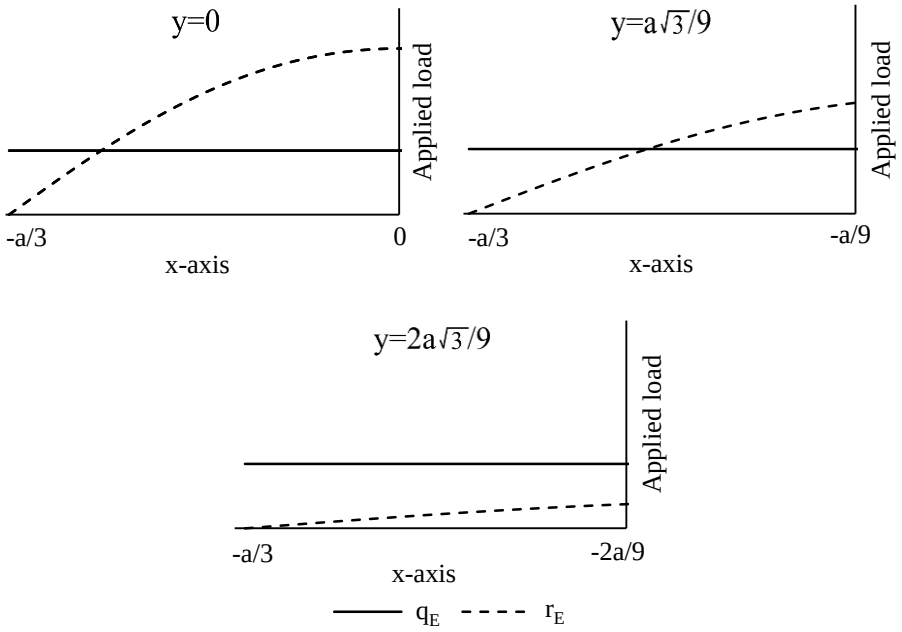


Figure 8: Plots of surface loads q_E and r_E tributary to the perimeter edge along $x = -a/3$.

Substituting (24) into (23) and multiplying the result by $\alpha(y)$ gives an expression for the peak dynamic edge reaction distribution during the elastic phase in terms of z_{max} along the edge at $x = -a/3$ and within the domain bounded by $y = 0$ and $y = a\sqrt{3}/3$:

$$Q_{x,e}\left(-\frac{a}{3}, y\right) = -\alpha(y)\frac{81Et^3 z_{\max}}{8a^5\left(\nu^2 - 1\right)}\left(2a^2 - 6y^2 + \nu a^2 - 9\nu y^2\right) \tag{31}$$

Full substitution of (28) into (31) results in an extensive formula and is not explicitly shown here. It is apparent that this distribution is mirrored in the domain bounded by $y = -a\sqrt{3}/3$ and $y = 0$. Also, the distribution is identical along the other two perimeter edges.

3.2 Edge Reaction Distribution in Plastic Phase

In the plastic phase the peak dynamic reaction Q_{max} is capped by the maximum resistance R_m expressed by (14). The corresponding edge reaction distribution is derived by disseminating an equivalent surface load $p_E(x,y)$ possessing a total magnitude equal to R_m in accordance with the tributary width of each perimeter edge. The boundaries delineating the tributary width of each perimeter edge are coincident with the assumed yield lines. The distribution of $p_E(x,y$ across the plate is proportional to the tetrahedron shape function described by $\Phi(x)$ and represents the capped inertia force imposed upon the plate during the plastic phase.

Considering the tributary area bounded by $x = -a/3$, $y = 0$, and the yield line along $y = x\sqrt{3}$, the equivalent surface load $p_E(x,y)$ is expressed in terms of the unit uniform surface load q_u and a constant C_p which controls the amplitude of $\Phi(x)$ at the center of the plate:

$$p_E(x,y) = q_u C_p \Phi(x) \tag{32}$$

It is apparent that the considered tributary area is representative of the remaining five divisions of plate tributary area because of the symmetry of an equilateral triangle. Therefore, the force imposed by $p_E(x,y)$ upon the plate within the considered tributary area must be equivalent to one-sixth of the maximum resistance R_m. The constant C_p is determined by integrating $p_E(x,y)$ over the surface of the considered tributary area and equating the result with one-sixth of R_m:

$$q_u C_p \int_0^{\frac{a\sqrt{3}}{3}} \int_{-\frac{a}{3}}^{-\frac{y\sqrt{3}}{3}} \Phi(x)\,dx\,dy = \frac{R_m}{6} \tag{33}$$

where the limits of integration vary as functions of the linear equations defining the yield lines. Substituting (14) and (15) into (33), resolving the integral, solving for C_p, and substituting the result into (32) gives an expression for the equivalent surface load:

$$p_E(x,y) = \frac{27t^2\sigma_Y}{a^2(1+\nu)}\Phi(x) \tag{34}$$

The dissemination of $p_E(x,y)$ into the perimeter edge along $x = -a/3$ is expressed by:

$$Q_{x,p}\left(-\frac{a}{3}, y\right) = \int_{-\frac{a}{3}}^{-\frac{y\sqrt{3}}{3}} p_E(x,y)\,dx \tag{35}$$

where the limits of integration vary as functions of the linear equations defining the yield lines. Sequentially substituting (15) and (34) into (35) and resolving the integral results in an expression for the dynamic edge reaction distribution during the plastic phase along the edge at $x = -a/3$ and within the domain bounded by $y = 0$ and $y = a\sqrt{3}/3$:

$$Q_{x,p}\left(-\frac{a}{3}, y\right) = -\frac{9t^2\sigma_Y}{2a^3(1+\nu)}\left(-3y^2 - a^2 + 2ay\sqrt{3}\right) \tag{36}$$

It is apparent that this distribution is mirrored in the domain bounded by $y = -a\sqrt{3}/3$ and $y = 0$. Also, the distribution is identical along the other two perimeter edges.

4 VALIDATION WITH DYNAMIC FINITE ELEMENT ANALYSES

4.1 Analytical Setup and Equivalent SDOF Analysis

The distribution of peak blast-induced dynamic reactions along the edges of three steel plate wall elements are determined using the equivalent SDOF model together with (31) and (36). The resulting reaction distributions are verified with dynamic FEA using ABAQUS/CAE. The three steel plate wall elements are assumed to be positioned within the braced frame lateral system of an imaginary 10-story building as shown in Figure 9. Each wall element is simply-supported in-plane of an inverted V-braced frame as shown in Figure 10a. Furthermore, each wall element is assumed to be formed in the shape of an equilateral triangle. It is noted that only equilateral triangular wall elements are assessed; other non-rectangular wall elements are deemed beyond the scope of this study but are considered no less important.

The wall elements are designed to protect the building's MEP systems from an assumed external street threat equivalent to the explosion of 1100 kg of TNT. The equivalent uniform peak pressure and impulse imposed upon each wall element are obtained using the blast effects software ConWep (see Table 2). The blast infill pressure is assumed to be described by a linearly decaying pressure-time history *q(t)* as shown in Figure 10b and does not account for the façade or any shielding by the floor slabs. The equivalent SDOF model is used to design the wall elements to meet the performance criteria recommended by ASCE 59-11 *Blast Protection of Buildings* [4] for moderate damage. Namely, the parameters summarized in Table 1 are used to develop an equivalent SDOF system for each wall element and the response of the system is solved for using the constant-velocity procedure [1]. Final wall element thicknesses are converged upon through iteration. The resulting geometric and material properties of each wall element are summarized in Tables 3 and 4, respectively. The resulting peak out-of-plane displacements at the center of each wall element during the free vibration phase are summarized in Table 5.

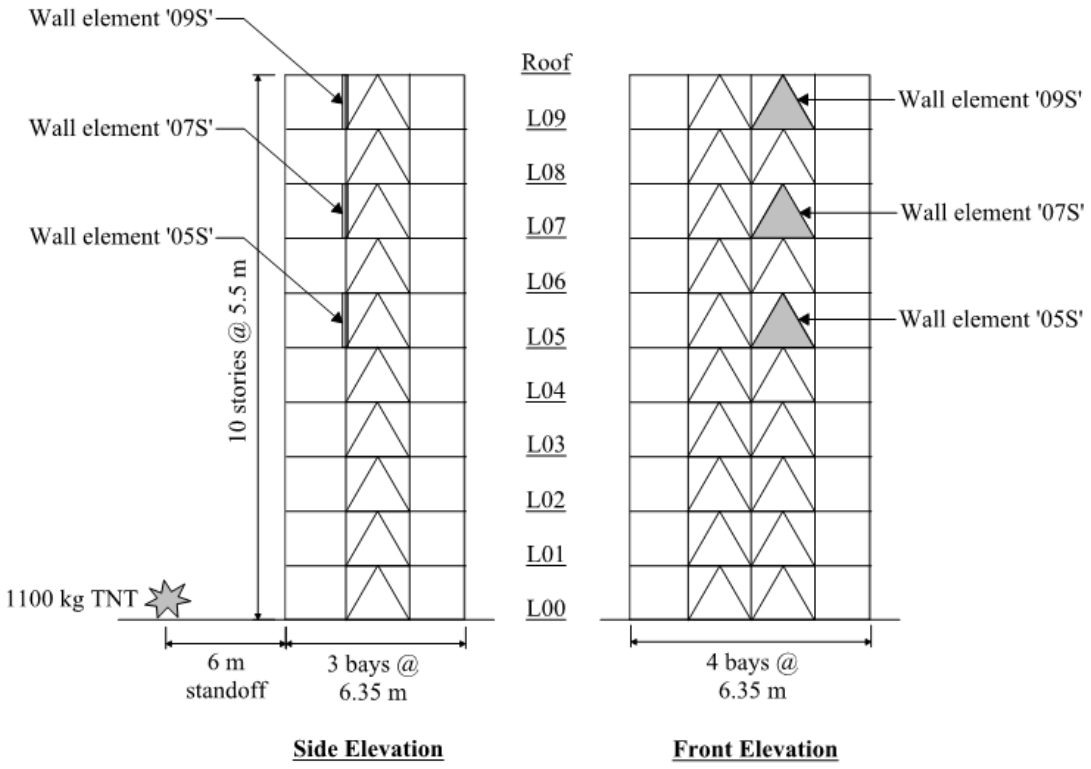


Figure 9: Imaginary 10-story building with steel plate wall elements in-plane of braced frame.

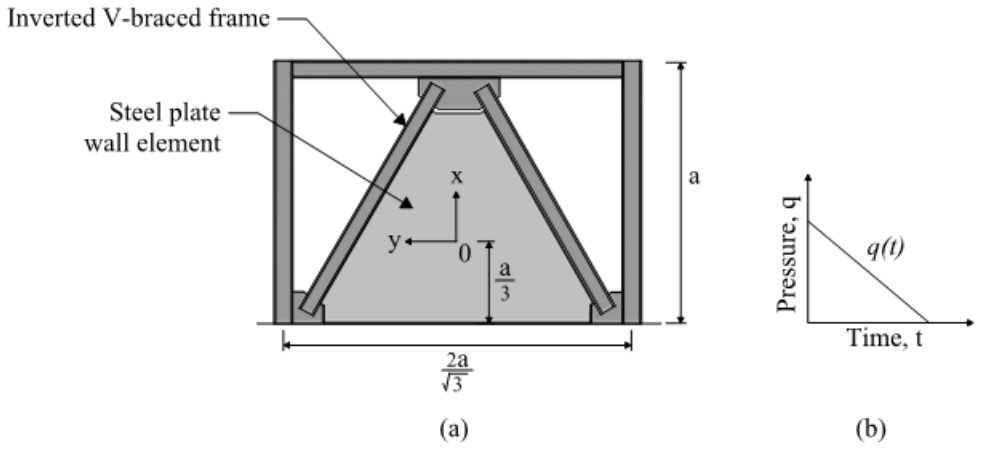


Figure 10: (a) Steel plate wall element loaded by (b) linearly decaying pressure-time history.

Wall element	Range to blast (m)	Uniform peak pressure (kPa)	Uniform impulse (kPa-msec)
05S	36	145	1158
07S	47	76	848
09S	58	48	669

Table 2: Blast pressure and impulse imposed on each wall element.

Wall element	Material[1]	Height (m)	Width (m)	Thickness, t (cm)
05S	Steel plate	5.5	6.35	4.5
07S	Steel plate	5.5	6.35	4.0
09S	Steel plate	5.5	6.35	3.5

1. Refer to Table 4 for material properties.

Table 3: Wall element geometric properties.

Material	Yield strength, σ_Y (MPa)	Young's modulus, E (GPa)	Poisson's ratio, ν	Density, ρ (kg/m^3)
Steel plate	250	200	0.30	7860

Table 4: Wall element material properties [3].

Wall element	Peak SDOF response, z_{max} (cm)	Rotation, θ	Ductility, μ	Rotation response limit, θ_{max}	Ductility response limit, μ_{max}
05S	5.8	1.82˚	0.90		
07S	5.2	1.63˚	0.71	2˚	8
09S	5.3	1.65˚	0.63		

Table 5: Peak SDOF responses and ASCE 59-11 response limits for steel plates considering moderate damage [4].

4.2 FEA Setup and Analysis

The three steel plate wall elements are modeled in ABAQUS/CAE in accordance with the geometric and material properties specified in Tables 3 and 4, respectively. Triangular shell elements are used to mesh the geometry, and simple supports are applied to the perimeter edges of each modeled wall element (see Figure 11). A finer mesh is developed in the vicinity of the three corner regions. Zero damping is assigned to the section properties.

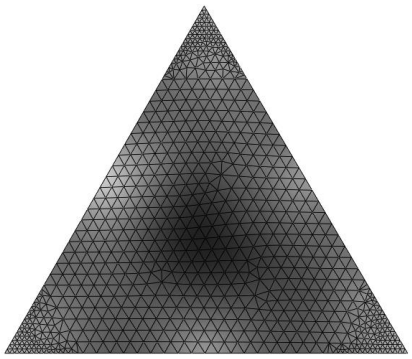


Figure 11: Typical steel plate wall element as modeled in ABAQUS.

The dynamic analysis for each wall element is carried out in two steps. First, an implicit dynamic load step is executed in which a uniform pressure is applied to the wall element. The uniform pressure is described by a linearly decaying ramp function (see Figure 10b) formulated in accordance with the pressure-time history described by the equivalent uniform peak pressure and impulse specified in Table 2. A second implicit dynamic load step is next performed in the absence of external loading to model the free vibration phase. History output requests are obtained at each time step for the out-of-plane nodal displacements and edge reactions.

Displacement-time history plots of the out-of-plane displacement at the center of each wall element as obtained from FEA and the equivalent SDOF system are displayed in Figure 10. The time history is displayed from the initial response to the blast infill pressure to 200 msec. The FEA and SDOF time history plots demonstrate very good correspondence.

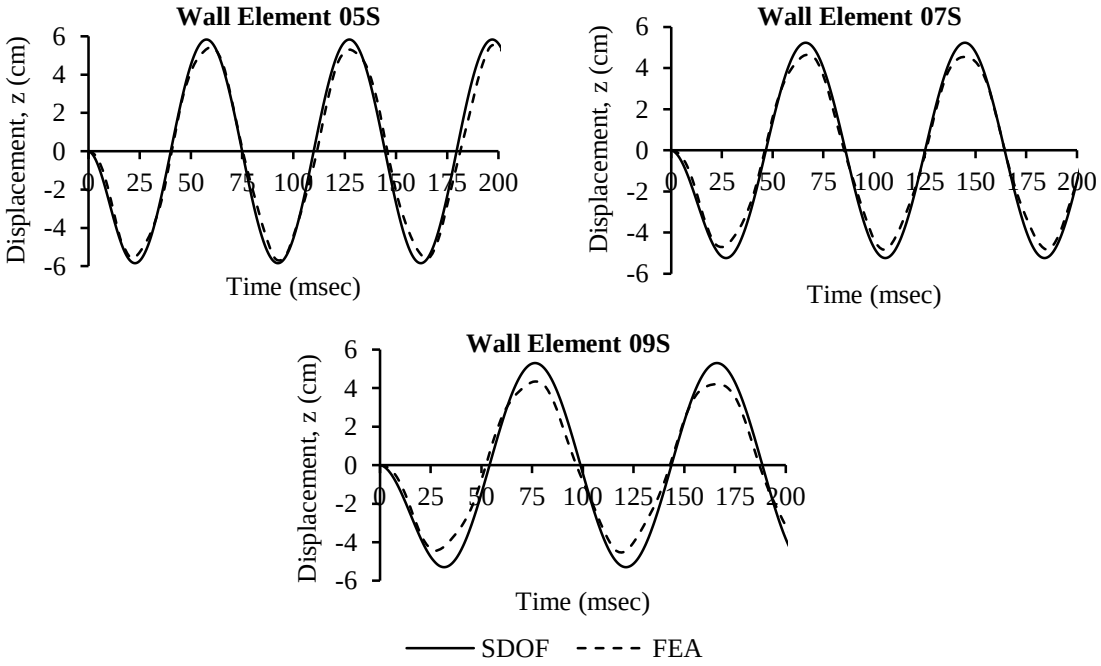


Figure 10: Displacement-time history plots of the out-of-plane displacement at the center of each wall element as obtained from SDOF and FEA.

4.3 Distribution of Peak Dynamic Edge Reaction

The responses of all three steel plate wall elements remain elastic throughout the free vibration phase (i.e. ductility, $\mu < 1.0$). As a result, the peak dynamic edge reaction distribution is determined using (31). Substituting into (31) the geometric and material properties listed in Tables 3 and 4, and the peak responses summarized in Table 5, allows for the computation of the peak dynamic edge reaction distribution for each wall element along the edge located at $x = -a/3$ and between $y = 0$ and $y = a\sqrt{3}/3$ (see Figure 11). The resulting edge reaction distribution is representative of the distributions along the remaining five edge segments owing to the symmetry of an equilateral triangle and the uniform blast infill pressure.

The peak dynamic edge reaction distribution for each wall element during the free vibration phase is determined in ABAQUS/CAE by way of three steps. First, a nodal path is created along an edge of each wall element. Specifically, the edge located at $x = -a/3$ and between $y = 0$ and $y = a\sqrt{3}/3$ is investigated. Next, the time step of peak displacement during the free vibration phase is determined from the displacement-time history plots. Finally, the edge reaction distribution along the nodal path is obtained from the history output requests at the time step of peak displacement (see Figure 11). The peak dynamic edge reaction distribution for each wall element as obtained from FEA and (31) demonstrates good correspondence.

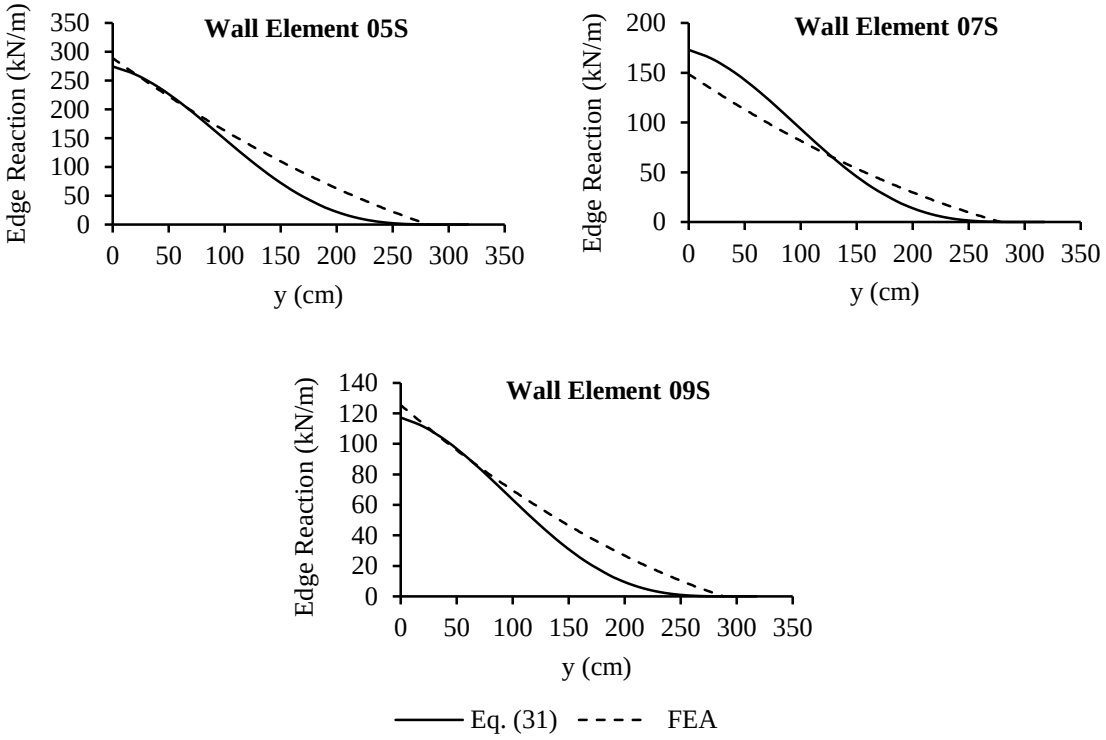


Figure 11: Time history plots of the out-of-plane displacement at the center of each wall element as obtained from SDOF and FEA.

5 CONCLUSIONS

Classical plate theory and mechanics of materials were employed to develop an equivalent SDOF model for describing the dynamic response of an equilateral triangular hardened wall element simply-supported in-plane of an inverted V-braced frame and uniformly loaded by blast infill pressure. Expressions (31) and (36) were then derived using the aforementioned theories for describing the distribution of the peak dynamic edge reactions along the perimeter supports of the wall element. The equivalent SDOF system and (31) were subsequently validated with dynamic FEA using ABAQUS/CAE.

The process of designing the lateral system of a tall building for wind and seismic loads is made more complex by also having to analyze and design individual braced frame members for out-of-plane blast loads as caused by blast-induced dynamic reactions of hardened walls. The presence of numerous quantities of hardened walls and supporting braced frame members within the tall building may make impractical the use of dynamic FEA for analyzing and designing the walls and braced frame members. The parameters summarized in Table 1 are highly applicable in that they may be easily incorporated into standard SDOF blast analysis spreadsheets for rapidly analyzing the dynamic response of hardened wall elements built in a form equal to or near to that of an equilateral triangle. The final design of wall elements may be refined by adjusting the wall geometric and material properties until an optimal design is converged upon through iteration. Importantly, the corresponding distributions of peak blast-induced dynamic reactions along the perimeter supports of wall elements, given by (31) and (36), may be used to efficiently analyze and design the supporting braced frame members.

Although this study dealt with homogeneous and isotropic equilateral triangular wall elements possessing elastic-perfectly plastic material properties, this work can indubitably be amended and expanded to include non-isotropic or non-homogeneous walls, such as reinforced concrete or stud walls. This would certainly involve modification of the assumed deformed shapes and out-of-plane stiffness. Closed-form solutions for wall elements built in the form of other non-rectangular shapes, such as wide isosceles triangles or right triangles, are increasingly complex. However, the analysis of an equilateral triangle is expected to envelop many cases seen in tall buildings.

REFERENCES

[1] J.M. Biggs, *Introduction to structural dynamics*. McGraw-Hill, 1964.

[2] S. Timoshenko, S. Woinowsky-Krieger, *Theory of plates and shells, 2nd Edition*. McGraw-Hill, 1987.

[3] F.P. Beer, E.R. Johnston, Jr., J.T. DeWolf, *Mechanics of Materials, 4th Edition*. McGraw-Hill, 2006.

[4] ASCE 59-11, *Blast Protection of Buildings*. American Society of Civil Engineers, 2011.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELLING STRATEGIES OF PRESTRESSING TENDONS AND REINFORCEMENT BARS IN CONCRETE STRUCTURES

Antoine Llau[1,2,3], Ludovic Jason[1,2], Frédéric Dufour[3,4], and Julien Baroth[3,4]

[1]SEMT, CEA DEN, 91191 Gif sur Yvette Cedex, France
e-mail: {antoine.llau,ludovic.jason}@cea.fr

[2]IMSIA, CEA, CNRS, EDF, ENSTA Paristech, Université Paris Saclay,
91762 Palaiseau Cedex, France

[3]Univ. Grenoble-Alpes, 3SR, 38041 Grenoble Cedex 9, France
e-mail: {julien.baroth,frederic.dufour}@3sr-grenoble.fr

[4]CNRS, 3SR, 38041 Grenoble Cedex 9, France

Keywords: Concrete cracking, 3D model, stress concentration, static condensation, steel.

Abstract. *This contribution presents an original approach to improve the modeling of steel rebars and prestressing tendons in concrete structures at a reduced cost. Classical 1D meshes and models typically used for civil engineering applications tend to provoke strain localization due to the geometrical singularity and are thus unable to reproduce local mechanical effects. Complete 3D models can be applied in some cases, however their accuracy at the local scale comes at the cost of engineering work on the meshes, especially for complex structures. The 1D-3D model presented in this contribution generates an equivalent volume for the steel bars, based on existing 1D models. Its 3D stiffness and stress state are computed, and then condensed on its interface with the concrete. The condensed degree of freedom are then linked to the surrounding concrete elements by kinematic relations. The presented approach is validated on different representative cases, and is able to predict the 3D effects of the bars and tendons at the local scale. In particular it provides the representativeness and mesh stability of a full 3D model, without the need for a complex mesh.*

1 INTRODUCTION

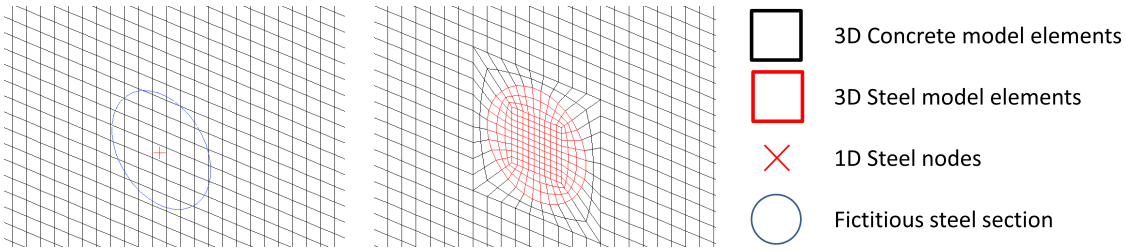
In civil engineering structures, concrete is usually reinforced and/or prestressed using passive and active steel reinforcements. These reinforcements constitute material heterogeneities, and have to be modeled to predict the behavior of the structure. They are usually considered as unidimensional inclusions, considering the ratio of their length to their cross section. Therefore, these reinforcements are typically modeled in FE analyses using 1D elements, with beam or truss kinematics [4]. In some cases, where the three-dimensional effects of the tendon or rebar section have to be taken into account, full 3D models using conform meshes can be applied [8].

In the case of 1D models, the cross section of the steel is neglected, compared to the concrete matrix FE size. The models are superimposed, and the cross section has no influence on stress and stiffness repartition in the concrete. This approach has no specific mesh requirements, and is thus very easy to apply on various (and particularly, large-scale) structures. Therefore, a geometrical singularity is created. Although the global effect of reinforcement and prestressing is well reproduced, this may lead to pathological mesh dependency effects due to stress concentration at the singularity [1]. Also, transversal and cover effects can hardly be modeled with this approach.

In the case of full 3D models, using a geometrical and mechanical representation of the inclusions, the kinematics is much richer. This approach is both representative at the small scale (transversal effects can be predicted) and stable toward the mesh. The counterpart is the additional mesh complexity required to match the geometry of the heterogeneity.

In this contribution, a new approach is proposed, called the 1D-3D model (1D inclusions / 3D finite elements). It aims to combine the advantages of both modeling strategies, without the need for a specific mesh. The proposed model is first described. Then, an application on the numerical model of a nuclear containment building RSV is presented, where the different models are compared [7].

2 MODELLING APPROACH



(a) 1D Model (b) Full 3D Model

Figure 1: Modelling of a tendon in a concrete volume.

Fig. 1 presents the meshes used to model a reinforcement in a concrete volume, for the 1D and 3D models. The 1D-3D approach aims at combining the upsides of these models. Its main objective is to conserve the ease of use and versatility of the 1D approach, while having the representative local results and robustness of a full 3D approach. It should provide a single model for the reinforcements in concrete, with a validity domain overlapping the 1D and 3D approaches. Moreover, existing FE models can be used with this new model, as it uses 1D meshes. The proposed "1D-3D" approach requires four steps:

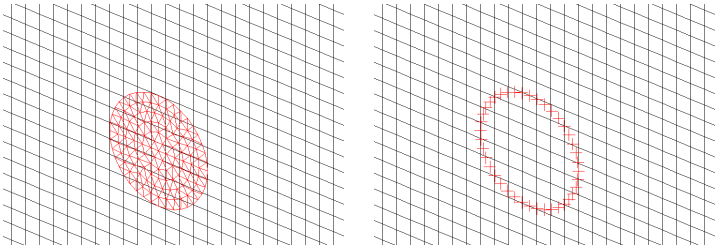
- Generation of the 3D equivalent mesh of the bar.

- Computation of the equivalent stresses.
- Condensation of the 3D stiffness and stresses.
- Construction of the steel-concrete interface.

2.1 Generation of the 3D equivalent mesh

To be able to represent the 3D effects of the steel bars on the concrete, a 3D equivalent mesh must be generated from existing 1D meshes. This allows to keep the adaptability of 1D meshes, as well as applying this new model to existing FE models of structures. The equivalent 1D-3D mesh is generated by projecting a circular section along the axis of the bar. The circular section is meshed with triangles using a concentrical circles method. The 3D equivalent volume is therefore meshed with prismatic elements. The 3D concrete domain is not modified.

Fig. 2 presents the 1D-3D equivalent mesh for a reinforcement in a concrete domain, and the nodes at the steel-concrete interface.



(a) Equivalent steel volume (b) Steel-concrete interface

Figure 2: Modelling of a tendon in a concrete volume.

2.2 Equivalent stresses

In the case of active reinforcements such as prestressing tendons, or pre-tensioned rods, stresses are applied to the steel member, which transmits it to concrete. The proposed 1D-3D model requires transposing a 1D stress state to the equivalent 1D-3D volume. The 1D stress state of the steel is first converted to nodal forces:

$$F_i = S \int_{L_i} \frac{d\sigma_{1D}}{dl} \mathrm{d}l \tag{1}$$

whereF_i is the nodal force vector at node i, σ_{1D} the stress value, S the bar's cross section and l the abscissa along the bar, integrated on the length L_i (depending of the finite elements form). These nodal forces are then applied to the 1D-3D model previously generated, creating a stress state such that:

$$F_i = \int_{V_i} \nabla\sigma_{1D3D} \mathrm{d}V \tag{2}$$

where σ_{1D3D} is the stress tensor, $\nabla\sigma_{1D3D}$ its divergence, V the volume of the bar, integrated on V_i (depending of the FE geometry and interpolation functions). Thus, the stress state computed using a 1D model can also be applied to the 1D-3D model.

2.3 Condensation

The 1D-3D model generates a large number of additional FEs and DOFs compared to classical 1D approaches. This may become an issue in terms of computational cost if large RC structures with high levels of reinforcements are considered. However, an approach is proposed to reduce the number of DOFs, similar to the condensation method presented in [9]. This method is based on Guyan's static condensation [6], and replaces the 1D-3D model by its reduced stiffness matrix and loading, applied on the envelope of the 1D-3D bar. This way, the volume is reduced to a surface.

However, the reduced stiffness matrices are smaller but also fuller than standard stiffnesses, which may impact computational performance wen solving the linear system. To avoid this problem, low-rank approximations of the condensed stiffnesses may be used. Depending on the configurations, the condensation may or may not improve computational performance. Since this step is not strictly necessary in the 1D-3D approach, it should be applied only in the situations where it improves performance.

2.4 Steel-concrete interface

The steel-concrete bond in ensured by the interface nodes of the bar, related to the adjacent concrete elements by kinematic relations. These relations transmit the stiffness and force field of the 1D-3D model to the adjacent concrete nodes. The displacement u_i of the steel node i located at x_i is such that:

$$u_i = \sum_j u_j N_j(x_i) \tag{3}$$

where u_j are the displacements of the concrete nodes j adjacent to i, and N_j the interpolation functions of the concrete domain. This model is similar to the perfect bond usually considered for 1D reinforcement bars (in the case of non-corresponding meshes). This relation could be improved to include more representative steel-concrete bond models.

As in the classical 1D model, the two materials are geometrically superimposed: the additional stiffness is neglected. This method has been developed to be used without additional adaptations and/or meshing effort on existing 1D meshes. It provides, for a small additional computational cost, more accurate local stress fields and removes the pathological mesh dependency.

3 MODELLING OF A PRESTRESSED RSV

3.1 Presentation

In this section, a structural application case is considered. It consists in a numerical model of a representative structural volume (RSV) of a prestressed and reinforced concrete containment building, on which a pressure test is performed. This particular structure is strongly dependent on the prestressing tendons, which affect damage localization and failure mode [8]. In particular, previous studies have demonstrated an inclusion effect around the vertical tendon. This effect is not predicted with 1D models, but can be analyzed with full 3D models. The aim is thus to evaluate the capacity of the 1D-3D approach to predict this effect.

Fig. 3 presents the geometry of the structure: it consists in a curved, horizontally prestressed concrete volume on which an internal pressure is applied. A non prestressed vertical tendon (the transverse inclusion), and passive reinforcements are also included. The prestress is first applied through the horizontal tendons, with an associated vertical compressive stress; the internal pressure is then applied.

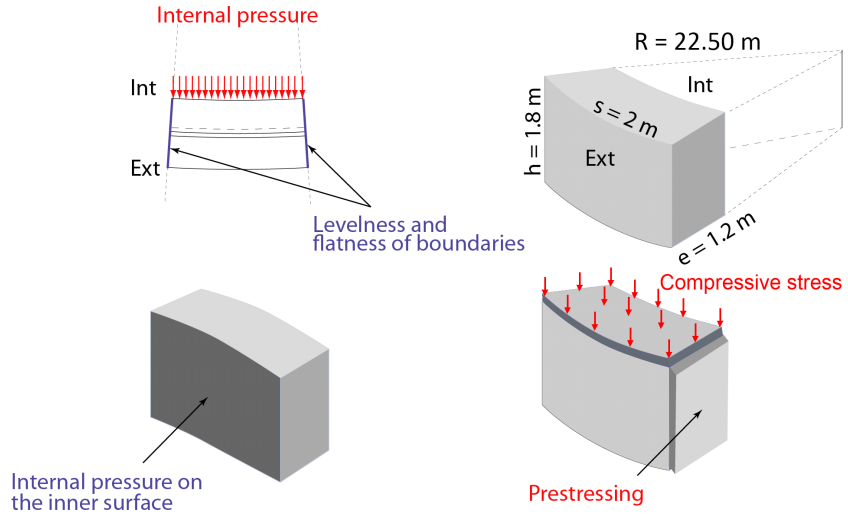


Figure 3: Geometry and loading of the prestressed concrete RSV [8].

The concrete is modelled using Mazars' isotropic damage model [10] (compressive strength $f_c = 37.6$ MPa, tensile strength $f_t = 3.57$ MPa). The stress-based nonlocal method is applied [5], with a nonlocal internal length $l_{C0} = 4$ cm. The passive steel reinforcements are modeled using equivalent shell FEs. All simulations are performed using Cast3M [2], with post-processing using Salome [3]. On this structure, 1D, 1D-3D and full 3D models are compared regarding global and local results.

3.2 Results

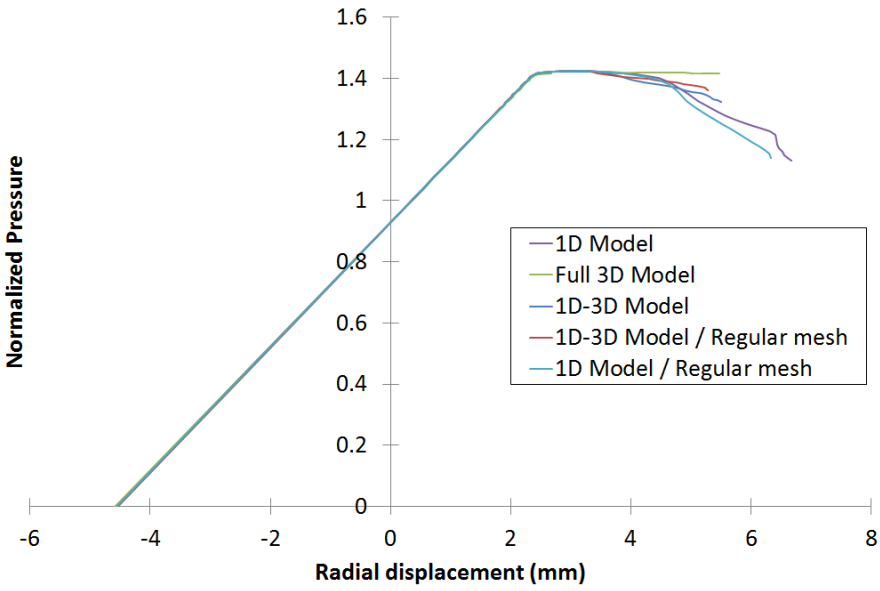
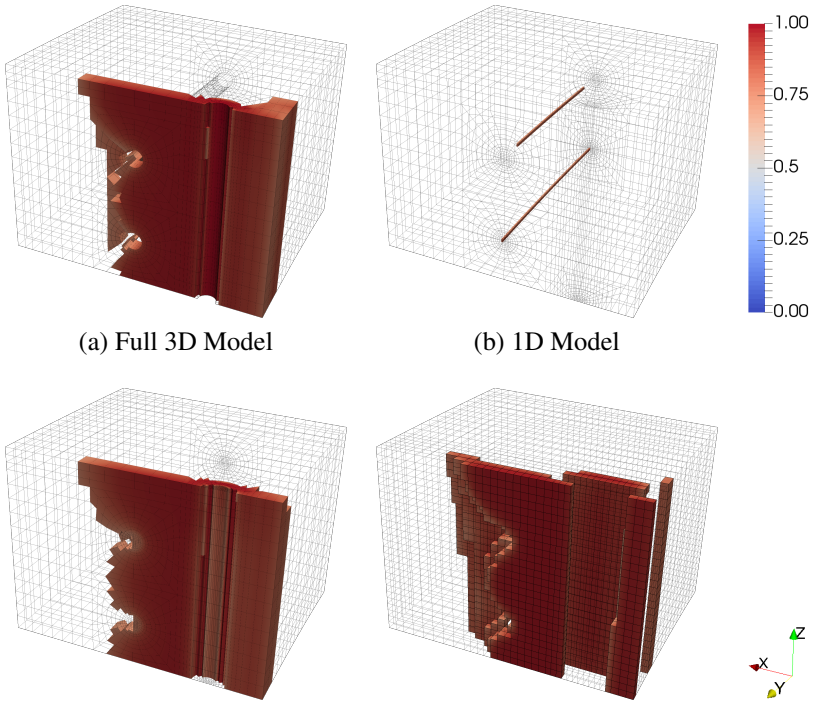


Figure 4: Pressure vs. Radial displacement evolution of the RSV with different tendon models.

To analyze the global nonlinear behavior, Fig. 4 presents the pressure-radial displacement evolutions for the different models. The prestress provokes an initial inward radial displacement. After the elastic part of the behavior, a nonlinear behavior is then observed: the pressure decreases while the (outward) radial displacement increases. All models provide very similar

results in this respect: the choice of the model used for the steel tendons (both horizontal and vertical) has little impact on the global behavior.

To analyze local effects, and in particular the inclusion effect of the vertical tendon, Fig. 5 compares for the different models the damage profiles in the RSV, for a radial displacement of 4.5 mm. The 1D-3D approach in particular, is applied using both the explicit 3D mesh with geometrical representation of the tendons, and a regular mesh, with 1D representation of the tendons. It appears that, excluding the inside of tendons (there is superimposition in the 1D-3D and not in the full 3D model), the distributions are nearly identical between 1D-3D, with both explicit and regular meshes, and the explicit 3D model. The effect of the vertical tendon on damage initiation and propagation is well reproduced. The 1D approach, however, does not reproduce this effect and features a more distributed damage in the whole volume, as well as damage localized around the horizontal singularities. Therefore, the proposed approach clearly improves the description of the mechanical behavior at the local scale. It is able to predict the effect of the vertical tendon without the additional meshing effort required to represent it explicitly.



(a) Full 3D Model

(b) 1D Model

(c) 1D-3D Approach (conform mesh)

(d) 1D-3D Approach (regular mesh)

Figure 5: Damage profiles for a radial displacement of 4.5 mm (only the areas with $D \geq 0.80$ are represented).

4 1D-3D APPROACH FOR UNCERTAIN GEOMETRY

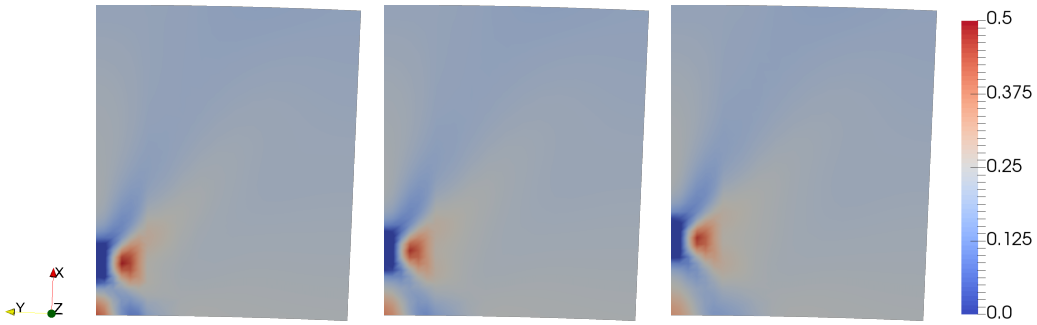
4.1 Presentation

As presented earlier, one of the main advantages of the 1D-3D approach in modeling reinforcement bars and prestressing tendons in concrete is its flexibility. As it is based on a simple 1D geometrical description of the reinforcements and can be used without a specific concrete

mesh, it is well suited for applications such as sensitivity analyses or uncertainty quantification (e.g SFE analyses), including geometrical uncertainty.

In this section, an example of application of the 1D-3D approach is presented, that uses the same structural case than the preceding section. In particular, the sensitivity of the damage profile in the prestressed RSV to the position of the vertical tendon will be studied. The 1D-3D approach, using a regular mesh for concrete, will be applied for this study. Three configurations are modeled and analyzed. In the first, the cover at which the vertical tendon (Ø 84 mm) is placed in the RSV is increased by 46 mm. In the second, the tendon is located at its nominal position (as in the preceding section), with a concrete cover of 208 mm. In the third, the cover is reduced by 46 mm. The damage profiles obtained with the three configurations are compared, in order to observe the local effects and how it is influenced by the tendon position.

4.2 Results



(a) Reduced cover (b) Nominal position (c) Increased cover

Figure 6: Damage profiles for a radial displacement of 3.0 mm with different positions of the vertical tendon (top view).

Fig. 6 presents and compares the damage profiles obtained with the three tendon positions. Damage profiles of each the increased/reduced configurations are compared locally to the reference configuration of the modeled RSV. In order to observe more easily the difference in localization of damage around the tendon, damage is observed at an earlier stage (radial displacement of 3.0 mm). We observe at first that the inclusion effect, leading to localized damaged areas, is still reproduced with all three positions using the 1D-3D model. In all three cases, damage seems to initiate on the side of the vertical tendon, following the perturbed position. Also, it appears that the concrete elements located inside the 1D-3D tendon are protected in all three simulations. This protected area follows quite correctly the perturbed position of the 1D-3D tendon. A second damaged area appears on the inside face of the RSV, in front of the vertical tendon. This area seems to be more damaged when the concrete cover of the tendon is reduced. The effect of the vertical tendon (and in particular the inclusion effect) is well reproduced in those configurations, and the influence of its position can be observed and quantified. This validates the applicability of the 1D-3D approach (in particular, using a regular mesh), for analyses in a uncertain geometry context.

5 CONCLUSIONS

A new approach, called 1D-3D, for the modeling of prestressing tendons and reinforcement bars in concrete structures has been presented. This method creates an equivalent volume from

the 1D mesh, with static condensation of the stresses and stiffness to limit the computational cost: no additional meshing requirements are necessary compared to classical 1D mesh. The reduced stiffness and nodal forces of the steel are applied on the envelope of the equivalent volume, and bonding to the concrete is then performed using kinematical relations. This method is validated on a representative structural volume of prestressed concrete containment buildings, on which geometrical uncertainties are considered. The 1D-3D approach avoids the well-known stress concentration around singularities observed with 1D models. It is also able to reproduce local structural effects, such as transversal effects. Therefore, the proposed 1D-3D approach combines in a single model the versatility and efficiency of classical 1D models at the large scale with the representativeness and the numerical stability of full 3D models.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding support from French National Research Agency through "MACENA" project (ANR-11-RSNR-012).

REFERENCES

[1] P. Badel and E. Lorentz. Critères de convergence en mécanique des solides. In *CSMA 2011 / 10e colloque national en calcul des structures*, Giens, France, May 2011.

[2] CEA. Description of the finite element code Cast3m. http://www-cast3m.cea.fr/, 2015.

[3] CEA, EDF, and OpenCascade. Salome : The open-source integration platform for numerical simulation. http://salome-platform.org/, 2015.

[4] J. A. Figueiras and R. H. C. F. Póvoas. Modelling of prestress in non-linear analysis of concrete structures. *Computers & Structures*, 53(1):173–187, Oct. 1994.

[5] C. Giry, F. Dufour, and J. Mazars. Stress-based nonlocal damage model. *International Journal of Solids and Structures*, 48(25–26):3431–3443, 2011.

[6] R. J. Guyan. Reduction of stiffness and mass matrices. *AIAA Journal*, 3(2):380–380, 1965.

[7] N. Herrmann, C. Niklasch, D. Kiefer, L. Gerlach, Y. Le Pape, and S. Fortier. PACE 1450: An Experimental Test Setup for the Investigation of the Crack Behaviour of Prestressed Concrete Containment Walls. In *16th International Conference on Nuclear Engineering*, volume 4, pages 1–8, Orlando, Florida, USA, May 2008.

[8] L. Jason, S. Ghavamian, and A. Courtois. Truss vs solid modeling of tendons in prestressed concrete structures: Consequences on mechanical capacity of a Representative Structural Volume. *Engineering Structures*, 32(6):1779–1790, 2010.

[9] A. Llau, L. Jason, F. Dufour, and J. Baroth. Adaptive zooming method for the analysis of large structures with localized nonlinearities. *Finite Elements in Analysis and Design*, 106:73–84, Nov. 2015.

[10] J. Mazars. A description of micro and macroscale damage of concrete structures. *Engineering Fracture Mechanics*, 25:729–737, 1986.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELING AND SIMULATION FOR AN OPTIMIZED DESIGN OF A DYNAMIC BENDING TEST

Rana Akiki[1], Fabrice Gatuingt[1], Cédric Giry[1], Nicolas Schmitt[1,2] and Lavinia Stéfan[3]

[1]LMT-Cachan
ENS Cachan/CNRS/Université Paris-Saclay
61, avenue du Président Wilson,
94235 Cachan Cedex, France
e-mail: {akiki,gatuingt,giry}@lmt.ens-cachan.fr

[2]Université Paris Est Créteil
61, avenue du Général de Gaulle
94000 Créteil, France
e-mail: nicolas.schmitt@u-pec.fr

[3]AREVA NC, D&S, Technical Department
1, Place Jean Millier
92084 Paris la Défense, France
e-mail: lavinia.stefan@areva.com

Keywords: dynamic bending test, quasi-brittle material, experiment design, parametric study, material properties.

Abstract. *Performance prediction of fibre-reinforced concrete structures under impact is of major importance in a wide range of industrial applications. Therefore, the study of the dynamic behaviour of the material is necessary to quantify the dynamic mechanical properties of the material in terms of both mechanical response and fracture behaviour : dynamic fracture energy and tensile strength. This study aims to design a modified Hopkinson bar bending test with a series of numerical simulations prior to the actual test in order to select a feasible experiment configuration. Results of a dynamic bending experiment on a steel fibre-reinforced concrete specimen loaded up to failure are presented.*

1 INTRODUCTION

Performance prediction of fibre-reinforced concrete structures under impact is of major importance in a wide range of industrial applications. Therefore, the study of material dynamic behaviour is necessary to quantify its dynamic mechanical properties in terms of both the mechanical and fracture behaviours. Tested under high strain rates, cement-based materials tend to be strain-rate sensitive, which causes the mechanical behaviour to be significantly different from what is observed under quasistatic conditions [1, 2]. To cite a few, tensile strength and fracture energy are both loading rate-dependent material properties.
Investigations of the behaviour of materials under high loading rates have resulted in new experimental techniques. The Hopkinson pressure bar was first introduced by Bertram Hopkinson as a way to measure the elastic wave propagation in a metal bar [3]. This technique was then refined by Kolsky [4] to measure stress and strain, by using two Hopkinson bars in series, and is now known as the split Hopkinson pressure bar (SHPB). The first tests investigated the dynamic compressive behaviour of materials.
Since then, many experiments have been performed for the characterization of the tensile strength over a wide range of strain rates, whereas only a few data are available for the fracture energy. The aim of this work is to provide a method for analyzing the Hopkinson bar bending test prior to the experimental procedure via numerical simulations. Such method allows us to examine the effects of many parameters (geometry and loading conditions) that experimentally, would be costly and time-consuming. First, different configurations for the dynamic bending test, including aspects such as loading methods, sample configurations and impactor speed are investigated. Second, based on the numerical simulations, the final configuration of the dynamic bending test is chosen. Using the Hopkison bars apparatus, the experiment is then conducted on a fibre-reinforced concrete specimen. Two high-speed cameras are used in order to capture the entire event from the onset of loading, through crack initiation and propagation. Full displacement fields are obtained with digital image correlation done on a number of relevant images. First results are presented and analyzed.

2 DESCRIPTION OF THE EXPERIMENTAL METHOD

2.1 Definition of the test configuration

The bending test has been employed in Hopkinson bar dynamic fracture testing in several configurations (Fig. 1). Loading methods are varied: with only one bar for a one-point bending test (one-bar/1PB) [5], or for a three-point bending test using one bar (one-bar/3PB) [6], two bars (two-bar/3PB) [7] or three bars (three-bar/3PB) [8, 9]. The three-bar/3PB loading configuration includes the impactor, the incident bar and two transmission bars which are the supports of the beam tested. The specimen is placed between the incident and transmission bars and the striker is propelled towards the incident bar. A stress wave travels then through the incident bar. Upon reaching the sample, due to the mismatch of mechanical impedances between the bar material and that of the specimen, part of the incident stress wave is reflected back into the incident bar, and the rest is transmitted through the sample to the transmission bars. The incident, reflected, and transmitted stress waves are then analyzed to deduce the sample's mechanical response by using one-dimensional wave propagation theory; considering that the bars

are sufficiently long.

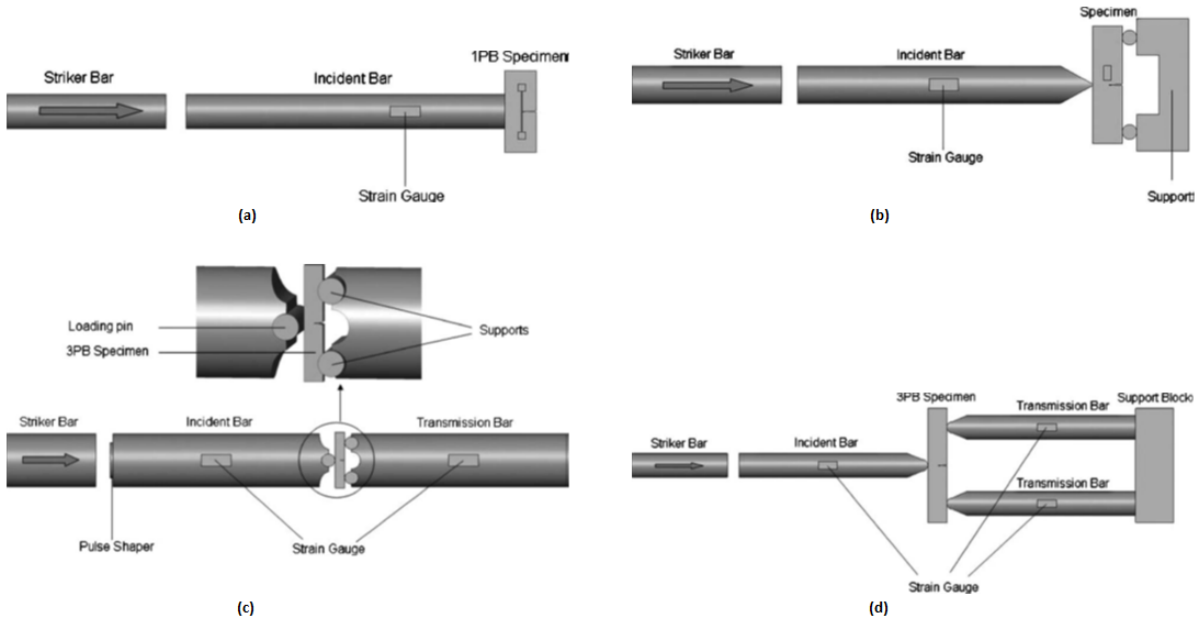


Figure 1: Different configurations of dynamic bending tests: (a) one-bar/1PB (b) one-bar/3PB (c) two-bar/3PB (d) three-bar/3PB [10].

2.2 Data reduction

Wave speed: The accurate measure of the wave speed in the bars is the first step in a valid data reduction for the Hopkinson bar experiment. This is done with a bar-to-bar test. By recording the time-of-arrival of both the incident and reflected waves at the incident bar strain gauge (t_1 and t_2 respectively), and knowing the length between the strain gauge and the end of the bar(L_{Ig}), the bar wave speed can be accurately calculated with equation 1:

$$c_b = \frac{2L_{Ig}}{t_2 - t_1} \tag{1}$$

The same is done between the signals measured from the incident gauge and the transmitted one, knowing the distance separating them. For the aluminium bars used in the lab's experimental apparatus, the mean value of the wave celerity is then computed: we find it to be equal to 5200 m/s.

Dispersion correction and time shifting: The waves during the test are measured using four strain gauges configured in a full Wheatstone bridge circuit. The strain gauges are used in pairs: the first pair is connected along the axis of the bar; one strain gauge is mounted axially and the other radially; the second pair of strain gauges is connected in the same manner but diametrically opposed. The electrical signals measured with the strain gauge bridges are converted to strain according to equation 2:

$$\varepsilon = \frac{V_{out}}{V_{in} \times f_a \times GF \times (1 + \nu_{alu})} \tag{2}$$

where, V_{out} is the electric potential measured, V_{in} is the electric potential applied across the strain gauge, f_a the amplification factor, ν_{alu} the Poisson's ratio of the aluminium and GF the gauge factor.
The time shifting of the waves consists of setting an origin matching the arrival of the incoming wave at the bar-specimen interface. Theoretically, it is estimated knowing the wave celerity and the distances between the gauges and the ends of the bars. Transport in space also requires taking into account the effect of dispersion, since waves with different lengths propagate in a medium at different speeds. This effect is observable in the Hopkinson bars, due to the three-dimensionality of the bar. Taking into account the dispersion correction, the measured waves are transported to the input bar-specimen interface using the software DAVID [11].

Sample boundary velocity: A measure of interest not typically reported, but can be readily calculated, is the bar end velocity. The velocity of the incident bar end v_i can be calculated using Equation 3:

$$v_i = c_b \times (\varepsilon_i - \varepsilon_r) \tag{3}$$

where, ε_i and ε_r are respectively the incident and reflected strain measures taken from the incident bar strain gauge and time shifted as exposed in the preceding paragraph.

Input force: The input force at the interface between the input bar end and the specimen F_i can be calculated using Equation 4:

$$F_i = E_b \times S_b \times (\varepsilon_i - \varepsilon_r) \tag{4}$$

where, E_b and S_b are the Young's modulus and the section of the aluminium bar, respectively.

3 MODELING OF THE HOPKINSON BAR EXPERIMENT

3.1 Numerical approach

A three dimensionnel finite element model of the Split-Hopkinson Pressure Bar (SHPB) test of a pre-notched prismatic beam is developed. The dimensions (diameter and length) of the striker and the incident bar are fixed (Fig. 2). Different dimensions of the beam are studied: 40×40×160, 60×60×240 and 100×100×400 (in millimeters). The numerical analysis is conducted using the explicit dynamic finite element code LS-DYNA [12].

Element type and mesh: Under-integrated eight-node solid hexahedron elements are selected for modeling the specimen and the bars. The specimen is described using fine mesh with finite elements having an aspect ratio of one in order to capture well the non-linear phenomena. The part of the experimental setting staying elastic are described using a coarser mesh with an aspect ratio larger than one. The summary of FE distribution in different portions of the FE model is given in table 1 and the finite element mesh in the specimen and the input bar end are shown in figure 3.

Material model: The selection of the appropriate material model is a critical step in the analysis since it determines the response of the structure to the type of loading it is subjected to. Experimental evidence [1, 2, 13, 14, 15] indicates that the material, which will be studied in the

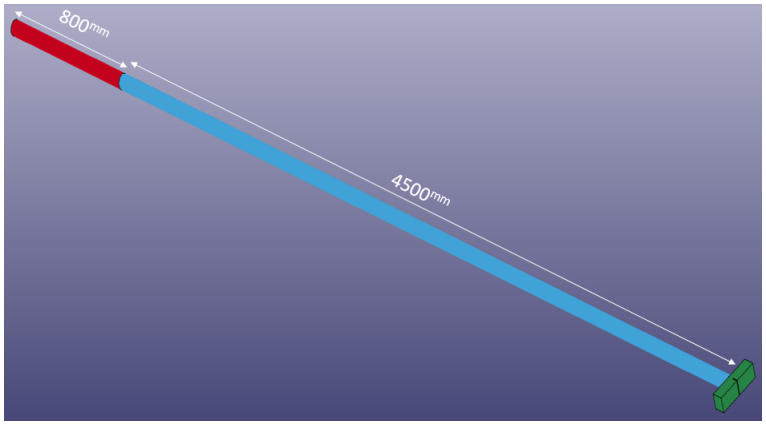


Figure 2: Model of the dynamic three-point bending test.

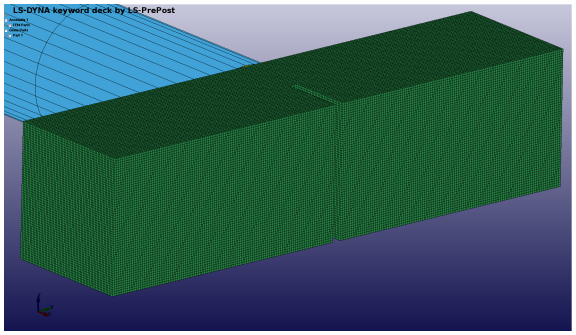


Figure 3: Mesh of the specimen and the input bar end.

Entity	Nodes	Elements
Striker	4329	3680
Input Bar	22126	20700
Specimen	115816	107100

Table 1: Mesh details of the components in the FE model.

future, is quasi-brittle and exhibits sensitivity to strain-rate. Therefore, an elasto-plastic damage model with rate effects, namely the continuous surface cap model (MAT-159) is chosen [16]. It takes into account strain softening, modulus reduction due to damage as well as rate effects. The striker and bar materials show a yield strength higher than the stress state expected during the test. Consequently, they are modeled as linear elastic materials.

Contact definition: The contact algorithm uses the penalty method, which allows penetration to occur but penalizes it by applying surface contact force models. Automatic surface to surface contact type category is used for the simulations, being adequate for the modified Hopkinson bar experiment which does not involve penetration of bodies.

Application of constraints: The axial direction of the striker and incident bar being along the X-direction of global cartesian coordinate system, the specimen displacement is constrained

in this direction to achieve a three-point bending configuration. As for the one-point bending configuration, no constraints are applied, and the specimen is in a free-free condition.

3.2 Results and discussion

The response of the beam under different loadings (i.e., different initial striker velocities) are studied and three specimen sizes are considered. The global energies of the FEM are checked for negligeable hourglass energies. The numerical simulations of the three-point bending test show that damage initiates before an equilibrium state occurs (Fig. 4). Smaller spans have also been tested, but the equilibrium was nevertheless still not reached. The supports in the three-point bending test configuration do not come into play initially which puts the beam in a free-free condition. Thus, cracking occurs before supports are engaged. The specimen's deformation is caused by striker loading and material inertia only. Since the mechanical state

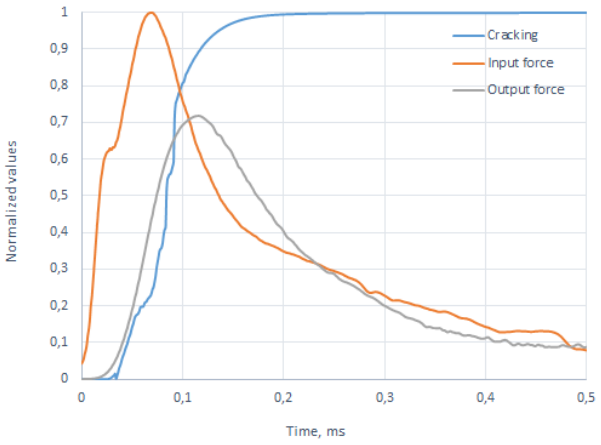


Figure 4: Input, output forces and cracking initiation with respect to time.

of the specimen is identical to the one of an unsupported beam, it is suitable to use a one-point bending test configuration. The force-displacement evolutions at the load point of the beam are plotted in figure 5 for various dimensions of the specimen. One can observe that the experimental device needed (bearing for specimen) and the response are influenced by the size of the specimen. As a matter of fact, the specimen is practically held with no supports which makes it hard to be kept in place with increasing weights. Thus, the specimen has to be kept relatively small. Depending on the striker velocities, the specimen breaks more or less early. As shown in figure 6, with higher velocities, the input force is higher for a given displacement as well as the maximum displacement. Also, damage and therefore cracking of the specimen is reached earlier for higher initial velocities (Fig. 7). Finally, the configuration chosen for the test is the one-point bending with one input bar and no output bars. As for the initial velocity of the striker, the range reachable in the LMT facilities is sufficient to lead to the failure of the specimen, whatever its size may be.

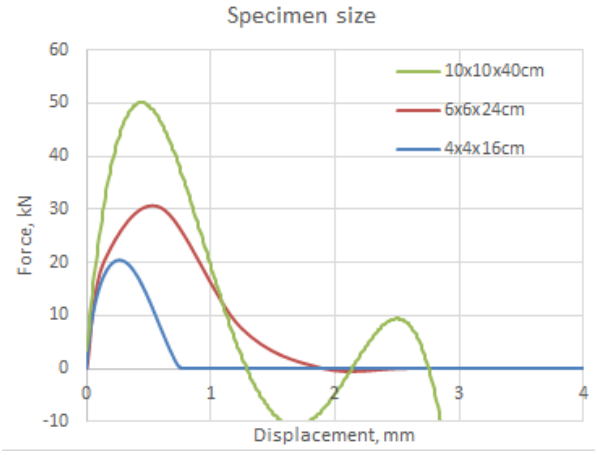


Figure 5: Force-displacement curves for different specimen sizes.

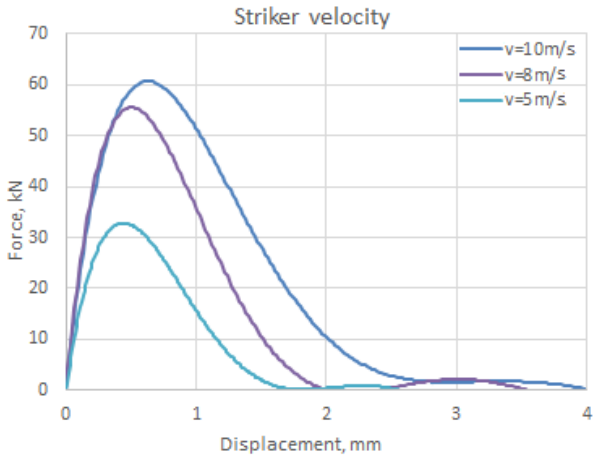


Figure 6: Effect on the striker velocity on the force-displacement curves.

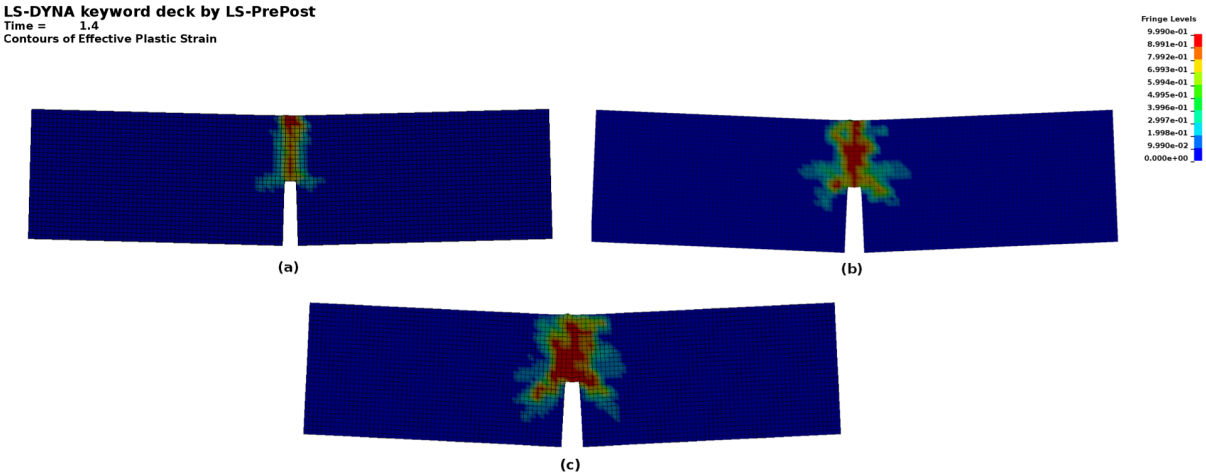


Figure 7: Damage in the specimen at t=1.4 ms for striker velocities of (a) 5 m/s (b) 8 m/s and (c) 10 m/s.

3.3 The experiment

3.3.1 Description of the apparatus and the measurement techniques

The apparatus consists mainly of a gas gun, a striker, an incident bar, and a computer-controlled data acquisition system. The incident bar is 4.5 m long, while the projectile used is 0.8 m long. Both the bar and the projectile are made of aluminum alloy and have a common diameter of 60 mm. The striker, input bar and specimen are positionned as mentionned in the previous section. The material tested in this work is a steel fibre-reinforced concrete with nominal specimen dimensions of $4\times4\times16$ cm and a 2-cm deep pre-notch. Figure 8 shows the final experimental configuration adopted for the dynamic one-point bending test and the measurement techniques used.

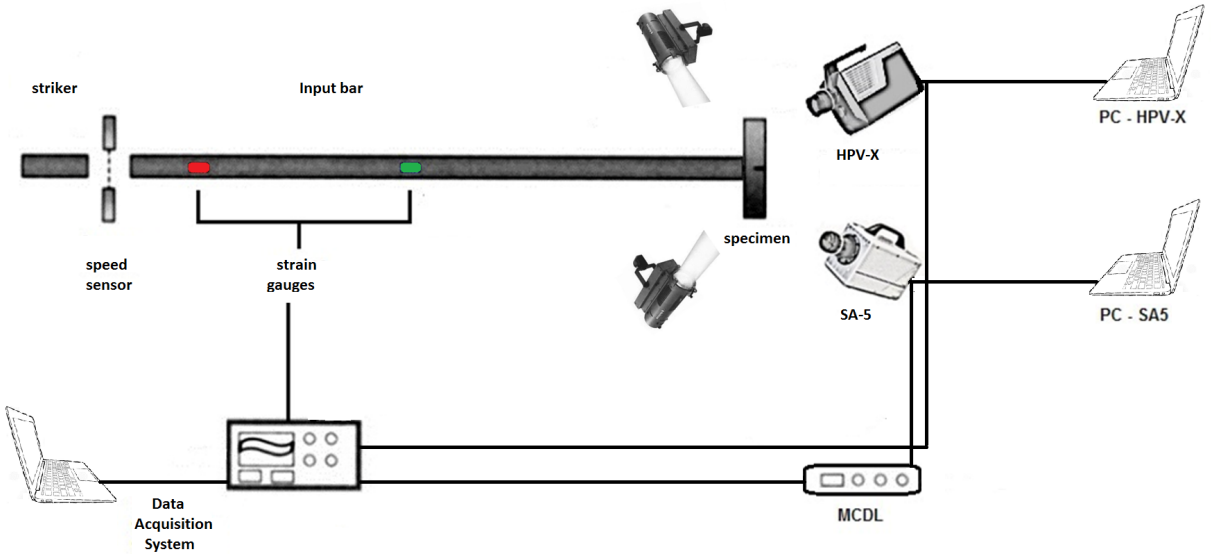


Figure 8: One-point bending test configuration.

Two strain gauge brigdes mounted at two locations on the incident bar provide strain measurements. The first one is used to trigger the camera's system and start recording, whereas the second one is used to determine the time histories of the stress and strain at the bar-specimen interface. Two high-speed cameras are used: the Photron Fastcam SA-5 and the Shimadzu HPV-X. The Photron Fastcam SA-5 camera is capable of recording images with a maximum resolution depending on the frame rate. It is used to capture the entire event from the onset of loading, up to crack initiation and propagation. To do it, the camera is run at a nominal frame rate of 50,000 frames per second resulting in a recording throughout the total time of the test with a resolution of 256×472 pixels. For the Shimadzu HPV-X, the maximum frequency can reach up to 10 million frames per second for a 50,000 pixels resolution with a memory capacity capable of recording up to 256 images in total. Thus, this camera has recorded images of the area near the crack tip in order to obtain precise information about the crack initiation and propagation throughout the duration of the loading. The camera is run at 200,000 frames per second resulting in a total recording of 1280 μs with a resolution of 250×400 pixels. The images are used to perform digital image correlation with the Correli Q4 code developed in

the LMT Cachan [17]. The artificial random pattern, obtained by spraying black spots on the surface of the beam, and the positionning of high-speed cameras are taken into account in order to measure a reliable displacement field.

3.3.2 Experimental results

Photos taken from the Fastcam SA5 follow the evolution of the crack at different times as shown in figure 9. Time reference corresponds to the triggering of the camera (i.e., arrival time of the wave at the location of the first gauge placed on the input bar).

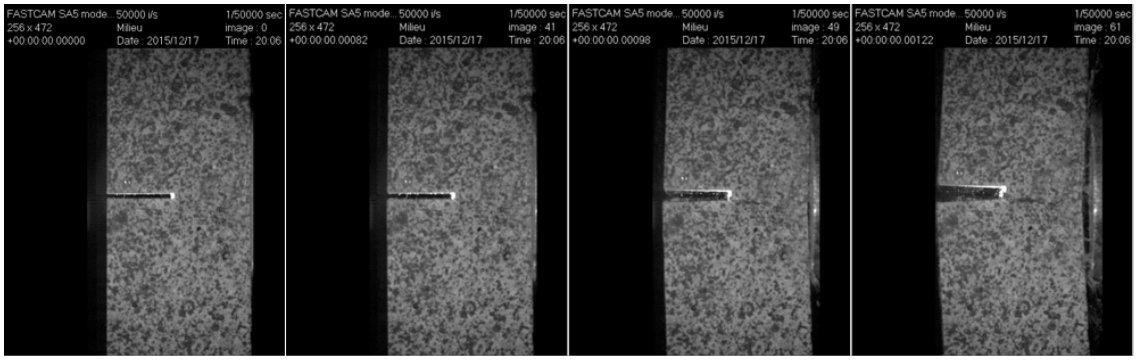


Figure 9: High speed photographs of specimen subjected to impact loading.

The incident and reflected strain waves obtained experimentally are presented in figure 10 after being corrected for dispersion and time shifting. The velocity at the interface between the input bar and the specimen is calculated from the measured waves and is shown in figure 11.

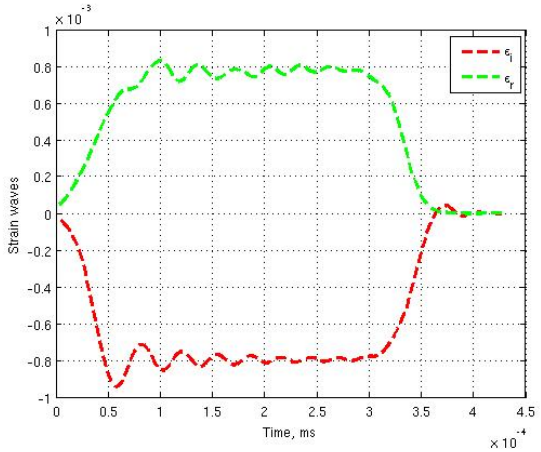


Figure 10: Incident (ε_i) and reflected (ε_r) strain waves after being shifted to the specimen's face.

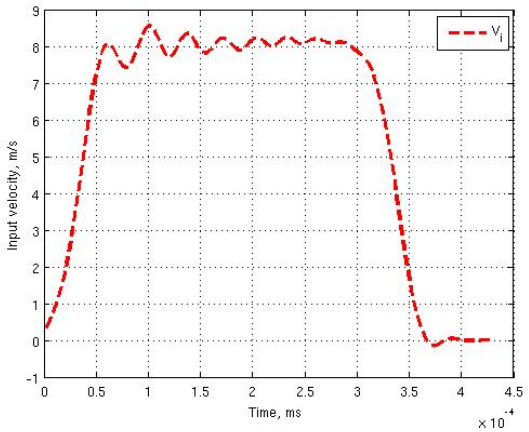


Figure 11: Velocity (V_i) at the interface input bar-specimen.

The input force calculated at the interface between the input bar and the specimen is shown in figure 12.

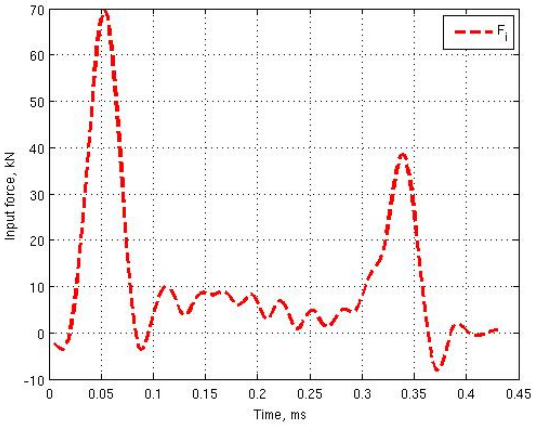


Figure 12: Evolution of input force (F_i) in time.

The digital image correlation gives access to the strains and displacement fields as shown in figure 13. The initiation of the crack is easily detected even when it is not discernible to the naked eye (Fig.9b and Fig.13b). It can be quantified via the discontinuity of the Y-displacement (Fig.13b, c, d) by calculating the displacement jump and it is found to be equal to 0.32 mm, 1.91 mm and 3.36 mm at the 3 respective times.

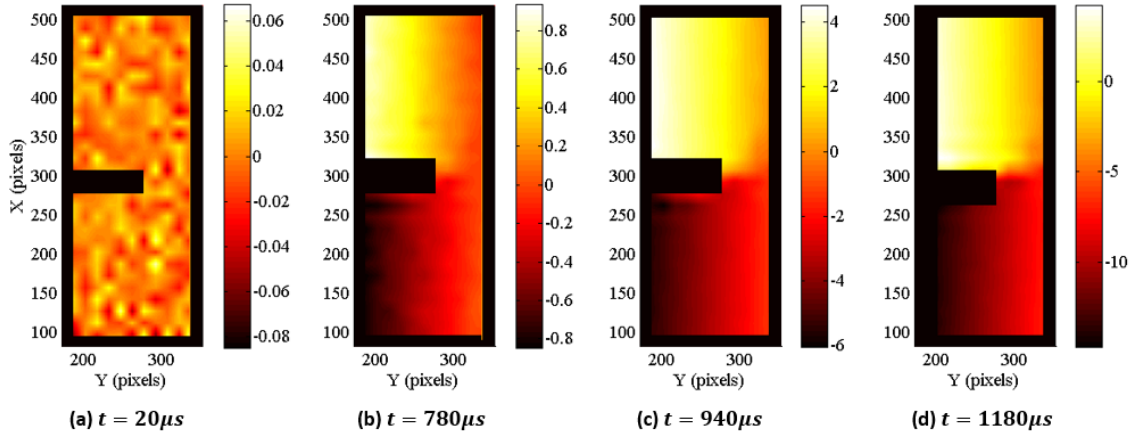


Figure 13: Displacement fields in the Y-direction measured with digital image correlation at different moments of the test.

4 CONCLUSIONS

This paper details the design of a dynamic bending test, through prior numerical analysis. The numerical simulations show that in the three-point bending test configuration, cracking occurs before the supports are engaged. Since the mechanical state of the specimen is identical to

the one of an unsupported beam, the latter can be used for the analysis of quasi-brittle materials for which failure occurs at the onset of the test. After having chosen the suitable parameters for the test configuration, the experiment is performed. First results of a test in which a steel fibre-reinforced concrete specimen loaded up to failure are presented and show the cracking initiation and propagation. A future iteration of this experimental procedure might attempt to have a more localized force application. Additional work should also be put into an accurate computation of a fracture energy for the dynamic case where no equilibrium state is obtained.

REFERENCES

[1] T. S. Lok and P. J. Zhao, "Impact Response of Steel Fiber-Reinforced Concrete Using a Split Hopkinson Pressure Bar," *Journal of Materials in Civil Engineering*, vol. 16, no. 1, pp. 54–59, 2004.

[2] J. I. Daniel, V. S. Gopalaratnam, and M. A. Galinat, "State-of-the-Art Report on Fiber Reinforced Concrete," *ACI Committe 544*, vol. 96, no. Reapproved, 2002.

[3] B. Hopkinson, "A Method of Measuring the Pressure Produced in the Detonation of High Explosives or by the Impact of Bullets," *Royal Society of London*, vol. 86, pp. 437–456, 1914.

[4] H. Kolsky, "An Investigation of the Mechanical Properties of Materials at very High Rates of Loading," *Proc. Roy. Phys. Soc.*, pp. 676–700, 1949.

[5] G. Weisbrod and D. Rittel, "A Method for Dynamic Fracture Toughness Determination Using Short Beams," *International Journal of Fracture*, vol. 104, pp. 89–103, 2000.

[6] C. Bacon, J. Färm, and J.-L. Lataillade, "Dynamic Fracture Toughness Determined from Load-point Displacement," *Experimental Mechanics*, pp. 217–223, 1993.

[7] F. Jiang, A. Rohatgi, K. Vecchio, and J. Cheney, "Analysis of the Dynamic Responses for a Pre-Cracked Three-Point Bend Specimen," *International journal of Fracture*, vol. 127, pp. 147–165, 2004.

[8] T. Yokoyama and K. Kishida, "A Novel Impact Three-Point Bend Test Method for Determining Dynamic Fracture-Initiation Toughness," *Experimental Mechanics*, vol. 29, no. 2, pp. 188–194, 1989.

[9] L. Rubio, J. Fernandez-Saez, and C. Navarro, "Determination of Dynamic Fracture-initiation in a Modified Hopkinson Pressure Bar," *Experimental Mechanics*, vol. 43, no. 4, pp. 379–386, 2003.

[10] F. Jiang and K. S. Vecchio, "Hopkinson Bar Loaded Fracture Experimental Technique: A Critical Review of Dynamic Fracture Toughness Tests," *Applied Mechanics Reviews*, vol. 62, no. 6, 2009.

[11] G. Gary, "DAVID - Instruction Manual," tech. rep., LMS, Ecole Polytechnique, Palaiseau (France), 2005.

[12] LSTC, *LS-DYNA Theory Manual*. 2015.

[13] A. E. Naaman and V. S. Gopalaratnam, "Impact Properties of Steel Fibre Reinforced Concrete in Bending," *The International Journal of Cement Composites and Lightweight Concrete*, vol. 5, no. 4, pp. 225–233, 1983.

[14] N. Banthia, K. Chokri, Y. Ohama, and S. Mindess, "Fiber-reinforced Cement Based Composites Under Tensile Impact," *Advanced Cement Based Materials*, vol. 1, no. 3, pp. 131–141, 1994.

[15] S. Wang, M.-H. Zhang, and S. T. Quek, "Mechanical Behavior of Fiber-Reinforced High-Strength Concrete Subjected to High Strain-Rate Compressive Loading," *Construction and Building Materials*, vol. 31, pp. 1–11, 2012.

[16] Y. Wu, J. E. Crawford, and J. M. Magallanes, "Concrete Constitutive Models," in *12th International LS-DYNA Users conference*, no. 1, pp. 1–14, 2012.

[17] G. Besnard, F. Hild, and S. Roux, ""Finite-element" displacement fields analysis from digital images: Application to Portevin-Le Châtelier bands," *Experimental Mechanics*, vol. 46, pp. 789–803, 2006.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON STABILITY LOSS OF A THIN-WALLED SPHERICAL SHELL SUBJECTED TO EXTERNAL PRESSURE AND INTERNAL HOMOGENEOUS CORROSION

Emmanuel M. Gutman[1], Rudolf M. Bergman[2], and Semyon P. Levitsky[2]

[1] Department of Material Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
e-mail: gutman@bgu.ac.il

[2]Mathematics Unit, Shamoon College of Engineering, 56 Bialik St., Beer-Sheva 84100, Israel
e-mail: rudolf@sce.ac.il; levits@sce.ac.il

Keywords: Spherical shell, Stability loss, Homogenous corrosion, Life-time.

Abstract. *Stability loss of a thin-walled elastic closed spherical shell, subjected to external pressure and internal corrosion, is studied. It is shown that critical time of stability loss of the shell can be established if the upper critical load value for static stability loss of the shell without corrosion and the corrosion rate law are known. Numerical results are obtained for carbon steel shells with different wall thickness at different temperatures.*

1 INTRODUCTION

Thin-walled elastic vessels, having the form of a closed sphere, are widely used in the industry, for example, in chemical engineering, as high-pressure reservoirs. In this relation the problem of stability loss of a thin closed spherical shell, subjected to a uniform external pressure, is of special importance. It was studied in a number of papers, particularly, in [1-4], where stability loss of thin-walled elastic shells, subjected to various loads, including external pressure, was investigated.

However, existing models don't take into account the fact that in the process of structure operation the shell interaction with an active environment is possible. Such interactions can change, for instance, the thickness of the shell's wall - it is known [5] that corrosion leads to sufficiently fast thinning of different elements of metal constructions. Experimental data indicate that corrosion rate depends also on the magnitude of mechanical stresses in the material. This effect must be taken into account when the service life of the structural elements operating in the presence of general corrosion is estimated [5].

The onset of breaking is related usually to approaching of certain limiting stress in the shell. Nevertheless, in certain cases of practical importance, the stability loss can occur in the shell before the breaking stress is reached. In such cases the structure life time will be completely determined by the criterion of the stability loss. For instance, the method of the life time estimation for elastic thin-walled closed circular cylindrical shell, subjected to simultaneous actions of uniform external pressure or longitudinal compressive forces, and homogeneous corrosion, was developed in [6, 7].

In the present paper this method is extended for the case of stability loss of a thin-walled elastic closed spherical shell, subjected to simultaneous action of a uniform external pressure and internal homogeneous (general) corrosion.

2 PROBLEM DESCRIPTION

The corrosion is considered here as a uniform dissolution of the internal surface of the shell (general corrosion). This suggestion leads to a model in which the internal wall of the shell is dissolved uniformly over the entire surface, and its thickness h decreases with time t. The rate of this process is equal to the corrosion rate. Experimental data for various metals and steels, received for pure elastic stressed state, have shown that corrosion rate, besides the temperature T, depends essentially from the stressed state of the material, which in the case of a spherical shell is characterized by compressive stress σ, caused by the external uniform pressure. As a result, the corrosion rate v follows the relation $v = f[\sigma(t), T]$. It was shown in [5] that this dependence can be approximated by the exponent function. Similar to [6, 7], it is possible to write this law in an explicit form, taking into account specific conditions of corrosion and using experimentally established parameters.

Below stability loss of a thin-walled elastic closed spherical shell, subjected to simultaneous action of uniform external pressure and internal uniform corrosion, is studied. The formulas for the critical time (live-time), corresponding to the moment of stability loss, are derived.

3 SHELL STABILITY LOSS IN THE ABSENCE OF CORROSION

Consider the problem of stability loss for a thin elastic closed spherical shell with thickness $h(t)$ and radius of the middle surface $R(t)$, subjected to external pressure of intensity Q. Assuming that changes of the wall thickness and, consequently, the radius $R(t)$, are quasi-static, the equations of static stability loss in linear approximation can be applied. Spherical coordinates of the shell middle surface are characterized by the radius $R(t)$ and the angles ϑ and ϕ

(Fig.1). It is assumed that the initial stress state of the shell is membrane like [8] and, therefore, from equations of the thin elastic shells theory it follows that the loadings, arising from the external pressure with intensity Q, in all normal cross-sections of the shell have the form [2, 4]:

$$S = S_1 = S_2 = -\frac{QR}{2} \tag{1}$$

Assuming that the loading S and the corresponding stress σ are compressing, we obtain from Eq. (1):

$$S = \frac{QR}{2}; \quad \sigma = \frac{QR}{2h} \tag{2}$$

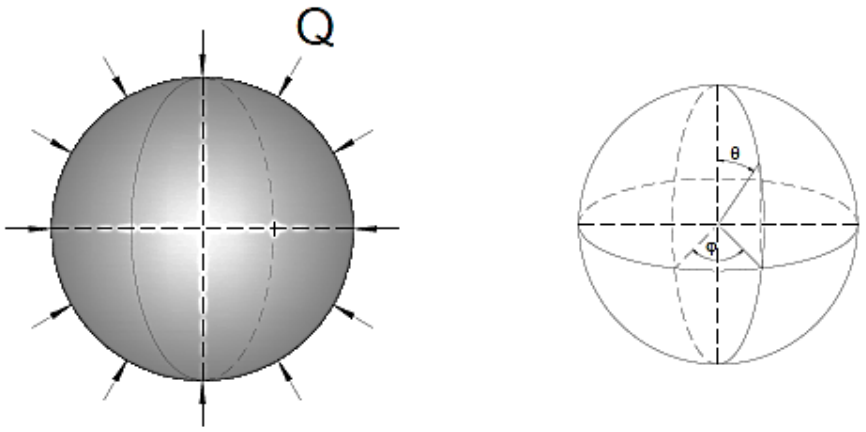


Figure 1: Spherical coordinate system and scheme of the shell's loading. Q - intensity of external loading, φ - polar angle, θ - spherical angle.

At solving of the linear problem for stability loss of elastic closed spherical shell we assume that on the middle surface of the shell many dents are formed. According to [2, 4, 8], the changes in the middle shell surface within one dent can be considered as minor. Taking into account Eq. (2) and using equations of the stability problem for thin elastic shallow spherical shells, we obtain the following homogenous differential equation of stability loss for a closed spherical shell, subjected to external uniform pressure of intensity Q [2, 4]:

$$\frac{Eh^2}{12(1-\nu^2)}\nabla^6 w + \sigma\nabla^4 w + \frac{E}{R^2}\nabla^2 w = 0 \tag{3}$$

Here E is the Young module, ν is the Poisson coefficient, and w is the normal deflection of the middle surface of the shell. Differential operators ∇^2, ∇^4 and ∇^6 are defined as follows:

$$\nabla^2 F(\theta,\phi) = \frac{1}{R^2}\left(\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial\phi^2}\right)F, \quad \nabla^4 F = \frac{1}{R^4}\left(\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial\phi^2}\right)^2 F, \quad \nabla^6 F = \frac{1}{R^6}\left(\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial\phi^2}\right)^3 F \tag{4}$$

where the function $F = F(\theta,\phi)$ is supposed to have partial derivatives of a corresponding order.

Following [9], we accept that solution of Eq. (3) must satisfy the relation:

$$\nabla^2 w = -\lambda^2 w \tag{5}$$

where λ is an indefinite parameter. Then we receive from Eq. (3):

$$\frac{E h^2}{12(1-\nu^2)} \lambda^4 - \sigma \lambda^2 + \frac{E}{R^2} = 0$$

which yields the relation:

$$\sigma = \frac{E h^2}{12(1-\nu^2)} \lambda^2 + \frac{E}{R^2 \lambda^2} \tag{6}$$

The value of λ, corresponding to minimum of σ, is equal to:

$$\lambda^2 = \frac{\sqrt{12(1-\nu^2)}}{R h} \tag{7}$$

Substituting Eq. (7) into Eq. (6), we obtain the upper critical stress σ^* and the corresponding upper critical pressure Q^*:

$$\sigma^* = \frac{E h}{R\sqrt{3(1-\nu^2)}}, \quad Q^* = \frac{2E h^2}{R^2\sqrt{3(1-\nu^2)}} \tag{8}$$

Note that the upper critical stress σ^* for a spherical shell, defined by Eq. (8), coincides with that one for a thin elastic circular cylindrical shell with the radius R of the middle surface, subjected to longitudinal compressive forces and external general corrosion [7].

Eq. (8) was derived at the assumption that the shell wall thickness is specified, whereas the critical value of the external pressure Q is the unknown quantity. It follows from Eq. (8) that if to assume, on the contrary, that the external pressure Q value is specified, then the shell wall thickness h^*, corresponding to stability loss, will be equal to

$$h^* = \frac{\sqrt[4]{3(1-\nu^2)} \cdot \sqrt{Q} \cdot R}{\sqrt{2E}} \tag{9}$$

4 STABILITY LOSS OF THE SPHERICAL SHELL IN THE PRESENCE OF HOMOGENEOUS INTERNAL CORROSION

It is assumed below that the shell is subjected to simultaneous action of a constant external pressure Q and homogeneous internal corrosion. The rate of the thickness decrease at each moment of time t is equal to the corrosion rate

$$\frac{dh}{dt} = -f(\sigma, T) \tag{10}$$

where $f(\sigma, T)$ is a sufficiently smooth function of the compressive stress σ, defined by the Eq. (2), and the temperature T.

The corrosion rate dependence on the stress value, according to experimental data, can be approximated by exponential function suggested in [5], while the temperature effect is de-

scribed usually by the Arrhenius type law [10]. Combination of both functions yields the following relation:

$$f(\sigma, T) = v_0 \exp[\bar{E}_{c0}(1-\bar{E}_c/\bar{T}) + V\sigma/(R_g T)],\ \bar{E}_{c0} = E_{c0}/R_g T_0,\ \bar{E}_c = E_c/E_{c0},\ \bar{T} = T/T_0 \quad (11)$$

Here R_g is the molar gas constant, V is the material molar volume, E_c, E_{c0} - the effective activation energy of the corrosion process and its reference value, respectively; T, T_0 - the absolute temperature and its reference value. The preexponential factor v_0 in (11) has the meaning of corrosion rate for a non-stressed shell ($\sigma = 0$) at $E_c = E_{c0}$ and $T = T_0$. Similar to [6,7], the critical time t^*, corresponding to the moment of the shell stability loss, can be obtained from the Eqs. (10), (11):

$$t^* = (h_0\sigma_0/v_0)\exp[-\bar{E}_{c0}(1-\bar{E}_c/\bar{T})]\cdot\int_{\sigma_0}^{\sigma^*}\sigma^{-2}\exp[-V\sigma/(R_g T)]\,\mathrm{d}\sigma \quad (12)$$

where

$$h_0 = h\big|_{t=0},\quad \sigma^* = \frac{Eh_0}{R\sqrt{3(1-\nu^2)}},\quad \sigma_0 = \frac{QR}{2h_0} \quad (13)$$

Here it is accounted for that $R \approx const$. Note that in Eqs. (12), (13) is supposed also that the external pressure Q for a given initial wall thickness h_0 and shell's parameters satisfies the inequality:

$$\sigma_0 \le \sigma^* \quad (14)$$

If we introduce the dimensionless variables according to

$$D^* = (t^* v_0)/R,\quad \tau = \sigma/\sigma^*,\quad N^* = \sigma_0/\sigma^*,\quad \varepsilon = h_0/R \quad (15)$$

then Eq. (12) takes the form

$$D^* = \varepsilon N^* \exp[-\bar{E}_{c0}(1-\bar{E}_c/\bar{T})]\int_{N^*}^{1}\tau^{-2}\exp[(-V\tau\sigma^*)/(R_g T)]\,\mathrm{d}\tau \quad (16)$$

Here, similar to [6, 7], the value of D^* can be treated as a relative durability of the shell, and value of N^* - as a safety coefficient for the stability.

5 NUMERICAL RESULTS

Numerical simulations were performed on the basis of Eq. (16) in order to study the relative durability D^* dependency on the safety coefficient for stability N^* for a thin–walled closed spherical shell, made from carbon steel, at different values of temperature and initial wall-thickness. The changes of D^* with N^* and ε are illustrated by the plots on the Fig. 2. Data for the Young modulus E temperature dependency for carbon steel (carbon content < 0.3%) were taken from the engineering toolbox [11] according to ASME B31.1-1995:

$$1)\ E = 2.049\times10^{11}\ \mathrm{N/m^2}\quad (-20^0\,\mathrm{C});\quad 2)\ E = 1.971\times10^{11}\ \mathrm{N/m^2}\quad (100^0\,\mathrm{C})$$

Carbon steel for the both temperature values is characterized by $\nu = 0.295$, $V = 7\times10^{-6}\ \mathrm{m^3/mol}$; the molar gas constant is equal to $R_g = 8.317\ \mathrm{N\cdot m/(mol\cdot K)}$. The effective activation energy for corrosion was chosen as $E_c = E_{c0} = 50\ \mathrm{kJ/mol}$; the reference temperature $T_0 = 293\mathrm{K}$.

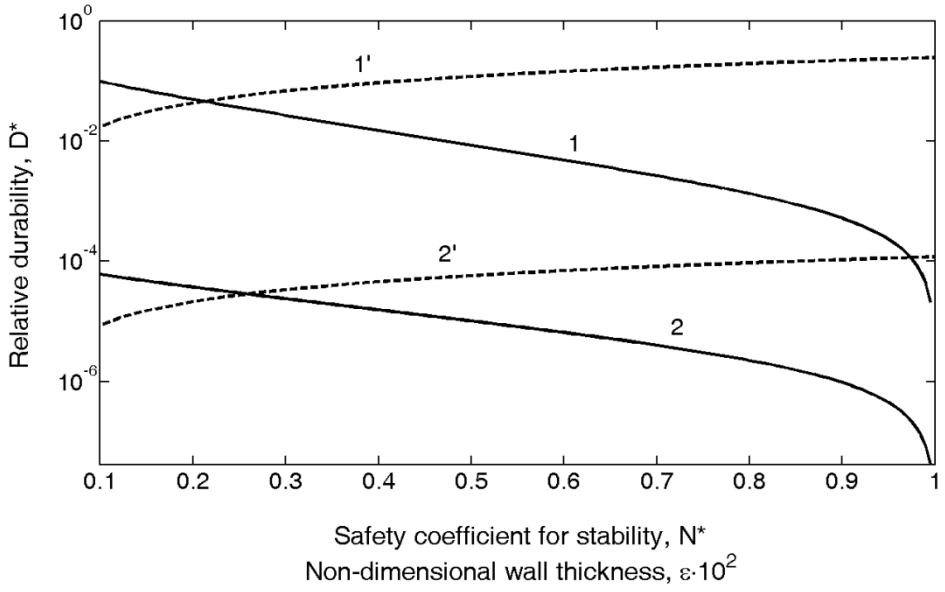


Figure 2: Dependency of the relative durability D^* from the safety coefficient for stability N^* (plots 1, 2) and the non-dimensional wall thickness ε (plots 1', 2') at different temperatures.

The curves 1, 1' on the Fig. 2 correspond to $T = -20^0\,C$, the curves 2, 2' - to $T = 100^0\,C$. The plots 1, 2 were calculated with a fixed initial non-dimensional wall-thickness $\varepsilon = 0.01$; the plots 1', 2' - with a fixed external load $Q = 5\cdot10^4\ \mathrm{N/m^2}$. It follows from the Fig. 2 that the relative durability decreases with the growth of the safety coefficient for stability, while the increase of the initial shell thickness has an opposite effect on the vessel life-time. The temperature increase yields the relative durability reduction because of two reasons – the corrosion enhancement and the elasticity modulus decrease. Simulations have shown that the first one is the most important factor.

6 CONCLUSIONS

- The critical time of stability loss of a thin-walled elastic closed spherical shell, subjected to uniform internal corrosion at a permanent uniform external pressure, can be predicted from the solution of the corresponding problem of static stability for the same shell, with account for the shell's wall thickness changes due to homogeneous corrosion, accelerated by mechanical stresses in the metal.
- The relative durability of the shell depends on the safety coefficient for stability; its value is highly influenced by the temperature and the initial shell thickness.
- It was shown that the relative durability decreases with the growth of the safety coefficient for stability.

- The temperature increase yields the relative durability reduction because of two reasons – the corrosion enhancement and the elasticity modulus decrease.
- Thicker shells at the same external load and temperature are characterized by larger life-time.

REFERENCES

[1] R.H. Donnel, *Beams, plates and shells*, McGraw-Hill, 1976.

[2] A.S. Volmir, *Stability of deformable systems*, Fizmatlit, 1967 (in Russian).

[3] P.E. Tovstik, A.L. Smirnov, *Asymptotic methods in the buckling theory of elastic shells*, World Scientific, 2001.

[4] E.I. Grigoljuk, V.V. Kabanov, *Stability of shells*, Fizmatlit, 1967 (in Russian).

[5] E.M. Gutman, *Mechanochemistry of solid surfaces*, World Scientific, 1994.

[6] E. Gutman, J. Haddad, R. Bergman, Stability of thin-walled high-pressure vessels subjected to uniform corrosion. *Thin-Walled Structures,* **38,** 43-52, 2000.

[7] R.M. Bergman, S.P. Levitsky, J. Haddad, E.M. Gutman, Stability loss of thin-walled cylindrical tubes subjected to longitudinal compressive forces and external corrosion. *Thin-Walled Structures*, **44**, 726-729, 2006.

[8] A.L. Goldenveizer, *Theory of thin elastic shells*, Pergamon Press, 1961.

[9] V.Z. Vlasov, Basic differential equations of the general theory of elastic shells, *National Advisory Committee for Aeronautics, Technical Memorandum* 1241, NASA, Washington, 1961.

[10] A. Popova, E. Sokolova, S. Raicheva, M. Christov, AC and DC study of the temperature effect on mild steel corrosion in acid media in the presence of benzimidazole derivatives, *Corrosion Science* **45,** 33-58, 2003.

[11] The Engineering Toolbox, www.EngineeringToolBox.com.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EFFICIENCY OF DAMAGE-PLASTICITY MODELS IN CAPTURING COMPACTION-EXPANSION TRANSITION OF CONCRETE UNDER DIFFERENT COMPRESSION LOADING CONDITIONS

Reza Mousavi[1], Masoud D. Champiri[2], and Kaspar J. Willam[3]

[1]University of Houston, Department of Civil Engineering
Houston , TX
e-mail: mmousavi@uh.edu

[2] University of Houston, Department of Civil Engineering
Houston , TX
e-mail: mdehghanichampiri@uh.edu

[3] University of Houston, Department of Civil Engineering
Houston , TX
e-mail: kwillam@central.uh.edu

Keywords: concrete, damage, plasticity, dilation, multiaxial behavior

Abstract. *The pre- and post-peak performance features of two widely used damage-plasticity constitutive formulations are assessed for modeling the multiaxial response behavior of pressure sensitive quasi-brittle materials such as concrete. Two benchmark problems have been solved using commercial software packages; ABAQUS which implements the two invariant Lee-Fenves damaged plasticity formulation as its main feature for concrete has been used along with LS-DYNA which is using Duvaut-Lion damage and plasticity model, a three invariant formulation. In order to have consistent results, the parameter identification and calibration for both models has been performed and response of both models has been compared to each other under uniaxial tension and compression. The two benchmark problem are triaxial compression CTC and triaxial tension CTE. The aim of this study is to observe the difference between two well-known material models in capturing the compaction to extension regime in the yielded material. The difference between the response of each material model and also the efficiency of input parameters in controlling of this transition has been studied in this paper. Based on these failure diagnostics, the main shortcomings of each formulation are discussed and possible enhancements and remedies are proposed for the pre- and post-peak feature.*

1 introduction

The usage of commercial packages is world wide and they have many applications in the he computational mechanics. Between these packages, for sure, ABAQUS [1] and LS-DYNA [2] have more interest than others. Since different material models are developed in these packages for same material like concrete, it is necessary to compare them together and see how the user can get similar results with implementing some parameters. This can help them to use material models widely and in different cases and situations.
Many concrete material models have been developed and implemented in these packages. Damage and plasticity are two important features where users are looking for them in order to use concrete material models. In ABAQUS, concrete damaged plasticity model (CDP) [3, 4], is widely used among researchers [5–8]. The constitutive formulation of this material model is reviewed in the following sections. Many concrete material models are available in LS-DYNA where they consider damage and plasticity [9]. Karagozian and Case model (concrete damage model, MAT-72) [10], the Riedel-Hiermaier-Thoma model (RHT, MAT-272) [11, 12], the Brannon-Fossum model (BF1) [13], and Continuous surface cap model (CSCM, MAT-159) [14] are some of those. Among them, MAT-159 has many applications in the literature [15, 16], because this material uses two parameters including uniaxial compressive strength and maximum size of aggregate in the simplest version for concrete structures, while other materials need many inputs. This material model uses the Duvaut-Lions formulation [17] and its constitutive model will be addressed in the section 3 of this paper. As a result, these two material models in ABAQUS and LS-DYNA have been selected and their behavior is compared together. It is tried to find a way to make their behavior closer to each other with implementing a parameter study. For this, their behavior is compared in uniaxial compression test, uniaxial tension test. Later on, with adopted parameters, the behavior of these material models is investigated in case of triaxial compression test (CTC), triaxial extension test (CTE) and bi-axial compression test. These tests can be also done experimentally as Kupfer et.al [18]. In order to compare the with the experimental results, one option to observe a more general localized behavior of concrete under different levels of confinement, full field measurment techniques like digital image correlation (DIC) method can be utilized [19–22]. A remarkable feature that can be pointed out about post processing of digital image correlation system is their feasibility in being implemented in expert systems. Champiri et al [23, 24]developed some expert systems for concrete structures and structural health monitoring with developing a fuzzy method. The parameter study is conducted in section 4 and the most efficient parameters are selected for performing different tests. Also it is noted that each material model has some deficiencies as well as some advantages in some points. Some concluding remarks are investigated in section 5.

2 Concrete Damaged Plasticity model

In this model, the admissible stress state has the condition of [4]:

$$F(\sigma, f_t, f_c) \leq 0 \tag{1}$$

In which, $f_t = f_t(\kappa_t), \quad f_c = f_c(\kappa_c)$ are tensile and comressive uniaxial strengths which are written as functions of damage parameter respectively. By rearranging uniaxial strength values as function of scalar degradation, the relation between the effective strength and strength is as follow:

$$f_t = (1 - D_t(\kappa_t))\bar{f}_t(\kappa_t), \quad f_c = (1 - D_c(\kappa_c))\bar{f}_c(\kappa_c) \tag{2}$$

The damage variable in CDP model is defined as scalar quantity which is different in compression and tension [3]. the uniaxial version of this parameter is defined as follow:

$$\kappa_\chi = \frac{\int_0^{\varepsilon^p} \sigma_\chi(\varepsilon^p) d\varepsilon^p}{\int_0^{\infty} \sigma_\chi(\varepsilon^p) d\varepsilon^p} \tag{3}$$

Here, $g_\chi = \int_0^{\infty} \sigma_\chi(\varepsilon^p) d\varepsilon^p$ is called dissipated energy density per unit volume. This quantity has been obtained by regularization of fracture energy in tensile and compressive state (G_χ) with respect to characteristic element size (l_χ) as $g_\chi = G_\chi / l_\chi$. the relation between stress an strain is written as follow:

$$\sigma_\chi = f_{\chi 0}((1 + a_\chi)e^{-b_\chi \varepsilon^p} - a_\chi e^{-2b_\chi \varepsilon^p}) \tag{4}$$

In which, a_χ, b_χ are constants. The relation between degradation and plastic strain is written as:

$$1 - D_\chi = e^{-d_\chi \varepsilon^p} \tag{5}$$

By taking into account degradation, the effective stress is calculated as follow:

$$\bar{\sigma}_\chi = f_{\chi 0}((1 + a_\chi)(e^{-b\varepsilon^p})^{1 - \frac{d_\chi}{b_\chi}} - a_\chi (e^{-b_\chi \varepsilon^p})^{2 - \frac{d_\chi}{b_\chi}}) \tag{6}$$

By combining eq.(4),(6) with eq.(5), we have:

$$g_\chi = \frac{f_{\chi 0}}{b_\chi}\Big(1 + \frac{a_\chi}{2}\Big) \tag{7}$$

$$\sigma_\chi = \frac{f_{\chi 0}}{a_\chi}\Big((1 + a_\chi)\sqrt{\phi_\chi(\kappa_\chi)} - \phi_\chi(\kappa_\chi)\Big) \tag{8}$$

By eq. (8) and (3), the damage evolution law in the uniaxial test is obtained by:

$$\dot{\kappa}_\chi = \frac{1}{g_\chi} f_\chi(\kappa_\chi)\dot{\varepsilon}^p \tag{9}$$

For the case of compressive stress, a damage value of κ_c' has been assumed when compressive stress is at its peak which gives corresponding degradation value of D_c as an input parameter for the material model which will be discussed in following. By putting this damage value into eq. ref$eq_stress2$, we will have following:

$$\sqrt{\phi(\kappa_c')} = \frac{1 + a_c}{2} \rightarrow D_c = 1 - \Big(\frac{1}{a_c}\Big(1 + a_c - \frac{1 + a_c}{2}\Big)\Big)^{\frac{d_c}{b_c}} \tag{10}$$

Since D_c is measured from experiment, we can obtain d_c/b_c by equation:

$$\frac{d_c}{b_c} = \frac{\log(1 - D_c)}{\log\left(\frac{1+a_c}{2a_c}\right)} \tag{11}$$

In the case of tensile loading, since the stress diagram is descending from the beginning of plastification, a tensile degradation value D_t will be considered at the point where stress value is half of the value of tensile strength $(\sigma_t = f_{t0}/2))$. for this case, we have:

$$\sqrt{\phi(\kappa_t)} = \frac{1 + a_t + \sqrt{1 + a_t^2}}{2} \tag{12}$$

$$\frac{d_t}{b_t} = \frac{\log(1 - D_t)}{\log((1 + a_t) - \sqrt{1 + a_t^2}) - \log 2a_t} \tag{13}$$

3 Material 159

The yield function for this material model reads as [14]:

$$\phi_{159}(I_1, J_2, J_3, \kappa) = J_2 - \Re^2 F_f^2 F_c \tag{14}$$

in which,$F_f(I_1), F_c(I_1, \kappa), \Re(J_2, J_3)$ are functions which are taking into shear surface, cap region and Rubin scaling factor [25], respectively.

$$F_f(I_1) = \alpha - \lambda e^{-\beta_0 I_1} + \theta I_1 \tag{15}$$

$$F_c(I_1, \kappa) = \begin{cases} 1 - \frac{(I_1-\kappa)^2}{(X-\kappa)^2} & : I_1 \geq \kappa_0 \\ 1 & : otherwise \end{cases} \tag{16}$$

$$\Re(J_2, J_3) = \frac{-b_1 + \sqrt{b_1^2 - 4b_1 b_0}}{2b_2} \tag{17}$$

$$L(\kappa) = \begin{cases} \kappa & : \kappa > \kappa_0 \\ \kappa_0 & : \kappa \leq \kappa_0 \end{cases} \tag{18}$$

$$X(\kappa) = L(\kappa) + R_0 F_f(L(\kappa)) \tag{19}$$

By Definition:

$$b_0 = -\frac{3 + b - a^2}{4} \tag{20a}$$

$$b_1 = a(\cos\beta - a\sin\beta) \tag{20b}$$

$$b_2 = (\cos\beta - a\sin\beta)^2 + b\sin^2\beta \tag{20c}$$

$$b = (2Q_1 + a)^2 - 3$$

$$a = \frac{-a_1 + \sqrt{a_1^2 - 4a_2 a_0}}{2a_2} \tag{20d}$$

$$a_0 = 2Q_1^2(Q_2 - 1) \tag{20e}$$

$$a_1 = \sqrt{3}Q_2 + 2Q_1(Q_2 - 1) \tag{20f}$$

$$a_2 = Q_2 \tag{20g}$$

$$\sin 3\hat{\beta} = \frac{3\sqrt{3}J_3}{2J_2^{3/2}} \tag{20h}$$

Q_1 and Q_2 are the parameters which determine the yield surface parameters that we want to fit. As a generalized form:

$$Q_1 = \alpha_1 - \lambda_1 e^{-\beta_1 I_1} + \theta_1 I_1 \tag{21a}$$

$$Q_2 = \alpha_2 - \lambda_2 e^{-\beta_2 I_1} + \theta_2 I_1 \tag{21b}$$

In formulation for material 159, the softening is modeled using damage formulation for tensile and in the case of compression with moderate confinement. It means that the softening is not activated unless damage formulation is activated. Unlike CDP model, in which, damage parameter had indirect effect on softening behavior, in material 159, softening behavior and stiffness degradation (modulus reduction) are in direct correlation with each other. The scalar damage parameter is categorized into two different categories: brittle damage, which happens

when stress state is tensile and ductile damage, in which, the stress state is compressive. The ductile and brittle damage evolution functions, are as follow:

$$d(\tau_b) = \frac{0.999}{D}\Big(\frac{1+D}{1+De^{-C(\tau_b - r_{0b})}} - 1\Big) \tag{22}$$

$$d(\tau_d) = \frac{d_{max}}{B}\Big(\frac{1+B}{1+Be^{-A(\tau_d - r_{0d})}} - 1\Big) \tag{23}$$

In which, d is scalar damage parameter, τ_b, τ_d are strain energy corresponding to current strain state and r_{0b}, r_{0d} are initial damage threshold for brittle and ductile damage,respectively.
$A,\ B,\ C,\ D$ are softening parameters for compression (A, B) and tension (C, D), respectively. d_{max} is maximum damage parameter which can be defined. (recommended to take $d_{max} < 1$ and is usually taken $d_{max} > 0.99$).
The fracture energy is calculated from the following:

$$G_f = \int_{u_0}^{\infty} (1-d) f' dx \tag{24}$$

in which, u_0 is displacement corresponding to the peak strength value.
By putting eq. (22) into eq. (24), relationship between fracture energy and softening parameters for brittle damage is calculated as follow:

$$G_f = r_{0b} L\Big(\frac{1+D}{CD}\Big) \tag{25}$$

based on eq. (25), a parameter reduction can be done by eliminating parameter C from the brittle damage evolution eq. (22):

$$C = r_{0b} L\Big(\frac{1+D}{G_f D}\Big)\log(1+D) \tag{26}$$

parameter reduction for ductile damage is obtained from following equation:

$$G_f = 2r_{0d} L\Big(\frac{1+B}{AB}\Big)\log(1+B) + 2L\Big(\frac{1+B}{A^2}\Big)\int_0^{\infty} \frac{ye^{-y}dy}{1+Be^{-y}} \tag{27}$$

In which, $y = -A(\sqrt{u} - \sqrt{u_0})\sqrt{\frac{f'}{L}}$. The integral for equation above doesn't have closed form solution. It is calculated numerically and the value is shown as $dilog$. By re-arranging eq. (27), the parameter A is obtained as following:

$$A = \frac{-b + \sqrt{b^2 - 4c}}{2} \tag{28}$$

$$b = \frac{-2(1+B)\log(1+B)Lr_{0d}}{BG_f} \tag{29}$$

$$c = \frac{-2L(1+B)dilog}{G_f} \tag{30}$$

Based on equations above after parameter reduction, the parameter that determines compressive behavior of concrete in the post peak regime, in addition to compressive fracture energy is B. Similarly, The parameter which shapes the post peak behavior of concrete under tensile loading condition, is D along with tensile fracture energy. The discussion about these parameters and values of each parameter are shown in the next section.

4 Parameter Evaluation for material models

Since finding analytical equivalence of each parameter is a very complicated task and needs a big amount of energy and time, here, the CDP parameters are calculated based on their numerical counter parts as used in material 159. The set of inputs for material are as shown in table 1. The set of parameters which are necessary to define a stress-strain curve and its compatible damage evolution law is mentioned in table 2.The material model has been assigned to a single cube element with displacement boundary conditions at one top side and symmetric boundary conditions in two perpendicular sides. the roller supports restraining bottom of the element from moving along load direction has been applied to the model. The analysis has been implemented in both packages in the case of uniaxial compression and tension and the results are compared in terms of stress vs strain, stress vs volumetric strain and damage evolution with respect to plastic strain as shown in figures 1 - 6

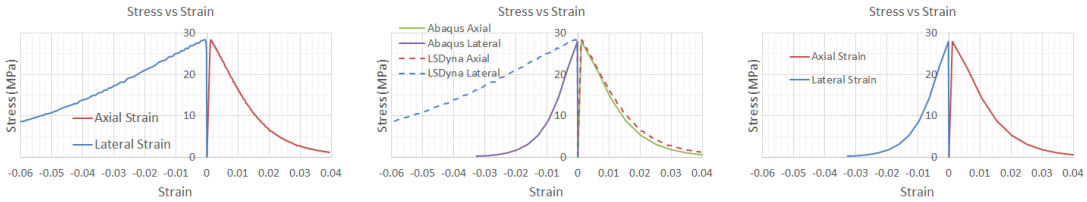


Figure 1: Uniaxial compression results of LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

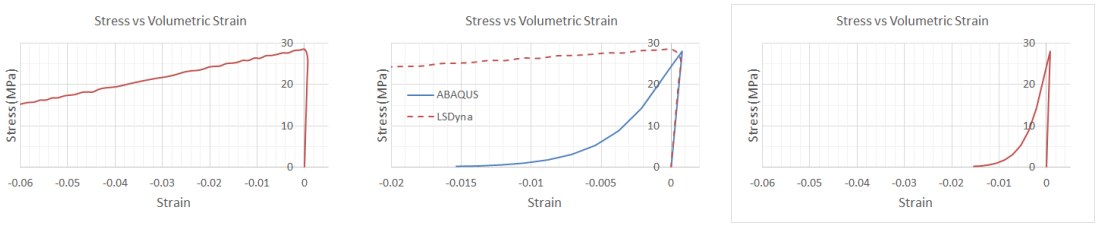


Figure 2: Compaction to extension behavior as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

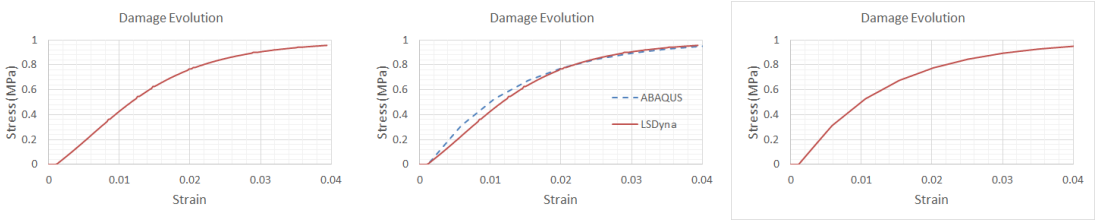


Figure 3: Compressive Damage evolution trend as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

The results of figs. 1-3 show that the models can be compatible with each other and give the same results in the case of uniaxial compression if the aim is to achieve identical stress vs. strain response in addition to observe same damage values and its corresponding plastic strain. However, it is clearly shown that lateral strains are not equal to each other in these two models. Because of using associate flow rule in material 159, it is expected to observe higher value of lateral expansion in this material which has been confirmed with the analysis results. The following figures are behavior of two material models under uniaxial tension:

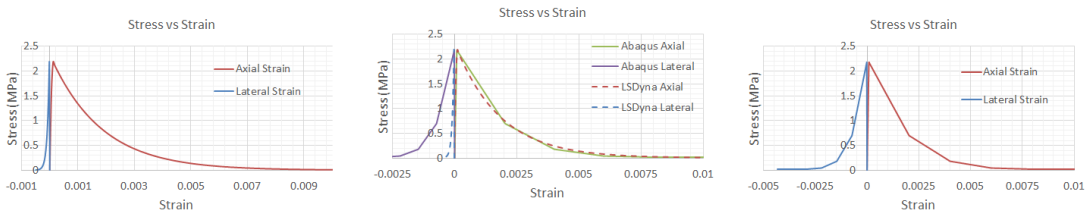


Figure 4: Uniaxial tension results of LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

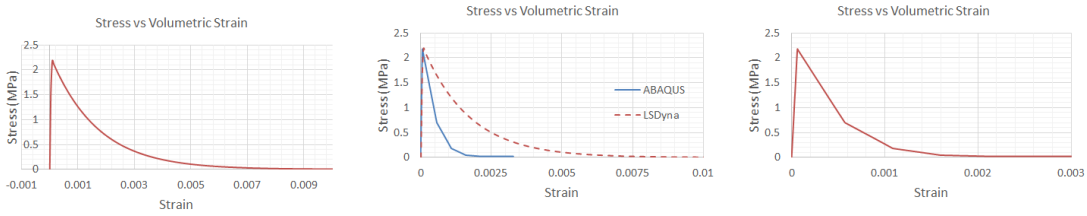


Figure 5: Extension to compaction behavior (there is not any!) as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

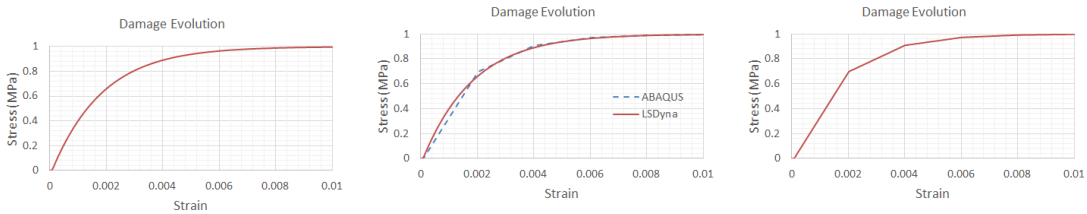


Figure 6: Tensile Damage evolution trend as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

After parameter matching of two material models in uniaxial tension and compression, some modelings have been done and comparisons are made for different load cases and the difference between response of material models has been discussed. The main focus of this section is on triaxial expansion vs triaxial compression behavior which are called CTE, CTC. The load path of CTE and CTC in $I_1 - J_2$ space is similar to each other but the direction is opposite ton each other. both are moving along the I_1 axis direction but CTE is moving towards most tensile point and CTC moves towards the most compressive point. The results of CTE has been shown in figures 7,8. Based on triaxial confined loading scenario, a tremendous increasing of

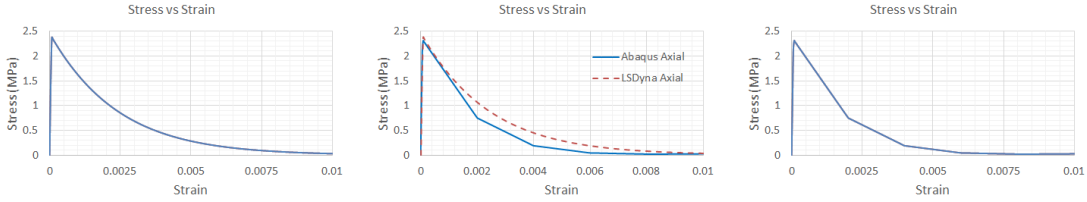


Figure 7: CTE results of LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

compressive strength has been observed in both models as show in fig. 9. In fact, according to expectation, no plastification has been observed in CDP model, which is based on an uncapped version of Drucker-Prager like behavior. Plastification has been observed in material 159 but like CDP, no damage was observed in the material.

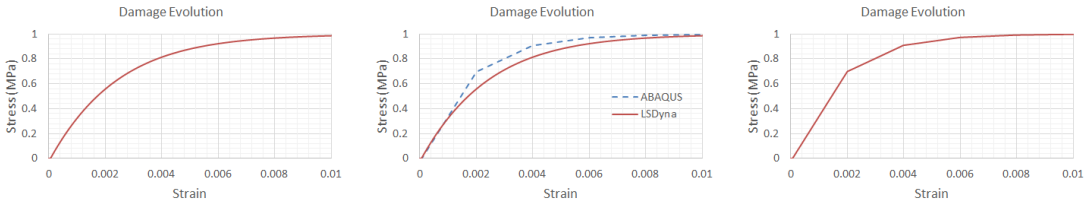


Figure 8: Tensile Damage evolution trend under CTE in LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

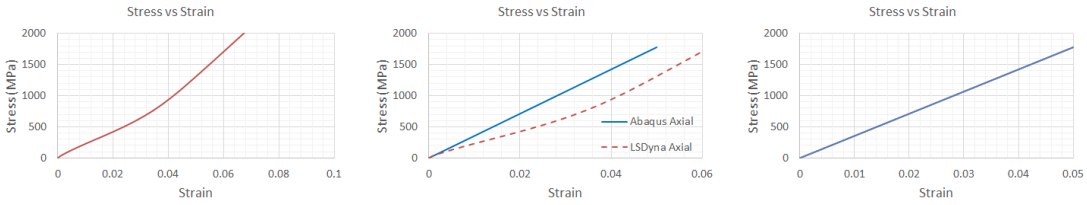


Figure 9: CTC results of LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

The results of compression under equibiaxial constant confining pressure for two material models have been obtained and compared in figures 10-12. As an example of higher level

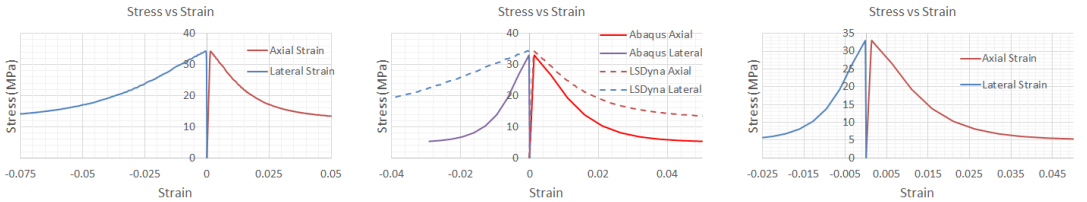


Figure 10: Results of low confinement pressure ($\sigma_{conf} = 1MPa$) with LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

of confinement, some modelings have been done when the model undergoes an equibiaxial confinement of $\frac{1}{2}f'_c$ prior to loading in he main axial direction. the results and comparisons are illustrated in figures 13-15. It can be observed from fig. 13 that the prediction of two models from biaxial behavior is different from each other while they are calibrated to each other in case of uniaxial loading condition. Since for both models, the shear is defined as an straight line, a parameter cannot be defined in order to match behavior of both models in three cases of loading (uni tension, uni compression and compression under low level of confinement). By looking at fig. 14, we can understand that even in the case of moderate confinements, a compaction to dilation behavior happens in material 159, while the behavior of CDP model is only compaction regime. It indicates the importance of considering non-associaticity in realistic modeling of material behavior. Being able to choose dilation angle, gives constitutive model the ability of performing more realistic and more complex modelings. It is also observed that in the higher levels of confinement, level of damage in models which associate damage with strength degradation is lower than the decoupled models due to lower amount of softening in material behavior.

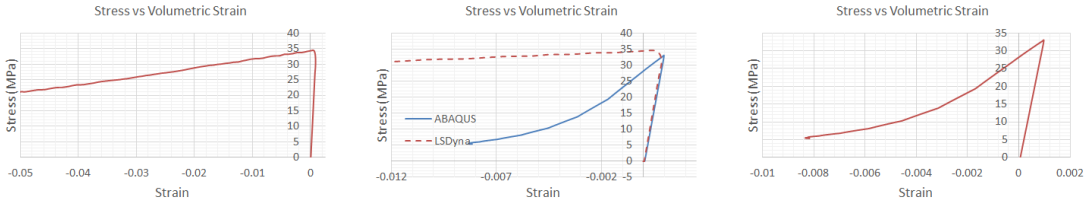


Figure 11: Compaction to Extension behavior (low confinement pressure $(\sigma_{conf} = 1MPa)$) as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

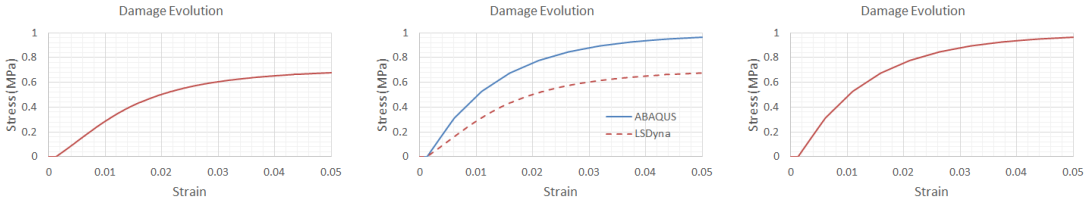


Figure 12: Compressive Damage evolution trend under low confinement pressure in LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

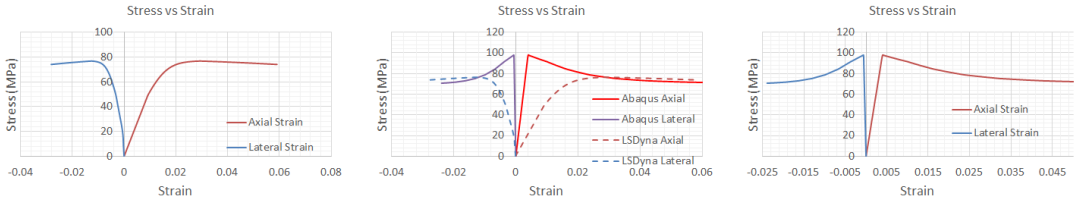


Figure 13: Stress vs. axial strain results for high confinement $(\frac{1}{2}f'_c)$ of LS-DYNA (Left), Abaqus (Right), and their stress vs strain comparison (Middle).

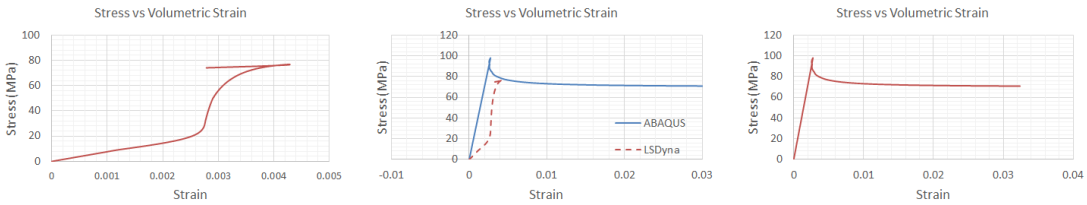


Figure 14: Compaction to Extension behavior (under high confinement $(\frac{1}{2}f'_c)$) as of LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

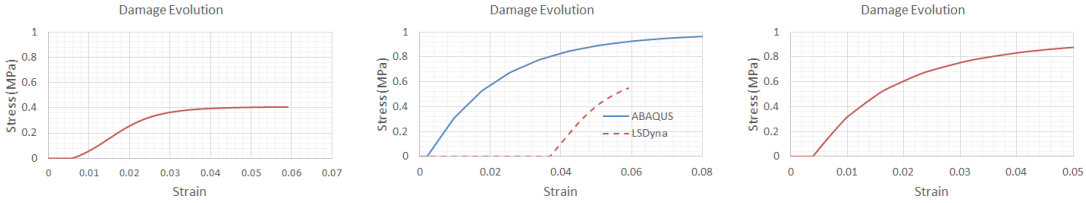


Figure 15: Compressive Damage evolution trend under high confinement $(\frac{1}{2}f'_c)$ in LS-DYNA (Left), Abaqus (Right) and their comparison (Middle).

Table 1: The material 159 parameters which are used to calibrate with CDP model

α	β_0	λ	θ	α_1	β_1	λ_1	θ_1
14.2	0.01929	10.51	0.2965	0.7473	7.25e-2	0.17	1.204e-3
α_2	β_2	λ_2	θ_2	f'_c	Aggregate size	G	K
0.66	7.25e-2	0.16	1.45e-3	28	16	1.15×10^4	1.26×10^4
R	X_0	D_1	D_2	W	G_c	G_T	G_s
5	90	2.5×10^{-4}	3.5×10^{-7}	0.05	10	0.1	0.1
B_1	B_2	B_3	B_4	D_1	D_2	D_3	D_4
500	100	50	10	0.05	0.1	1	10

Table 2: The CDP parameters

a_c	b_c	d_c	f'_c
0.8	110	75	28
a_t	b_t	d_t	f_t
0.4	700	600	2.2

5 Conclusions and Remarks

This paper has critically looked at two well known concrete material models (Lee-Fenves, Duvaut-Lion) which are implemented in two most widely used FEA software packages among researchers (ABAQUS, LS-DYNA, respectively) from a practical point of view to give general insights into modeling strategies to help reader choose a proper material model based on abilities and limitations of each model by emphasizing on difference between response of these material models under more complex loading paths, while their response under simple loading histories are identical to each other by choosing proper parameters. The models have been compared to each other initially under uniaxial compression and tension and after calibrating parameter values together, response of models in the cases of CTE, CTC and confined compression under low and high confining pressure has been obtained and compared to each other.

It has been observed that even in the case of uniaxial compression, there is a difference in dilatant behavior of models. The reason for that is CDP model uses a non-associate flow rule which can control the dilatation of response, while material 159 uses associate flow rule and it has been observe clearly that the volumetric strain is higher under associate flow condition which might not be a realistic response.

Under CTE condition, since both models are using a linear function for frictional area in $(I_1 - J_2)$ space, it is clear that none of the models are able to incorporate tiraxial extension strength as an independent material parameter. It can be solved by choosing a hyperbolic regime for shear regime that is asymptotic to linear curve in by moving towards compressive region in $(I_1 - J_2)$ space.

The behavior of two models in CTC conditions shows that proper cap behavior of CDP model can avoid linear elastic behavior in very high stress values. In addition, since for both material behaviors, the value of damage remains zero, which requires further analysis to find out how realistic is this value. It can be concluded from the confined load cases that under this condition, the peak value that both material models are giving has a small difference in addition to

dramatically different behavior in transition from compaction to expansion behavior in all cases of compressive loading.

REFERENCES

[1] K. Hibbitt and Sorensen, *ABAQUS/CAE User's Manual*. Hibbitt, Karlsson & Sorensen, Incorporated, 2002.

[2] L. Manual *Livermore Software Technology Corporation*, vol. 1, 2007. cited By 1.

[3] J. Lee and G. Fenves, "Plastic-damage model for cyclic loading of concrete structures," *Journal of Engineering Mechanics*, vol. 124, no. 8, pp. 892–900, 1998. cited By 739.

[4] J. Lubliner, J. Oliver, S. Oller, and E. Oate, "A plastic-damage model for concrete," *International Journal of Solids and Structures*, vol. 25, no. 3, pp. 299–326, 1989. cited By 729.

[5] T. Jankowiak and T. Lodygowski, "Identification of parameters of concrete damage plasticity constitutive model," *Foundations of civil and environmental engineering*, vol. 6, no. 1, pp. 53–69, 2005.

[6] S. V. Chaudhari and M. A. Chakrabarti, "Article: Modeling of concrete for nonlinear analysis using finite element code abaqus," *International Journal of Computer Applications*, vol. 44, pp. 14–18, April 2012. Full text available.

[7] M. R. Mousavi and A. Khaloo, "Effect of steel plate jacketing of columns in seismic behavior of concrete beam-column connections," in *6th National Congress on Civil Engineering*, 2011.

[8] G. Xotta, S. Beizaee, and K. Willam, "Localization analysis of coupled plasticity and damage models for dissipative materials," *Computational Modelling of Concrete Structures*, p. 439, 2014.

[9] R. M. Brannon and S. Leelavanichkul, "Survey of four damage models for concrete,"

[10] L. Malvar, J. Crawford, J. Wesevich, and D. Simons, "A plasticity concrete material model for dyna3d," *International Journal of Impact Engineering*, vol. 19, no. 9-10, pp. 847–873, 1997. cited By 245.

[11] H. Hansson and P. Skoglund, "Simulation of concrete penetration in 2d and 3d with the rht material model," *Swedish Defense Research Agency*, 2002.

[12] T. Borrvall and W. Riedel, "The rht concrete model in ls-dyna," in *Preceedings of the 8 th European LS-DYNA Users Conference, Strasbourg*, 2011.

[13] A. Fossum, R. Brannon, A. F. Fossum, and R. M. Brannon, "The sandia geomodel: theory and users guide," 2004.

[14] Y. D. Murray, "Users manual for ls-dyna concrete material model 159," tech. rep., 2007.

[15] M. D. Champiri, A. Attar, M. Hanifehzadeh, K. Willam, and B. Gencturk, "Long-term performance of dry storage structures," in *CONCREEP 10*, pp. 1593–1602.

[16] M. D. Champiri, A. Attar, M. Hanifehzadeh, K. Willam, and B. Gencturk, "Tipover simulation of degraded dry cask storage structures," in *Proceedings of the ASME 2015 International Mechanical Engineering Congress & Exposition -IMECE15*, ASME (Manuscript accepted for publication).

[17] J. Simo, J. Kennedy, and S. Govindjee, "Non-smooth multisurface plasticity and viscoplasticity. loading/unloading conditions and numerical algorithms," *International Journal for Numerical Methods in Engineering*, vol. 26, no. 10, pp. 2161–2185, 1988.

[18] H. Kupfer, H. K. Hilsdorf, and H. Rusch, "Behavior of concrete under biaxial stresses," in *Journal Proceedings*, vol. 66, pp. 656–666, 1969.

[19] G. Yang, M. Zomorodian, A. Belarbi, and A. Ayoub, "Uniaxial tensile stress-strain relationships of rc elements strengthened with frp sheets," *Journal of Composites for Construction*, p. 04015075, 2015.

[20] K. Willam, A. Mohammadipour, R. Mousavi, and A. S. Ayoub, "Failure of unreinforced masonry under compression," *Proceedings of the Structures Congress*, pp. 2949–2961, 2013.

[21] G. Xotta, S. Beizaee, and K. Willam, "Bifurcation investigations of coupled damage-plasticity models for concrete materials," *Computer Methods in Applied Mechanics and Engineering*, vol. 298, pp. 428–452, 2016.

[22] S. Beizaee, G. Xotta, and K. Willam, "Perforation studies on flat bars for xfem-based failure analysis," *Engineering Fracture Mechanics*, 2016.

[23] M. D. Champiri, S. H. Mousavizadegan, and F. Moodi, "A decision support system for diagnosis of distress cause and repair in marine concrete structures," *Computers & Concrete*, vol. 9, no. 2, pp. 99–118, 2012.

[24] M. D. Champiri, S. H. Mousavizadegan, and F. Moodi, "A fuzzy classification system for evaluating the health condition of marine concrete structures," *Journal of Advanced Concrete Technology*, vol. 10, no. 3, pp. 95–109, 2012.

[25] M. Rubin, "Simple, convenient isotropic failure surface," *Journal of engineering mechanics*, vol. 117, no. 2, pp. 348–369, 1991.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A SIMPLIFIED MODEL OF A STEEL COLUMN SUBJECTED TO IMPACT

Piseth Heng[1,2], **Mohammed Hjiaj**[1], **and Jean-Marc Battini**[2]

[1]INSA de Rennes, LGCGM/Structural Engineering Research Group, Université Européenne de Bretagne
20 avenue des Buttes de Coësmes, CS 70839, 35708 Rennes Cedex 7, France
e-mail: Piseth.Heng@insa-rennes.fr, Mohammed.Hjiaj@insa-rennes.fr

[2] Department of Civil and Architectural Engineering, KTH, Royal Institute of Technology
Department of Civil and Architectural Engineering, KTH, Royal Institute of Technology, SE-10044 Stockholm, Sweden
e-mail: Jean-Marc.Battini@byv.kth.se

Keywords: Impact; Steel Structures; Progressive Collapse; Generalized Plastic Hinges; Catenary Action; Energy-Conserving Scheme.

Abstract. *The impact response of structures becomes of great interest for recent research activities due to the concern for public safety regarding the accidents (e.g. a vehicle collision onto the building structure). In the event of impact, the structure can be expected to undergo large plastic deformation which is required to absorb kinetic energy generated from the moving vehicle. In this paper, the nonlinear response of a steel column subjected to transversal vehicle collision is studied. The deformation of the column is assumed to occur at critical locations where the generalized plastic hinges are formed. A simplified model which consists of rigid bars connected by elasto-plastic hinges is presented. To include the effect of the surrounding structure, an elastic spring and a mass are attached to the top end of the column where a constant compressive load is also applied. The elliptical yield surfaces in a stress resultant space that account for the interaction between bending moment and axial force are used. Regarding the impact load, the unilateral contact is adopted while friction is ignored. Because of the reason that the velocities and the contact force cannot be defined at impact point of time, the differential measures are written to combine the smooth and non-smooth parts. The set-value force laws including Newton's impact law are adopted to describe the contact condition. Concerning with the time integration, the classical midpoint rule scheme is employed to integrate the equations of motion and the flow rule of the hinges. The paper also presents numerical applications of the model to show its ability to describe the nonlinear dynamic behavior of a steel column under transversal vehicle collision.*

1 INTRODUCTION

Typical frame buildings, when subjected to extreme loading such as explosion or impact, may have high risks to experience progressive collapse. It is suggested that such progressive collapse is initiated by local damage of the supporting elements, i.e. columns. There are at least two directions to follow in order to prevent the progressive collapse. The first direction is to assume that one of the columns has failed due to the mentioned exceptional loading and thus removed. The response of the remaining structures are then investigated. This type of analysis can be found in [1, 2, 3, 4]. The other method is to focus on the response of the suffered column. Several publications as found in [5, 6, 7, 8] and the references cited therein have worked on the behaviour of reinforced concrete structures against extreme loading. However, due to the fact that concrete has low ductility compared to steel, the design concept that bases on stiff member behaviour brings about massive concrete columns. Yet, the design of steel building against impact loading is not clearly explained in current practical norms [9, 10, 11, 12]. With high ductility, steel members are allowed to have large deformation, which is beneficial to the response of the structure under impact. The structure with large displacement has the ability to absorb the kinetic energy generating from the moving vehicle. Some authors such as in [13, 14, 15] used an equivalent single degree of freedom system which is proposed by Biggs in [16] in order to study the dynamical response of the column subjected to extreme loading. Despite of the practical convenience, the method is not able to include membrane effect and the geometric nonlinearity. The influence of axial force on the plastic bending moment is also not included. This paper describes a simplified nonlinear model of a steel column subjected to a compression load and a transverse vehicle collision at any point along the column length for any support conditions. The inelastic behaviour of the column system is taken into account by employing the generalized plastic hinge method. With this model, the deformation of the column is assumed to concentrate at the generalized plastic hinges that form at critical locations. The column elements between the hinges are considered as rigid. The non-smooth problem of impact contains the discontinuity of the velocity. A set of measure - differential equations are used to describe the impact and impact - free motion. The numerical integration of the equations are solved by adopting a classical midpoint rule.

2 DESCRIPTION OF THE SIMPLIFIED MODEL

Described by Fig. 2, the model of the steel column consists of two rigid bars and four generalized elastic-plastic hinges. The top of the column is attached to a spring and applied by a compressive load. Illustrated in Fig. 1.a, the generalized plastic hinges which are modelled by a combined axial-rotational spring as shown in Fig. 1.b situate at both ends of the column and at the location where the force is applied. The material is assumed to be elastic perfectly plastic and the strain rate is ignored. Only axial and flexural deformations are considered whereas shear deformations are neglected. The elongation and shortening of the column are assumed to occur along the column.

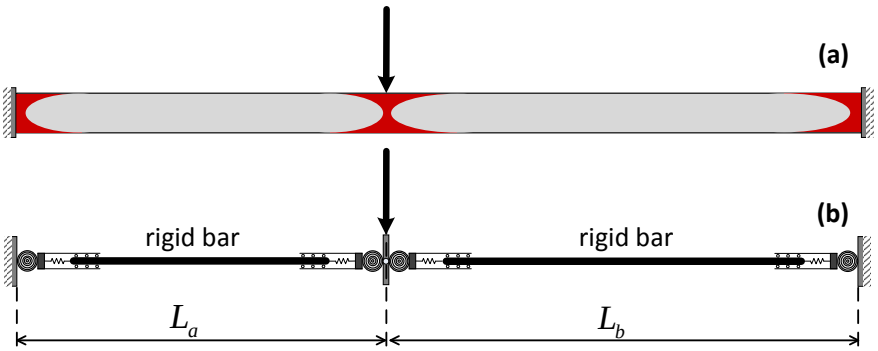


Figure 1: (a): plastic hinge formation. (b): simplified model.

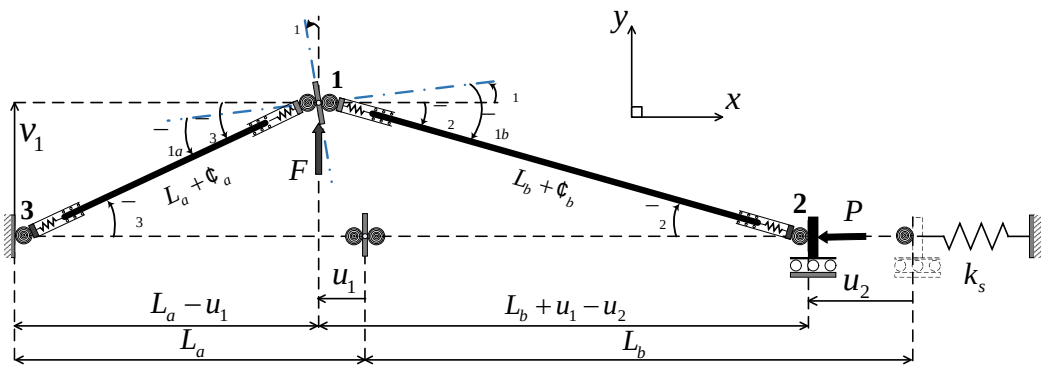


Figure 2: Presentation of simplified model

3 KINEMATIC OF THE MODEL

The kinematic of the model is shown in Fig.2. The model has four degrees of freedom, and the displacement vector is defined by:

$$\boldsymbol{U} = [\, v_1,\, u_1,\, \beta_1,\, u_2 \,]^{\mathrm{T}} \tag{1}$$

The deformation vector $\boldsymbol{E}$ is obtained:

$$\boldsymbol{E} = [\, \delta_a,\, \theta_{1a},\, \delta_b,\, \theta_{1b},\, u_2,\, \theta_2,\, \theta_3 \,]^{\mathrm{T}} \tag{2}$$

The components of the deformation vector $\boldsymbol{E}$ are related to the components of the displacement vector through:

$$\theta_2 = \arctan\left(\frac{v_1}{L_b + u_1 - u_2}\right) \tag{3}$$

$$\theta_3 = \arctan\left(\frac{v_1}{L_a - u_1}\right) \tag{4}$$

$$\theta_{1a} = \theta_3 - \beta_1 \tag{5}$$

$$\theta_{1b} = \theta_2 + \beta_1 \tag{6}$$

$$\delta_a = \sqrt{v_1^2 + (L_a - u_1)^2} - L_a \tag{7}$$

$$\delta_b = \sqrt{v_1^2 + (L_b + u_1 - u_2)^2} - L_b \tag{8}$$

The incremental relation between the displacements and deformations is obtained by differentiating the above equations:

$$\delta \boldsymbol{E} = \frac{\partial \boldsymbol{E}(\boldsymbol{U})}{\partial \boldsymbol{U}} \delta \boldsymbol{U} = \boldsymbol{B}\, \delta \boldsymbol{U} \tag{9}$$

in which

$$\boldsymbol{B} = \begin{bmatrix} \sin\theta_3 & -\cos\theta_3 & 0 & 0 \\ \dfrac{\cos\theta_3}{\tilde{L}_a} & \dfrac{\sin\theta_3}{\tilde{L}_a} & -1 & 0 \\ \sin\theta_2 & \cos\theta_2 & 0 & -\cos\theta_2 \\ \dfrac{\cos\theta_2}{\tilde{L}_b} & -\dfrac{\sin\theta_2}{\tilde{L}_b} & 1 & \dfrac{\sin\theta_2}{\tilde{L}_b} \\ 0 & 0 & 0 & 1 \\ \dfrac{\cos\theta_2}{\tilde{L}_b} & -\dfrac{\sin\theta_2}{\tilde{L}_b} & 0 & \dfrac{\sin\theta_2}{\tilde{L}_b} \\ \dfrac{\cos\theta_3}{\tilde{L}_a} & \dfrac{\sin\theta_3}{\tilde{L}_a} & 0 & 0 \end{bmatrix} \tag{10}$$

where $\tilde{L}_a = L_a + \delta_a$ and $\tilde{L}_b = L_b + \delta_b$. By inserting Eqs. (3-8) into Eq. (10), the transformation matrix $\boldsymbol{B}$ is completely in function of $\boldsymbol{U}$.

4 STATIC EQUILIBRIUM OF THE SYSTEM

The equation of the virtual work can be expressed as:

$$\delta \boldsymbol{E}^{\mathrm{T}} \boldsymbol{S} = \delta \boldsymbol{U}^{\mathrm{T}} \boldsymbol{f}^{ext} \tag{11}$$

where $\boldsymbol{S}$ is the internal force vector (see Fig. 3) conjugate to the deformation vector $\boldsymbol{E}$. It is given by:

$$\boldsymbol{S} = [\, N_a, M_{1a},\, N_b,\, M_{1b},\, F_2,\, M_2,\, M_3]^{\mathrm{T}} \tag{12}$$

in which N_a and N_b are the normal forces in elements a and b, respectively. F_2 is the force in the spring attached to node 2.

$$F_2 = k_s u_2 \tag{13}$$

The external force vector(See Fig.2) is given by

$$\boldsymbol{f}^{ext} = [\, F\,,\, 0\,,\, 0\,,\, P\,]^{\mathrm{T}} \tag{14}$$

where F and P denote external transversal load and external static compressive load, respectively. Inserting Eq. (9) into Eq. (11) gives the relation between the internal force vector and the external force vector, that is the equilibrium equations:

$$\boldsymbol{f}^{ext} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{S} \tag{15}$$

Denoting

$$\boldsymbol{f}_g^{int} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{S} \tag{16}$$

the global internal force vector, the equilibrium equation becomes:

$$\boldsymbol{f}^{ext} = \boldsymbol{f}_g^{int} \tag{17}$$

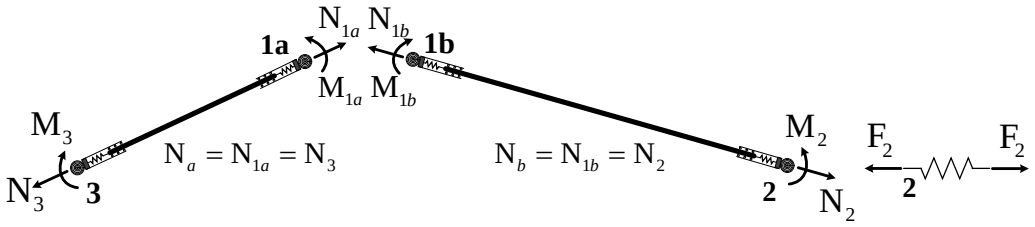


Figure 3: Stress resultant diagram

The elongations associated to the normal forces at each hinge are calculated from:

$$\delta_{1a} = \alpha_a \delta_a \tag{18}$$

$$\delta_{1b} = \alpha_b \delta_b \tag{19}$$

$$\delta_2 = (1 - \alpha_b)\delta_b \tag{20}$$

$$\delta_3 = (1 - \alpha_a)\delta_a \tag{21}$$

The parameters α_a and α_b describe the repartition of the total elongations of bars a and b at their ends.

5 COMPLEX ELASTO-PLASTIC MODELS FOR GENERALIZED PLASTIC HINGE

The elastic behavior of the generalized hinge is uncoupled whereas axial-moment interaction is considered in the plastic range. We adopt the total generalized strain rate decomposition into elastic and plastic parts

$$\dot{\boldsymbol{d}} = \dot{\boldsymbol{d}}^e + \dot{\boldsymbol{d}}^p \tag{22}$$

where $\dot{\boldsymbol{d}} = \left[\dot{\theta}, \dot{\delta}\right]^{\mathrm{T}}$. For an associated flow rule, the direction of the generalized plastic strain rate vector is given by the gradient to the yield function, with its magnitude given by the plastic multiplier rate $\dot{\lambda}$:

$$\dot{\boldsymbol{d}}^p = \dot{\lambda} \frac{\partial \Phi}{\partial \boldsymbol{s}} \tag{23}$$

where $\boldsymbol{s} = [\, M, N \,]^{\mathrm{T}}$ is the generalized stress vector containing the bending and axial forces in the hinge. The plastic multiplier $\dot{\lambda}$ is determined by the classical complementary conditions:

$$\dot{\lambda} \geq 0, \qquad \Phi(M, N) \leq 0, \qquad \dot{\lambda}\Phi(M, N) = 0 \tag{24}$$

The above relations characterize the loading/unloading conditions. For $\dot{\lambda} > 0$, persistency of the plastic state during plastic flow requires:

$$\dot{\lambda}\, \dot{\Phi}(M, N) = 0 \tag{25}$$

In this paper, we consider a family of symmetric and convex yield surface of superelliptic shape:

$$\Phi(M,N) = \|\boldsymbol{s}\|_{\beta_p} - 1 = 0 \tag{26}$$

where $\| \bullet \|_{\beta_p}$ denotes the superelliptic norm:

$$\|\boldsymbol{s}\|_{\beta_p} = \left(\left| \frac{M}{M^p} \right|^{\beta} + \left| \frac{N}{N^p} \right|^{\beta} \right)^{\frac{1}{\beta}} \tag{27}$$

with $1 \leq \beta < \infty$. The curve intersects x-axis at N^p and $-N^p$; it intersect the y-axis at M^p and $-M^p$. The coefficient β, called roundness factor, controls the shape of the yield criterion. This shape evolves from a parallelogram ($\beta = 1$) to a rectangular($\beta \to \infty$). The limit curve becomes a pinched diamond for $\beta < 1$ and is no longer convex. When β goes over 2, the curve become a rounded rectangle. The above family of yield surfaces is further widened by considering generalized superelliptic shapes described by

$$\Phi(M,N) = \left(\left| \frac{M}{M^p} \right|^{\alpha} + \left| \frac{N}{N^p} \right|^{\beta} \right)^{\frac{1}{\gamma}} - 1 \tag{28}$$

For $\alpha = 1$ and $\gamma = 1$ we recover the yield function to the one proposed by Duan and Chen in [17] for a doubly symmetrical steel cross section which is mirrored to give a complete envelop by:

$$\Phi(M,N) = \left| \frac{M}{M^p} \right| + \left| \frac{N}{N^p} \right|^{\beta} - 1 \tag{29}$$

An associated flow rule is adopted where plastic deformation (axial elongation and rotation) are constrained to occur in a direction normal to the yield surface

$$\dot{\theta}^p = \dot{\lambda} \frac{\partial \Phi}{\partial M} = \dot{\lambda} \left(\left| \frac{M}{M^p} \right|^{\alpha} + \left| \frac{N}{N^p} \right|^{\beta} \right)^{\frac{1-\gamma}{\gamma}} \frac{\alpha}{\gamma \left(M^p\right)^{\alpha}} |M|^{\alpha-1} \, sgn(M) \tag{30}$$

$$\dot{d}^p = \dot{\lambda} \frac{\partial \Phi}{\partial N} = \dot{\lambda} \left(\left| \frac{M}{M^p} \right|^{\alpha} + \left| \frac{N}{N^p} \right|^{\beta} \right)^{\frac{1-\gamma}{\gamma}} \frac{\beta}{\gamma \left(N^p\right)^{\beta}} |N|^{\beta-1} \, sgn(N) \tag{31}$$

where the *signum* function $sgn(\bullet)$ is defined by

$$sgn(x) = \begin{cases} 1 & for \quad x > 0 \\ 0 & for \quad x = 0 \\ -1 & for \quad x < 0 \end{cases} \tag{32}$$

6 NONSMOOTH DYNAMICS: IMPACT LOADING

6.1 Contact model

The model, see Fig. 4, considers the collision between a vehicle, modelled by a rigid point mass m_c, and a column represented by the model proposed in Section 2. The mass of the column is assumed to be concentrated at the impact point. Since the column is modelled by two

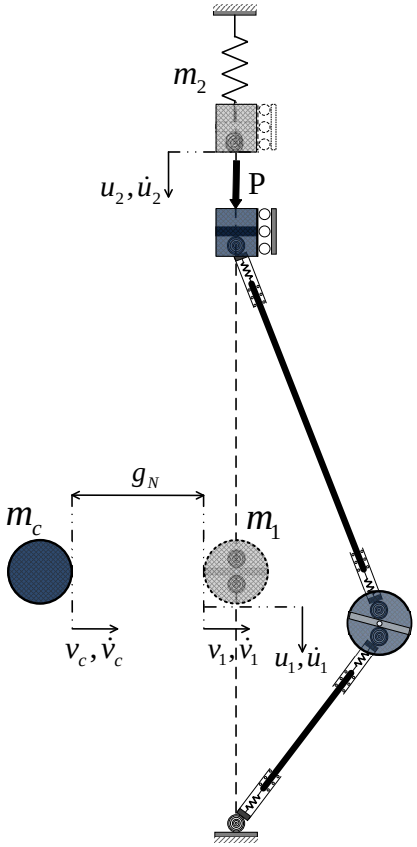


Figure 4: Contact model

rigid bars, the concentrated mass m_1 can be taken as a third of the total mass of the column. Unilateral contact is assumed with a gap function given by:

$$g_N = v_c - v_1 \tag{33}$$

where v_c is the position of the mass m_c and v_1 is the position of the mass m_1. The relative velocity is given by:

$$\gamma_N = \dot{v}_c - \dot{v}_1 \tag{34}$$

6.2 Constraint law

Before the equations of motion of the impacted bodies can be written, the constraint laws must be imposed so that one can detect when contact and impact occur. The constraint law for a unilateral contact is described by a Signorini's law at a position level below:

$$g_N \geq 0 \; ; \; \lambda_N \geq 0 \; ; \; g_N \lambda_N = 0 \tag{35}$$

To accommodate energy dissipation during collision and proposed the following impact law:

$$\gamma_N^+ = -\varepsilon \gamma_N^- \tag{36}$$

where γ_N^+ denotes relative velocity after impact and γ_N^- denotes relative velocity before impact. $\varepsilon \in [0, 1]$ is the restitution coefficient; $\varepsilon = 1$ for a purely elastic contact and $\varepsilon = 0$ for

a completely inelastic contact. By inserting the Newton's impact law, the constraint law at the velocity level can be written as:

$$\xi_N \geq 0, \qquad -\Lambda_N \leq 0, \qquad \xi_N \Lambda_N = 0 \tag{37}$$

where $\xi_N = \gamma_N^+ + \varepsilon \gamma_N^-$ and Λ_N denotes the percussion force. The above relations (37) can be combined together to give the following variational inequality:

$$(-\Lambda_N) \in \mathbb{R}_0^-, \qquad \xi_N \left(\Lambda_N^* - \Lambda_N\right) \geq 0, \qquad \forall\, (-\Lambda_N^*) \in \mathbb{R}_0^- \tag{38}$$

6.3 Equations of Motion

Depending on the value of the gap g_N, see Eq. (35), the equations of motion must be considered into two cases, open contact ($g_N > 0$) and closed contact ($g_N = 0$).

6.3.1 Open contact

The model is illustrated in Fig. 4. In addition to the masses m_1 and m_c, the mass m_2 of the surrounding structure at the top of the column is also considered. The dynamic effect associated with the rotation β_1 (see Fig. 2) of mass m_1 is neglected. Hence, the equations of the open contact motion are:

$$\begin{aligned}
&m_c\, \ddot{v_c} = 0 \\
&m_1\, \ddot{v_1} + \boldsymbol{f}_g^{\text{int}}(1) = 0 \\
&m_1\, \ddot{u_1} + \boldsymbol{f}_g^{\text{int}}(2) = 0 \\
&m_2\, \ddot{u_2} + \boldsymbol{f}_g^{\text{int}}(4) - P = 0 \\
&\boldsymbol{f}_g^{\text{int}}(3) = 0
\end{aligned} \tag{39}$$

where $\boldsymbol{f}_g^{\text{int}} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{S}$ is the global internal force vector defined in Eq. (16). In this work, midpoint scheme is used to solve the dynamic equations. By introducing midpoint scheme in the equations of motion (Eqs. (39)) at the time $n + \frac{1}{2}$, it is obtained:

$$\begin{aligned}
&\frac{2m_c}{\Delta t}\left(\frac{{v_c}^{n+1} - {v_c}^{n}}{\Delta t} - \dot{{v_c}}^{n}\right) = 0 \\
&\frac{2m_1}{\Delta t}\left(\frac{{v_1}^{n+1} - {v_1}^{n}}{\Delta t} - \dot{{v_1}}^{n}\right) + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(1) = 0 \\
&\frac{2m_1}{\Delta t}\left(\frac{{u_1}^{n+1} - {u_1}^{n}}{\Delta t} - \dot{{u_1}}^{n}\right) + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(2) = 0 \\
&\frac{2m_2}{\Delta t}\left(\frac{{u_2}^{n+1} - {u_2}^{n}}{\Delta t} - \dot{{u_2}}^{n}\right) + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(4) - P = 0 \\
&\boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(3) = 0
\end{aligned} \tag{40}$$

Since $\boldsymbol{f}_g^{\text{int}}$ is nonlinear, Newton Raphson iterations are required in order to solve Eq. (40).

6.3.2 Closed contact

During closed contact, the velocity may jump at the time instances when the impacts occur. At those times, the velocity is not differentiable and the contact force is impulsive. A set of differential measures are adopted to write a combined equation of motion so that one does not

have to write separately the equation for smooth part and impact part of the close contact. The equations of motion of the colliding masses, m_c and m_1 for a closed contact motion using a set of measure equation is now written as:

$$\begin{aligned} & m_c\, d\dot{v}_c = d\Lambda_N \\ & m_1\, d\dot{v}_1 + \boldsymbol{f}_g^{\text{int}}(1)dt = -d\Lambda_N \\ & g_N = 0\ ,\ \xi_N \geq 0 \end{aligned} \tag{41}$$

where the sets of measure for the velocity and contact force are given by:

$$d\dot{v} = \dot{v}\, dt + (\dot{v}^+ - \dot{v}^-) \tag{42}$$

$$d\Lambda_N = \lambda_N dt + \Lambda_N d\eta \tag{43}$$

By using

$$\int_n^{n+1} \boldsymbol{f}_g^{\text{int}} dt = \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}} \Delta t\ ;\ \int_n^{n+1} d\Lambda_N = P_N\ ;\ \int_n^{n+1} d\dot{v} = \dot{v}^{n+1} - \dot{v}^n \tag{44}$$

Eqs. (41) can be integrated between t_n and t_{n+1}. By considering also the equations associated with other degrees of freedom, it is obtained:

$$\begin{aligned} & m_c(\dot{v_c}^{n+1} - \dot{v_c}^n) = P_N \\ & m_1(\dot{v_1}^{n+1} - \dot{v_1}^n) + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(1)\Delta t = -P_N \\ & m_1\, \ddot{u_1} + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(2) = 0 \\ & m_2\, \ddot{u_2} + \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(4) - P = 0 \\ & \boldsymbol{f}_g^{\text{int},n+\frac{1}{2}}(3) = 0 \end{aligned} \tag{45}$$

Before solving Eqs. (45), the constraint law in Eq. (37) is considered in order to determine if the contact (or impulsive) force P_N is present or not. The following methodology is adopted. First, the contact(or impulsive) force is assumed to be zero and Eqs. (45) are solved by using the midpoint scheme. Then, ξ_N, defined in Eq. (37) is calculated as:

$$\xi_N = \gamma_N^{n+1} + \varepsilon\gamma_N^n \tag{46}$$

If ξ_N is positive, then the prediction of no contact force is true. If ξ_N is negative, then the contact or impulsive force exists and has a positive value. In such case, the equation $\xi_N = 0$, is added to Eqs. (45) in order to calculate both the kinematic quantities and the contact (or impulsive) force.

7 NUMERICAL EXAMPLES

7.1 Parameters

The purpose of the examples presented in this section is to show the ability of the proposed model in capturing the main features of the steel column's behavior under the vehicle impact. A numerical example of a build - in column of type HEB 220 impacted by a rigid mass is given. The parameters of the problem are given in Table 1. The other necessary parameters

are described as follows. The stiffness of the elastic spring, the compressive load and the mass attached to top end of the column are respectively provided: $k_s = 10\,\mathrm{MN/m}$, $P = 0.5\,\mathrm{MN}$ and $m_2 = 0\,\mathrm{kg}$.

Table 1: Parameters of the model

Material properties		
Young modulus	E	$210\,000\,\mathrm{MPa}$
Yield strength	f_y	$355\,\mathrm{MPa}$
Column Type: HEB 220		
Section area	A	$91\,\mathrm{Cm}^2$
Moment of inertia	I	$8\,091\,\mathrm{Cm}^4$
Nominal weight	m_l	$70.1\,\mathrm{kg/m}$
Length	L_a	$1\,\mathrm{m}$
	L_b	$3\,\mathrm{m}$
Bending resistance	M^p	$0.307\,\mathrm{MNm}$
Axial resistance	N^p	$3.23\,\mathrm{MN}$
Impactor		
Mass	m_c	$1\,500\,\mathrm{kg}$
Initial velocity	v_c^0	$10\,\mathrm{m/s}^2$
initial gap	g_N	$0\,\mathrm{m}$
Restitution coefficient	ε	0

The elastic stiffness of the axial and rotational springs for a build - in column are provided as:

$$k_{1a}^{\theta} = k_3^{\theta} = \frac{6EI}{L_a} \tag{47}$$

$$k_{1b}^{\theta} = k_2^{\theta} = \frac{6EI}{L_b} \tag{48}$$

$$k_{1a}^{\delta} = k_3^{\delta} = \frac{2EA}{L_a} \tag{49}$$

$$k_{1b}^{\delta} = k_2^{\delta} = \frac{2EA}{L_b} \tag{50}$$

Fig. 5 illustrates the evolution of the horizontal displacement of both masses with time. Immediately from the beginning, both masses come into contact and remain in such motion until the energy absorbing capacity overcomes the motion and both masses are pushed back and then lose contact. Fig. 6 show the nonlinear changes of vertical displacements with time. The normal forces in each hinges start with an initial compression value as shown in Fig. 7. As long as the impact loading is applied, the normal forces begin to decrease their values toward zero. From the evolution of the bending moments in Fig. 8, one can infer that each hinge plasticizes at a different time.

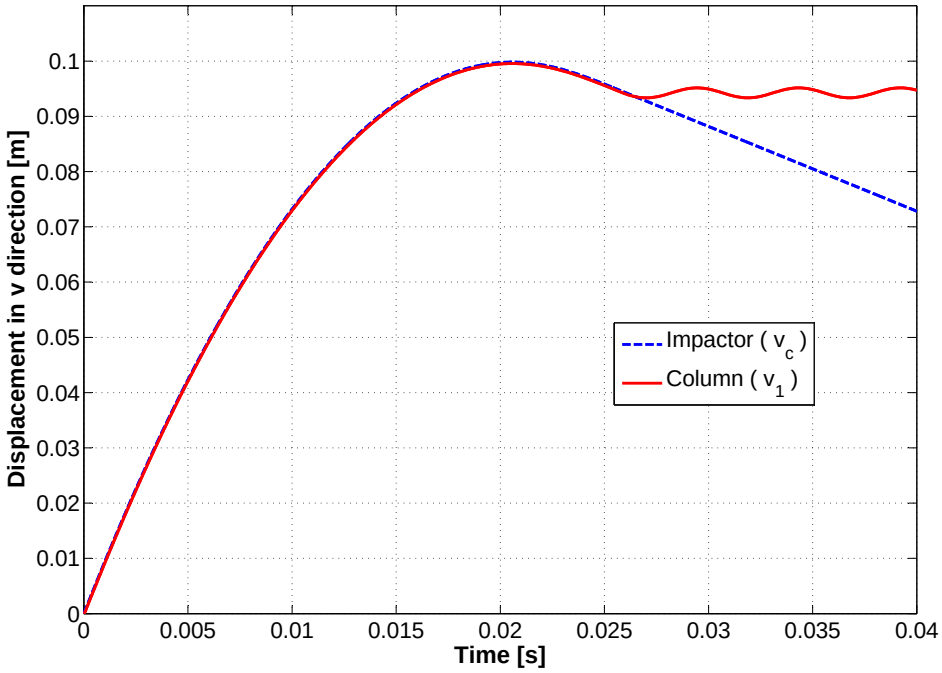


Figure 5: Evolution of horizontal displacement v_1

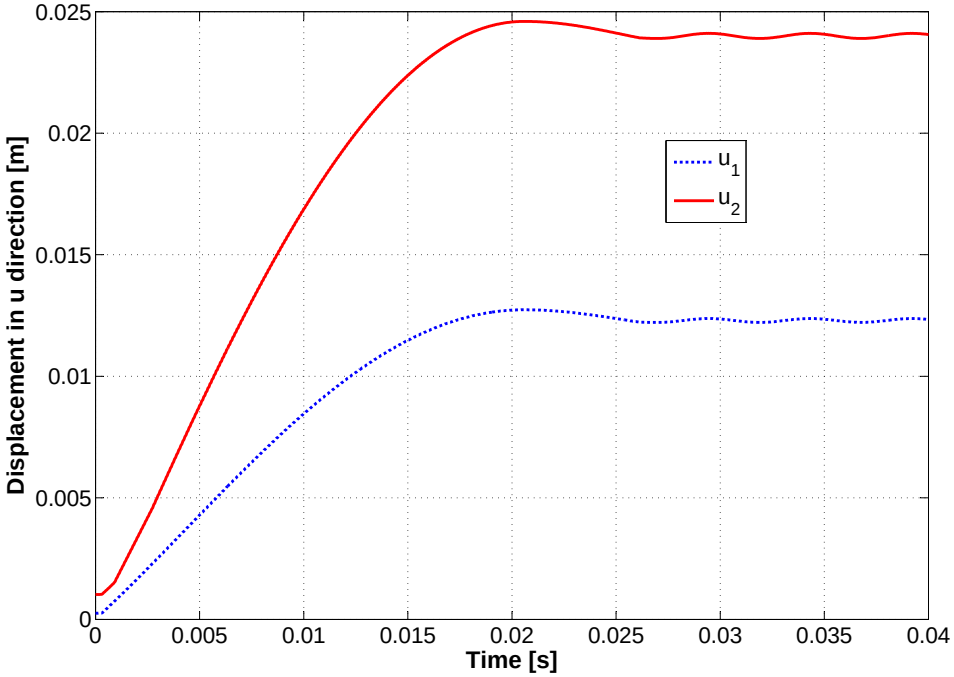


Figure 6: Evolution of vertical displacement u_1 and u_2

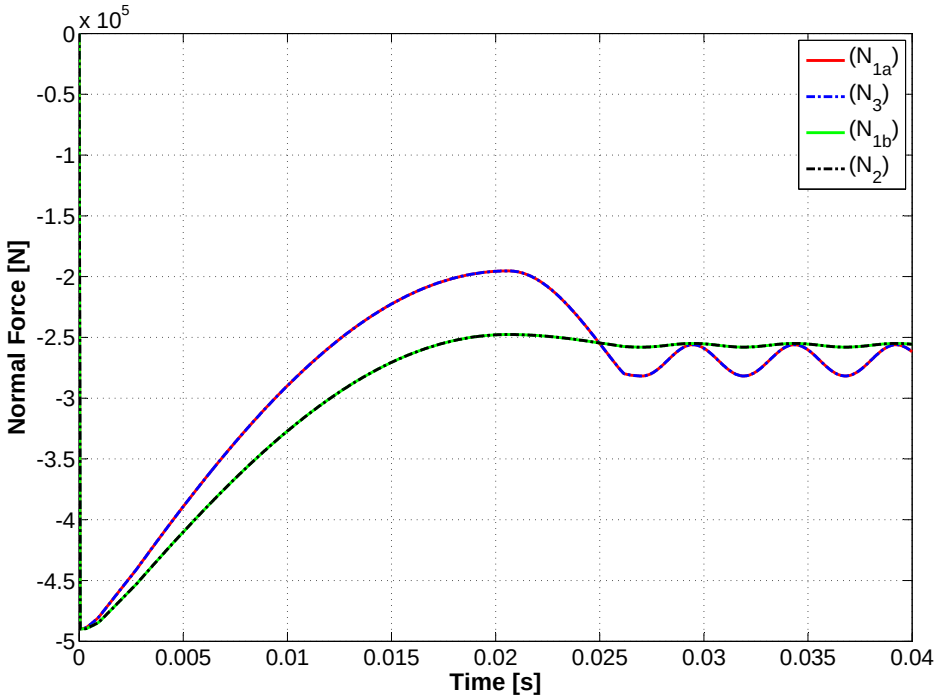


Figure 7: Evolution of normal forces

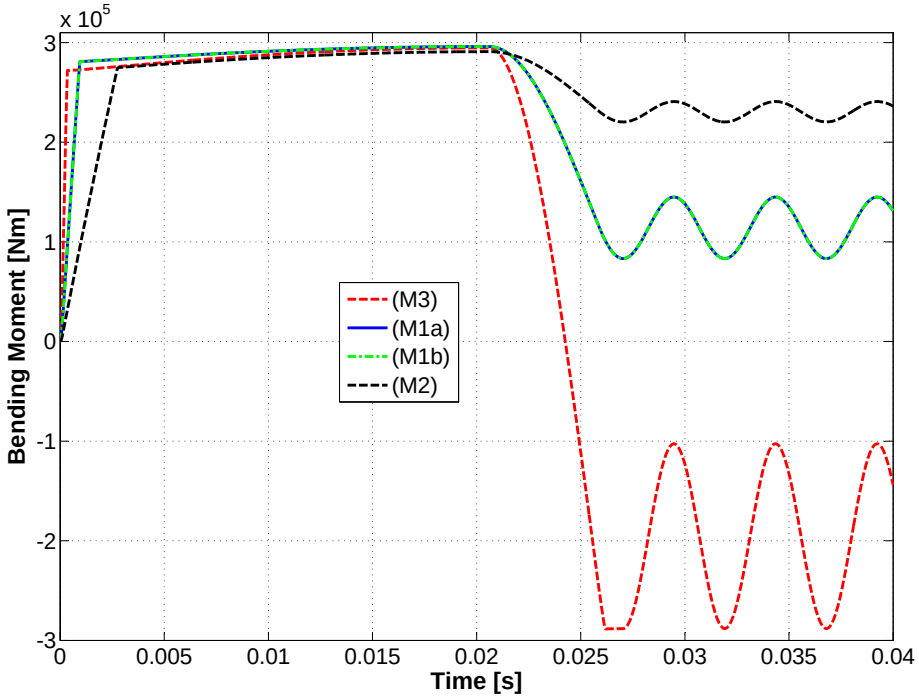


Figure 8: Evolution of bending moments

8 CONCLUSION

In the paper, the new nonlinear model of a steel column subjected to a transversal vehicle mass was presented. The nonlinear geometry and material are included in the model. The generalized plastic hinge method is adopted in order to model a column by using two rigid elements connected with elasto - plastic hinges. The yield criterion od the hinges are governed by a variety of super - elliptic shape functions. For the non -smooth problem, the constraint laws such as Signorini's force law and Newton impact law are imposed to write the equations of motion. A set of differential measure equations are adopted to write overall equations of close - contact motion in order to describe both the impact and impact - free motion. The integration of the equations of motion are executed using the classical mid - point rule scheme.

Acknowledgements

The authors gratefully acknowledge financial support by the European Commission (Research Fund for Coal and Steel) through the project ROBUSTIMPACT under grant agreement RFSR-CT-2012-00029.

REFERENCES

[1] B.A. Izzuddin, A.G. Vlassis, A.Y. Elghazouli, D.A. Nethercot, Progressive collapse of multi-storey buildings due to sudden column loss — Part I: Simplified assessment framework. *Engineering Structures*, **30(5)**, 1308-1318, 2008.

[2] L. Comeliau , B. Rossi , J-F. Demonceau, Robustness of steel and composite buildings suffering the dynamic loss of a column. *Struct Eng Int J*, **22(3)**, 323–9, 2012.

[3] Clara Huvelle, Van-Long Hoang, Jean-Pierre Jaspart, Jean-François Demonceau, Complete analytical procedure to assess the response of a frame submitted to a column loss. *Engineering Structures*, **86**, 33–42, 2015.

[4] J-F. Demonceau, Steel and composite building frames: sway response under conventional loading and development of membrane effects in beams further to an exceptional action. *PhD thesis, University of Liege*, 2008.

[5] Luan Huynh, Stephen Foster, Hamid Valipour, Robert Randall, High strength and reactive powder concrete columns subjected to impact: Experimental investigation. *Construction and Building Materials*, **78**, 153-171, 2015.

[6] H.M.I. Thilakarathna, D.P. Thambiratnam, M. Dhanasekar, N. Perera, Numerical simulation of axially loaded concrete columns under transverse impact and vulnerability assessment. *International Journal of Impact Engineering*, **37(11)**, 1100-1112, 2010.

[7] H. Sharma, S. Hurlebaus, P. Gardoni, Performance-based response evaluation of reinforced concrete columns subject to vehicle impact. *International Journal of Impact Engineering*, **43**, 52-62, 2012.

[8] M.A.K.M. Madurapperuma, Anil C. Wijeyewickrema, Response of reinforced concrete columns impacted by tsunami dispersed 20' and 40' shipping containers. *Engineering Structures*, **56**, 1631-1644, 2013.

[9] British Standards Institution, *BS 5950: Structural use of steelwork in buildings, Part 1: code of practice for design – rolled and welded sections*. London (UK), 2001.

[10] European Committee for standardization, *EN 1991-1-7:2006, Eurocode 1: Actions on structures – Part1-7: general actions – accidental actions*. Brussels: Belgium, 2006.

[11] Unified Facilities Criteria (UFC)-DoD, Department of Defense, *Design of buildings to resist progressive collapse.*, 2009.

[12] American society of civil engineers (ASCE), *Minimum design loads for buildings and other structures*, SEI/ASCE 7-05. Reston (VA, USA), 2005.

[13] NORSOK N004, *Design of steel structures*, Rev. 2., 2004.

[14] Amr A. Nassr, A. Ghani Razaqpur, Michael J. Tait, Manuel Campidelli, Simon Foo, Single and multi degree of freedom analysis of steel beams under blast loading. *Nuclear Engineering and Design*, **242**, 63-77, 2012.

[15] K.-J. Ek, P. Mattsson, Design with Regard to Blast- and Fragment Loading. *Institutionen för bygg- och miljöteknik, Chalmers tekniska högskola*, no: 2009:81. Göteborg, Sweden, 2009.

[16] J.M. Biggs, Introduction to structural dynamics. New York: McGraw-Hill, 1964.

[17] L. Duan, W-F. Chen, A yield surface equation for doubly symmetrical sections. *Engineering Structures*, **12(2)**, 114-119, 1990.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MATERIAL MODELLING OF CAST ALUMINIUM BY APPLICATION OF THE WILKINS DAMAGE MODEL

Christian Mühlstätter, Matthias Hartmann

[1] Leichtmetallkompetenzzentrum Ranshofen GmbH, Austrian Institute of Technology
Christian.Muehlstaetter@ait.ac.at
Matthias.Hartmann@ait.ac.at

Keywords: Aluminium cast alloy, Damage model, Crashworthiness, Triaxiality, Lode Parameter

Abstract. *Since cast light metal components gain more relevance in automotive application, the development of adequate material models including consideration of damage is necessary. A key issue in material modelling of metals is the consideration of material damage in addition to the elasto-plastic behaviour. The damage behaviour is a specific property of a material which characterises the crashworthiness of a structural material, especially for cast alloys.*
These have a significant low ductility compared to wrought alloys. Numerical simulations of cast alloy panels under crash loadings show noticeable overestimated results if material damage is not considered.
A damage model for ductile fracture introduced by Wilkins is discussed, as well as the appropriate strategy to characterise the material parameters for this model. In these studies an elasto-plastic material model is set up for a squeeze cast Aluminium alloy which includes the Wilkins approach for consideration of damage. The material model is characterised and validated on flat tensile tests. The final material model is then used in models of crushed hat profiles and validated by experiments.
The comparison of the experiment and the numerical simulation demonstrate a good accordance of the force displacement data and a difference in energy absorption of 3 %.

1 INTRODUCTION

Cast aluminium alloy achieve an increasing relevance in automotive application, due to their light weight design properties. Especially, the relative low density and the well-established manufacturing processes of these components constitute this field of application. For design purpose, numerical simulations build a state of the art method. In there, appropriate material models are mandatory to capture the structural behaviour in a correct manner not only in the elasto-plastic regime, but also during damage until fracture. A common way to perform failure modelling of a material is the micromechanical consideration which observes the phenomena of void nucleation, growth and coalescence. A material model including a micromechanical approach is introduced by Gurson [1]. Further models, as the approach of Needleman and Tvergaard [2] are resulted of modifications of the Gurson Damage Model. One of the limitations of a formulation according to Gurson is the neglected influence of damage caused by shear load cases. The modification by the approach of Nahshon and Hutchinson [3] overcomes this problem.
The phenomenological approach of Johnson-Cook [4] takes temperature and strain rate for damage evolution into account. Mathematical damage models are available in finite element (FE) simulation tools. These use a damage description by the definition of curves, which constitutes the coherence of load dependent parameters, e.g. plastic strain and triaxiality, concerning their damage influence. This paper focuses on the material modelling of an Aluminium cast alloy under crushing load. Principles of failure modelling and especially the Wilkins Damage Model are introduced. It will then go to the investigation and validation of the same. A material model including Wilkins Damage approach will be defined for an AlSi cast alloy on flat tensile specimen. This material model will be validated on hat profile crushing tests.

2 FAILURE MODELLING

Material definitions are an essential input for numerical simulations. Besides the definitions of elasticity, flow rule and hardening type the additional definition of material damage are essential. Especially, where crash loadings induce high deformations it is important to correctly capture material failure and to have a convenient method for parameter characterisation. Every stress condition induced by mechanical loads or temperature causes material damage. Once damage reaches a critical, material dependent level failure is observed. In numerical simulation an incremental damage value is determined in each time step and cumulated for the description of failure.
To link the actual stress state to the damage effect, several parameters are introduced. These parameters reduce the stress tensor to equivalent values for modelling of failure. The most common parameter in this context is the stress triaxiality η, which is defined according to (1) as function of the hydrostatic stress σ_H and the equivalent stress of Mises σ_{Mises}.

$$\eta = \frac{\sigma_H}{\sigma_{Mises}} \tag{1}$$

The stress triaxiality is used in numerous damage and failure models [[4], [2], [1]], e.g. in the approach of Gurson. The purpose is the characterisation of the stress state according to the influence of material damage. Stress triaxiality is considered as a crucial factor for ductile fracture [5]. An additional way to consider damage in a material model is the formulation of the Lode Stress Parameter μ_σ.
The Lode Stress Parameter is determined by the deviatoric part of the stress tensor with the

deviatoric principal stresses $s_1 > s_2 > s_3$ in the form

$$\mu_\sigma = \frac{2\,s_2 - s_1 - s_3}{s_1 - s_3}\ . \tag{2}$$

The Lode Stress Parameter is defined in the range of $[-1, 1]$, whereas the value of $\mu_\sigma < 0$ denotes tensile stress state, $\mu_\sigma = 0$ denotes pure shear stress state and $\mu_\sigma > 0$ indicates a compression stress condition. A characterisation of the stress state using the Lode stress Parameter in addition to the stress triaxiality is a convenient way for modelling ductile fracture behaviour. The Lode Stress Parameter distinguishes the stress state in axisymmetric and shear dominated stress conditions [6].

3 WILKINS DAMAGE MODEL

The damage model introduced by M. L. Wilkins [7] uses a phenomenological approach for the formulation of damage and failure. It takes into account that material damage which is introduced by the weighting of two factors in a mechanical loading case. First the hydrostatic part and second the deviatoric part of a stress tensor. During loading, both of them imply incremental damage until failure occurs. The damage effect of hydrostatic pressure is observed as spalling mechanism in the material. The deviatoric stress part leads to linkage of voids in a material [7]. Strain rate effects are not included in the approach of Wilkins. The cumulative damage formulation by Wilkins is implemented in the numerical simulation code LS-Dyna [8] as

$$D = \int \omega_1\, \omega_2\, d\varepsilon_{pl,equ}\ , \tag{3}$$

where

$$\omega_1 = \left(\frac{1}{1 - \gamma\, \sigma_H}\right)^{\alpha} \tag{4}$$

and

$$\omega_2 = (2 - A_D)^{\beta}. \tag{5}$$

In equation (3) the cumulative damage *D* is a function of the two weighting terms ω_1, ω_2 and, furthermore, of the equivalent plastic strain increment $d\varepsilon_{pl,equ}$. The parameters α, β and γ in equations (4) and (5) are material dependent values according to the particular damage behaviour. These values have to be identified for the specific AlSi cast alloy of this investigation.
The approach of Wilkins calculates damage history by hydrostatic stress σ_H and the deviatoric part of the stress tensor. The deviatoric part of the stress tensor is determined by the deviatoric principal stresses s_1, s_2 and s_3 in parameter A_D as follows [8]:

$$A_D = \max\left(\left|\frac{s_2}{s_3}\right|, \left|\frac{s_2}{s_1}\right|\right) \tag{6}$$

Equations 3-6 describes material damage *D* caused by mechanical stresses. An additional criterion is needed for the definition of material failure.

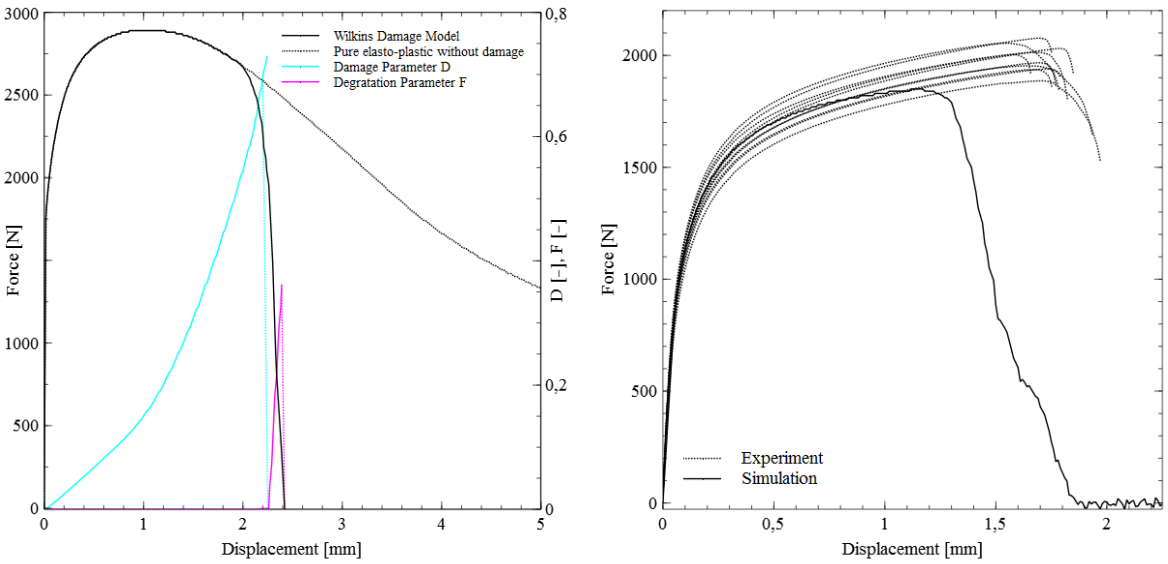
The following approach is implemented in LS-Dyna [8] by coupling of damage evolution and material degradation in the form

$$\frac{D}{D_c} > 1 \tag{7}$$

with D_c being a critical damage value for the material in use and the material degradation parameter F defined as

$$F = \frac{D - D_c}{D_s} \tag{8}$$

with the material dependent parameter D_s. To link the theory of failure modelling in the present work to the approach of Wilkins, a closer view into the damage formulation is necessary. In common material models, the influence of hydrostatic stress related to material damage is carried out by the stress triaxiality parameter η. Obviously, in the approach of Wilkins the unscaled amount of the hydrostatic stress is applied in equation (4). Furthermore, the Lode Stress Parameter is implemented to consider the effects of material damage [6]. It can be shown that the parameter A_D shows a similar characteristic as the absolute value of the Lode Stress Parameter [9]. Equations 3-8 include the material dependent parameters β, γ, D_c and D_s. The determination of these parameters asks for a material characterisation strategy. Therefore, four flat tensile specimen with different geometries are used, e.g. Merklein Geometry [10]. A pure tensile load is applied on these geometries until failure occurs. The different geometries cause failure by various stress triaxiality histories. On basis of the experimental force-displacement curves a reverse engineering process is done, which delivers the parameters for the Wilkins Damage Model [9].
To point out the principle of the implemented Wilkins Damage Model including failure criterion in LS-Dyna, a brief simulation of a flat tensile test is performed (Figure 1a, black full line) and compared to a model without consideration of damage and failure (Figure 1a, dotted line). The force-displacement curve of the simulation is extracted.



(a) (b)

Figure 1: (a) Investigation of the implemented Wilkins Damage Model in LS-Dyna by simulation of a tensile test geometry, (b) Experimental force-displacement data of the Merklein test geometry in comparison to a simulation curve including the Wilkins Damage Model

In Figure 1a the damage history variable D and the material degradation parameter F of the element with initial failure is plotted in parallel to the force-displacement history. It is apparent from Figure 1a that the damage parameter D increases steadily. At the point of critical damage D_c the material degradation parameter F is initiated. Once F reaches a value of 0.7 the element is deleted from the simulation. The selected element of the tensile geometry is the last element that failed, before fracture of the cross section is fully established. Hence, it is a gap between initiation of material softening and initiation of material degradation due to damage.
Figure 1b shows ten experimental force-displacement curves of a Merklein Geometry in dotted representation. In comparison a numerical simulation of the Merklein Geometry is considered and depicted in Figure 1b, full line.
The reverse engineering process of the test geometries delivers a set of parameters for the investigated AlSi cast alloy. The identified parameters in this study are listed in Table 1.

Table 1: Identified Parameters of the Wilkins Damage Model for the AlSi cast alloy of this study

Wilkins Damage Model Parameter	Determined value for an AlSi cast alloy
α	900
β	0.2
γ	10^{-5}
D_s	0.3

4 EXPERIMENTAL VALIDATION

A validation of the material is carried out to ensure that the calibrated material model is applicable to FE simulations of structural aluminium alloy. For this purpose, an experimental crushing test of a squeeze cast hat profile is performed in parallel simulation by means of FEM (Figure 2, cross section geometry). The length of the depicted hat profile is 250 *mm*.

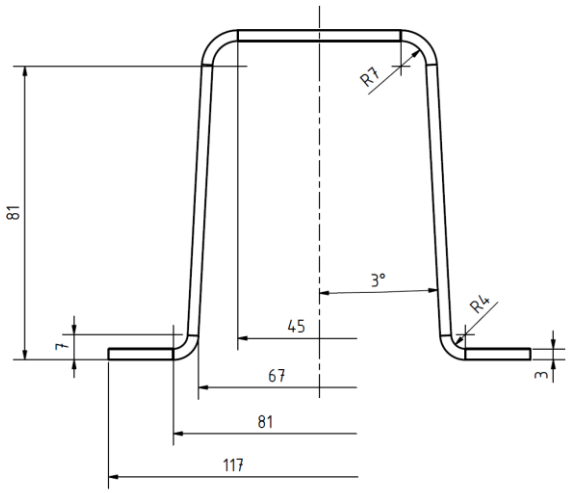


Figure 2: Cross section of the hat profile for the validation of material models

The experimental test setup is apparent from Figure 3 (a) in a 90° view rotated representation. The hat profile is crushed at a test speed of 5 $\frac{mm}{s}$ for a length of 100 mm in an ITC Interlaken 1000 *kN* press.

A total number of six experiments are performed. In parallel a numerical simulation is performed (Figure 3a). The hat profile is modelled by solid hexahedrons with an edge length of 1 *mm*. The test rig is simplified in the numerical model with two rigid shell planes. Contact definitions between shells and profile are applied and an additional contact for self-penetration is defined for the profile. Figure 3 shows the test geometry after an applied crushing displacement of 25 *mm*, simulation (a) and experiment (b). Under this test load, the middle sections of the hat profile observe material failure. The visual comparison of experiment and simulation shows a similar deformation mode of the model. Furthermore, the areas of the middle material failure are predicted properly in the simulation. On both edges of the experimental test geometry cracks are observed, which are not apparent in the simulation at this timestep. This effect may be caused by the ejector stamps and resulting differences in material properties in these areas. The appearance of these surface imperfections is a result of the casting process and is not considered in the FE simulation.

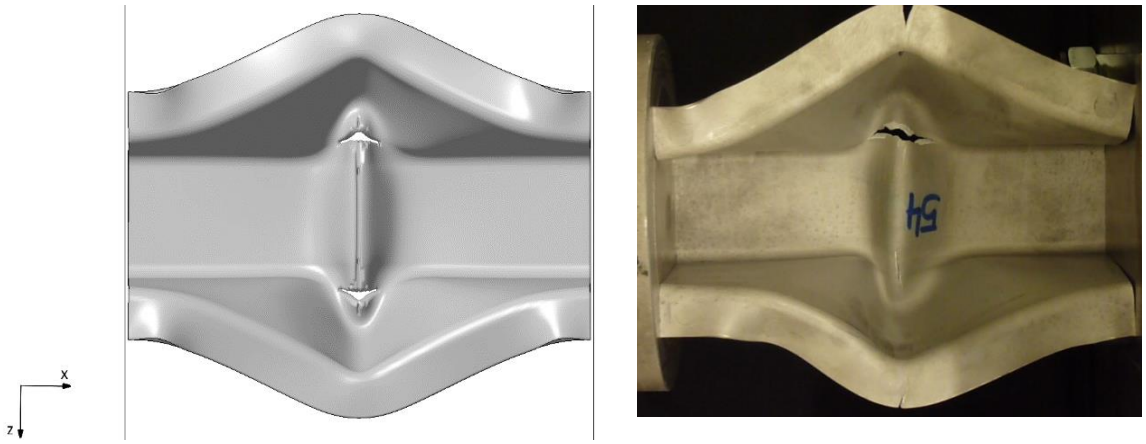


Figure 3: Comparison of deformation and failure in a crushing test after 25 mm displacement. A hat profile of AlSi cast with a not deformed lenth of 250 mm

In Figure 4, the force-displacement curves of the experiments and two different simulation models are plotted. The results of the experimental data is reduced to the upper and lower deviation bounds. One simulation model uses a pure elasto-plastic material mode. The second simulation model uses the same elasto-plastic material model with enhanced Wilkins Damage approach.
The initial loading phase including force peak and force drop shows a proper agreement between simulation and experiments. Only after displacements of 20 *mm* the numerical model results start to deviate significantly from each other.
The curve of the Wilkins Damage Model runs within the bounds of the experiments up to a displacement of 90 mm. In the remaining displacement area the difference to the experiments increases. The occurrence of ruptures through the entire cross section is observed in experiments. This behaviour is also predicted in the FE simulation and is apparent by the zero force level at a displacement of 90 *mm* in the curve of the Wilkins Model. Compared to the results of simulation including the Wilkins Damage Model, the pure elasto-plastic model leads to far higher force in the phase after 20 *mm* of displacement. The resulting energy absorption of the models is highly effected by this deviation.

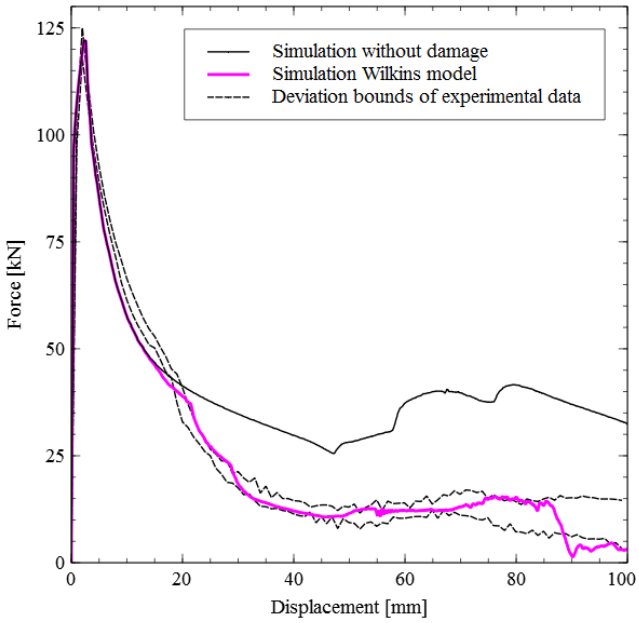


Figure 4: Force-Displacement curves of experimental data and simulations

Table 1 illustrates the relative deviation of force peaks and the absorbed energy from the simulations compared to the mean experimental value. The energy absorption is overestimated by +66.94 % if material damage is not considered with the pure elasto-plastic model. In the case of the Wilkins Damage Model, an excellent agreement of the energy absorption can be achieved with a deviation of -3.63 %.

Table 2: Relative deviation of the numerically calculated maximum force peak and the absorbed energy compared to the experimental mean curve

Model approach	Elasto-plastic + Wilkins Damage	Elasto-plastic
Force peak [%]	-0,66	-0,66
Absorbed energy [%]	-3,63	+66,94

5 CONCLUSIONS

The mayor aspects of this paper can be summarised as follows:

1. FE simulations of cast aluminium with pure elasto-plastic material models lead to distinctively overestimated results under crush loading.
2. FE simulations of cast aluminium with elasto-plastic material models and Wilkins Damage Model lead to very good results under crush loading.
3. The results of the investigation show that the Wilkins Damage Model is applicable for AlSi cast alloys.
4. The regions of rupture are well predicted by the Wilkins Damage Model. This grants applicability in numerical crash simulation.
5. The findings are limited, as the influence of the strain rate on the damage behaviour of AlSi cast alloys is not included in the approach of Wilkins.

REFERENCES

[1] A. L. Gurson, Porous rigid-plastic materials containing rigid inclusions - yield function, plastc potential and void nucleation, The Physical Metallurgy of Fracture, 1978.

[2] A. Needleman and V. Tvergaard, *An analysis of ductile rupture in notched bars,* Journal of the Mechanics and Physics of Solids, 1984.

[3] K. Nahshon and J. Hutchinson, "Modification of the Gurson Model for shear failure," *European Journal of Mechanics A/Solids 27 (2008) 1-17,* 2007.

[4] G. R. Johnson and W. H. Cook, *Fracture characteristics of three metals subjected to various strains, strain rates, temperatures and pressures,* Engineering Fracture Mechanics, 1985.

[5] Y. Bao and T. Wierzbicki, *On fracture locus in the equivalent strain and stress triaxiality space,* International Journal of Mechanical Sciences, 2004.

[6] I. Barsoum, J. Faleskog and S. Pingle, The inuence of the lode parameter on ductile failure strain in steel, Procedia Engineering, 2011.

[7] M. L. Wilkins, R. D. Streit and J. E. Reaugh, Cumulative-strain-damage model of ductile fracture: Simulation and prediction of engineering fracture tests, 1980.

[8] Livermore Software Technology Corporation; (LSTC), *LS Dyna Keyword User's manual Volume 2 Material models R 7.1,* 2014.

[9] C. Mühlstätter, *Versagensmodellierung von Aluminium Gusswerkstoffen, Master Thesis,* Wels: Upper Austria University of applied sciences, 2015.

[10] Q. Yin, B. Zillmann, S. Suttner, G. Gerstein, M. Biasutti, A. Tekkaya, M. Wagner, M. Merklein, M. Schaper, T. Halle and A. Brosius, *An experimental and numerical investigation of different shear test configurations for sheet metal characterization,* International Journal of Solids and Structures, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

FINITE ELEMENT SIMULATION OF CRACK PROPAGATION AND DELAMINATION IN LAYERED SHELLS DUE TO BLADE CUTTING

Federica Confalonieri, Aldo Ghisi, and Umberto Perego

Department of Civil and Environmental Engineering, Politecnico di Milano,
piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: {federica.confalonieri, aldo.ghisi, umberto.perego}@polimi.it

Keywords: Blade cutting, crack propagation, delamination, explicit dynamics, finite elements, directional cohesive elements.

Abstract. *This work is focused on some computational issues concerning the simulation of blade cutting of thin walled structures. In comparison to plain crack propagation problems, the presence of the blade brings in additional complexity, mainly due to the interaction between the blade and the material in the crack process zone. The blade sharpness introduces in the problem an extremely small geometrical scale (orders of magnitude smaller than a typical element in-plane size) that is here resolved using the so called directional cohesive elements, recently proposed in [6, 12]. A special feature of this type of problems is that the crack path is driven by the blade trajectory, which is prescribed and hence known in advance. Crack propagation is therefore modeled by adjusting the mesh in such a way that shell element edges are disposed along the expected main crack path and then by interposing directional cohesive elements between the sides of separating elements. The transition from a continuous mesh to a mesh containing a crack with a cohesive interface is well known to be critical for the solution accuracy. Nodal equilibrium is in general violated during the transition, with subsequent generation of spurious stress oscillations that, in view of the non-reversible nature of the problem, can lead to significant deviations in the stress response. This aspect is investigated in this contribution. The effect of the number of introduced directional cohesive elements per opening face is critically assessed and a simple correction, based on an automatic adaptation of the cohesive model, is proposed. Numerical tests taken from the literature are used to validate the proposed approach.*

1 INTRODUCTION

For the cutting simulation of thin shells, standard cohesive elements are not well suited to correctly address the interaction of the cohesive zone with a sharp blade. When a cutting tool interacts with a shell structure made of a ductile material, the blade curvature radius (sometimes of the order of micrometers) turns out to be far smaller than the typical size of a computationally acceptable in-plane discretization of the process zone, so that the interaction cannot be properly resolved. The blade can interfere with the correct transmission of the cohesive forces, leading to severe underestimation of the dissipated energy. A possible remedy to this inconvenience has been proposed in [6, 12], with the introduction of the so called directional cohesive elements (DCEs). When a crack propagation criterion is met at a node, the node is duplicated and cohesive string elements are attached to the opening faces. The main feature of these string elements is that they are able to detect contact with the cutting blade. Upon contact, the string element interacts with the cutting blade and deforms, transmitting cohesive forces to the two crack flanks in the correct directions. The correct amount of cohesive energy is therefore dissipated through the string elongation (see Figure 1).

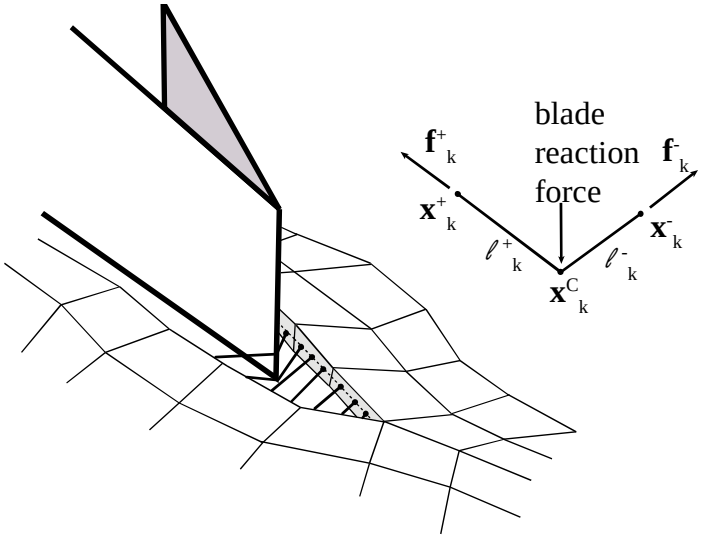


Figure 1: distribution of cohesive forces via the “directional” concept.

Despite the improved energy balance in the presence of a cutting blade, the issue of the accuracy of the cohesive zone discretization remains open. Another important issue arising during crack propagation is the oscillation in the stress field generated by the sudden release of a node at the crack tip. The node separation with the simultaneous introduction of the directional cohesive elements may generate a propagating shock, with severe stress oscillations. In the exact instant of the (directional or classical) cohesive element insertion, the nodal force balance is not automatically assured, because of the configuration change due to the appearing of new surfaces where cohesive forces have to be applied. These problems, which are present also with standard cohesive elements, are well known and have been discussed extensively in the literature on finite element cohesive models. Shellekens and De Borst [14] interpreted the oscillations as due to the combination of high tractions gradients and of the chosen integration scheme: in their examples, Gaussian integration (GI) for linear plane interface elements resulted in non-zero energy modes with coupling between degrees of freedom of individual node sets.

They found that the Newton-Cotes integration (NCI) rule performed better even for quadratic cohesive elements and ascribed the responsibility to the integration point located at the face centroid. Alfano and Crisfield [1], however, showed that a higher order of the numerical integration is not sufficient to allow for a coarser mesh discretization neither guarantee the robustness of the solution algorithm which, conversely, could be negatively affected.

Recently, Gillormini and Diani [8] discussed the effect of the numerical integration rule in case of finite strain. While finite strain formulations for cohesive elements have been widely used (see e.g. [11, 7]), these authors point out the influence of the choice of the reference configuration (original or current) where the numerical integration is carried out, and the consequence of neglecting the local basis mobility in the finite element formulation. The latter choice, in particular, leads to completely neglect a term in the traction-separation law: this is legitimate only in small deformations or whether tractions and displacements are collinear, such as in the present work. Their numerical peeling experiment shows no sensible differences between NCI and GI rules in nonlinear regimes, while a less precise result has been observed for NCI in the elastic part of the cohesive law.

The possibility to exploit numerical integration to improve the cohesive element computational efficiency has been investigated by Yang *et al.* [18], who introduced a cohesive element featuring a multiple subdomain integration: the numerical integration for each cohesive finite element is subdivided in zones whose dimensions are smaller than the cohesive zone size and inside which GI or NCI schemes are carried out. This approach makes it possible to increase the mesh size from a fraction (typically $\frac{1}{3} \div \frac{1}{5}$) of the cohesive zone to a size comparable and even slightly larger of the cohesive zone. Moreover, the modifications in a standard finite element implementation are quite limited, see [13]. In addition, it has been shown that it is possible to move the evolving crack front within a partially failed cohesive element if an improved stress integration scheme (GI or NCI) is implemented. Roughly speaking, the integration points are in this case moved ahead of the crack tip position, and the integrals for the cohesive model are then calculated on the uncracked domain restriction of the finite element traversed by the crack front. Even in this case, the outcome is the relaxation of the restraint on the minimum mesh size with respect to the cohesive zone.

To mitigate the oscillations due to unbalanced nodal forces upon the insertion of a new cohesive segment, Menouillard and Belytschko [10], for extended finite elements (XFEM), proposed to add artificial correction nodal forces. In their correction method, to reduce the jump due to the additionally injected enriched dofs, fictitious forces are added to the interested nodes to re-establish equilibrium. These forces are then scaled down when the crack tip reaches the next element edge in the propagation path.

It is well known that in finite element cohesive propagation, the cohesive energy is the critical parameter to be respected in the modeling of the cohesive properties, whereas the cohesive strength, the limit opening and the exact shape of the cohesive traction-separation law have a far smaller impact on accuracy. The adaptive definition of these latter parameters has also been proposed as a technique to improve accuracy in the presence of coarse finite element discretizations of the process zone [17, 16]. In view of these considerations, a correction similar to the one proposed in [9, 10] is considered in this work. The required correction force is calculated at the node in correspondence of the newly separating element crack tip, so as to compensate the missing force contribution coming from the adjacent element before separation. The additional contribution is however provided by a modification of the maximum traction and limit opening in the cohesive law, while preserving the cohesive energy.

The correction has been implemented in an explicit dynamics solid-shell finite element code

together with the DCE technique. The effect on accuracy of the number of introduced string elements and of the proposed correction is investigated by application to test cases taken from the literature. In the next section 2 the basics of the DCEs, the rules for the numerical integration leading to the cohesive forces for multiple cables, and the adopted correction scheme are briefly summarized, while in section 3 we exemplify through simple cases the effects of our choices. Our conclusions are drawn in section 4. In all the paper, bold characters are reserved for vectors or matrices.

2 FORMULATION

Let us consider a continuum with the thickness smaller than the other dimensions. The body occupies a volume Ω_0 in a reference configuration $\mathcal{B}_0$; its boundary is called $\partial\Omega_0$. The map of the deformation consequent to a change of configuration is $\boldsymbol{\chi}(\mathbf{X}, t)$, $t \in [0, T]$, $\mathbf{X}$ being the original coordinates of a generic point in Ω_0, and T the time duration of of the process. After the change, the body occupies the volume Ω in the current deformed configuration $\mathcal{B}$. A cohesive crack through the thickness is described via the original interface Γ_0, which evolves into Γ because of the configuration change. Across the surface Γ, displacement discontinuities are defined as $\boldsymbol{\delta} = \mathbf{x}^+ - \mathbf{x}^-$, being $\mathbf{x}^+$ and $\mathbf{x}^-$ the current coordinates of the material points belonging to the flank surface originally linked and then separating into Γ^+ and Γ^-. A variational balance in dynamics, neglecting damping, accounting for the change in kinetic, internal energy and external work is written as:

$$\delta\Pi^e_{\text{kin}} + \delta\Pi^e_{\text{int}} - \delta\Pi^e_{\text{ext}} = \left[\int_\Omega \rho\, \ddot{\mathbf{u}} \cdot \delta\mathbf{u}\, \mathrm{d}\Omega\right] + \left[\int_{\Omega_0 \setminus \Gamma_0} \mathbb{S} : \delta\mathbb{E}\, \mathrm{d}\Omega_0 - \int_\Gamma \mathbf{T} \cdot \delta\boldsymbol{\delta}\, \mathrm{d}\Gamma\right] - \delta\Pi^e_{\text{ext}} = 0 \quad (1)$$

where $\mathbb{F}$ is the deformation gradient, $\mathbb{E}$ is the Green-Lagrange strain tensor, $\mathbb{S}$ is the second Piola-Kirchhoff stress tensor, ρ is the material mass density, $\mathbf{u}$ are the displacements, $\ddot{\mathbf{u}}$ the accelerations, $\mathbf{T} = \sigma\mathbf{m}$ is the Cauchy traction vector on the interface (σ is the Cauchy stress tensor and $\mathbf{m}$ is the normal to the flank surface).

A solid-shell finite element discretization is adopted, by using the element proposed in [15] and successive modifications (in particular we use the selective mass scaling described in [3] and [4]). In the paper we assume the following numbering for the reference element: since the thickness is small, it is always possible to identify nodes 1-4 for the lower surface, and nodes 5-8 for the upper surface. By gathering the nodal coordinates in two vectors $\mathbf{X}_{1-4}$ and $\mathbf{X}_{5-8}$, it is possible to define as a corner fiber the segment connecting two corresponding nodes (one at the lower and one at the upper surface), henceforth coinciding with the element smaller edges. Fiber mid-points are therefore defined as:

$$\mathbf{X}^m = \frac{\mathbf{X}_{1-4} + \mathbf{X}_{5-8}}{2} \quad (2)$$

The DCE approach is based on the introduction, when a fracture criterion is satisfied at a solid-shell finite element node, of a “cable” element either between two separating nodes (one of which is a duplicate of the original) or in correspondence of integration points along each one of the two crack flank middle surfaces (see figure). The DCE, or “cable”, element can interact with a blade, and transfer the contact forces to the opening faces in the correct direction, as shown in figure 1.

2.1 Cohesive force calculation

The contribution to the internal energy of the cohesive forces acting on a separating element side can be expressed in terms of the nodal displacement jumps as

$$\delta\Pi^e_{\text{coh,int}} = \int_{-1}^{+1}\int_{-1}^{+1} \mathbf{T}\cdot\delta\left[\sum_{a=1}^{4} N_a(\theta,\zeta)\,\boldsymbol{\delta}_a\right]^e J_\Gamma\,\mathrm{d}\theta\,\mathrm{d}\zeta \tag{3}$$

where $\mathbf{T}$ is the traction vector obtained from the cohesive law, J_Γ is the Jacobian determinant of the area transformation between the original and master reference system on the crack surface, θ and ζ are the intrinsic coordinates in the master reference system, along the crack direction and along the thickness direction, respectively. Therefore, $N_a(\theta,\zeta)$ represents the shape function of node a restricted to the element face corresponding to the crack surface.

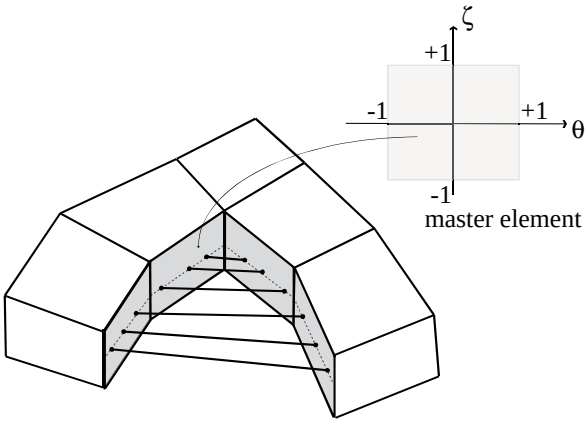


Figure 2: master element reference system for the DCEs inserted at the opening flank surfaces.

Cohesive forces are transmitted to the element side through a discrete number N of of DCEs, introduced in correspondence of the adopted integration points. Only one integration point at $\zeta = 0$ is assumed (i.e. in correspondence of the middle surface in the thickness direction) in view of the assumed small thickness, while along θ different choices are considered, as summarized in table 1. The implementation used in [6, 12] connects the DCEs to the fiber nodes at the crack flanks. This can be interpreted as adopting a NCI with a trapezoidal rule. In contrast, in this paper several alternatives for GI are explored. The contribution of the cracked element side to the cohesive internal work can be henceforth written as (without summation on repeated indices):

$$\delta\Pi^e_{coh,int} = \sum_{a=1}^{4} \mathbf{f}^e_a\cdot\delta\boldsymbol{\delta}^e_a = \sum_{a=1}^{4}\sum_{k=1}^{N} w_k J^e_{\Gamma_k}\mathbf{T}^e_k\,N_a(\theta_k,\zeta=0)\cdot\delta\boldsymbol{\delta}^e_a \tag{4}$$

where $\mathbf{f}^e_a = \sum_{k=1}^{N} w_k J^e_{\Gamma_k}\mathbf{T}^e_k\,N_{a_k}$ are the equivalent cohesive forces at each node of the opening face, T_k is the cohesive traction transmitted by the cable at the k-th integration point, w_k is the integration weight, $N_a(\theta_k,\zeta=0)$ is the shape function of node a evaluated at the same position and $\delta\boldsymbol{\delta}_a$ is the virtual displacement jump at node a.

Table 1: Gaussian or Newton-Cotes integration rules considered in the paper. N is the number of integration points; θ_k, $k = 1, N$ are the master reference system element coordinates along the crack flank direction ($-1 < \theta < +1$); w_k are the ordered weights used for the numerical integration.

Integration rule	N	θ_k	w_k
NCI	2	$\pm$1	1,1
GI1	1	0	2
GI2	2	$\pm$0.577350269	1,1
GI3	3	$\pm$0.774596669,0	5/9, 8/9, 5/9
GI4	4	$\pm$ 0.861136311594053,	w_a, w_b, w_b, w_a
		$\pm$0.339981043584856	w_a=0.347854845137454,
			w_b=0.652145154862546
GI5	5	$\pm$0.906179845938664,	w_c, w_d, w_e, w_d, w_c
		$\pm$0.538469310105683, 0	w_c=0.236926885056189,
			w_d=0.478628670499366,
			w_e=0.56888888888889

The equivalent cohesive forces are then assembled together with the contributions of all the elements sharing the same node

$$\mathbf{f}_a = \boldsymbol{\mathcal{A}}\, \mathbf{f}_a^e = \boldsymbol{\mathcal{A}} \sum_{k=1}^{N} w_k J_{\Gamma_k} \mathbf{T}_k^e N_a(\theta_k, \zeta = 0) \tag{5}$$

being $\boldsymbol{\mathcal{A}}$ the assembly operator.

To guarantee force balance at the node at the time of node duplication, this force should be exactly equal to the force transmitted to node a by the element on the other side of the crack, before node duplication and separation. At the insertion of the DCE, the nodes of the opening flanks are duplicated but they are still geometrically coincident. In the original implementation of the directional cohesive elements [12], the cohesive strength was initialized to match the component of the Cauchy stress normal to the crack flank at the integration point, consistent with the assumed mode I dominated opening. However, this procedure in general does not respect the force balance at the node. The sudden release of the unbalanced internal nodal forces upon node separation leads to oscillations in the stress field. A correction procedure in a single explicit dynamic time step has been envisaged as follows.

When the fracture activation criterion is met at a node, the opening face is identified as the one normal is closest to the direction of maximum principal stress. The node itself and the face are then duplicated and the topology of the elements belonging to the node support is updated. A prescribed number of directional cohesive elements is introduced between the two opening faces, depending on the chosen integration scheme. As already noticed in [9] and in [10], the duplication of a node causes a spurious stress oscillation due to the element-wise description of the crack propagation and to the release of internal nodal forces, which is only partially balanced by the cohesive force transmitted by the cable elements. In the framework of the XFEM method, to avoid this unphysical effect Menouillard and Belytschko [9, 10] proposed to introduce a fictitious nodal correction force exactly equal to the unbalanced part and linearly decaying in time.

The proposal in this work is to recompute the maximum tensile stress of the newly added directional cohesive elements, such that the equilibrium is restored at the time t_{nd} of node duplication. Modifying this parameter has been shown [1, 17, 16] to not significantly alter the

overall mechanical response, since the dissipative process is mainly governed by the fracture energy. At time t_{nd}, the initial cable lengths are zero and the directional cohesive elements are assumed to transmit a cohesive traction directed as the normal $\mathbf{m}_k$ to the opening element side at the k-th integration point and of magnitude $\mathbf{T}_{0_k}$, being $\mathbf{T}_{0_k}$ the maximum tensile stress at point k (figure 3). Since at t_{nd} the two duplicated nodes have the same velocity and acceleration, there are no jumps in the inertia and viscous forces. As a consequence, the equilibrium of node a writes:

$$\mathbf{f}_a = \boldsymbol{\mathcal{A}} \sum_{k=1}^{N} w_k J_{\Gamma_k} \mathbf{T}_{0_k} N_{a_k} = (\mathbf{F}_a^{\text{int}} - \mathbf{F}_a^{\text{cohes}}) \cdot \mathbf{m} \tag{6}$$

where $\mathbf{F}_a^{\text{int}}$ and $\mathbf{F}_a^{\text{cohes}}$ represent the internal nodal force and the cohesive force due to previously existent cables, computed after node duplication.

The cohesive forces are normal to the flank surfaces at the beginning; however, when the flanks start separating, then the cable tilts and shear forces are transmitted together with the normal ones. Morevover, the way the cohesive forces are transmitted to the crack flanks depends on whether the blade is interacting or not with the DCE. When the generic cable interacts with the blade through contact, the "directional" concept for the force calculation is triggered (see figure 1): if $\mathbf{x}_k^c$ is the position of the contact point, $\square^+$ and $\square^-$ indicate quantities at each (middle plane) flank surface, then the effective crack opening at k is computed as

$$\ell_k = \left\|\mathbf{x}_k^c - \mathbf{x}_k^+\right\| + \left\|\mathbf{x}_k^c - \mathbf{x}_k^-\right\| = \ell_k^+ + \ell_k^-. \tag{7}$$

The directions of the cohesive tractions transmitted by the cable element are aligned with the lines joining the contact points $\mathbf{x}_C$ to the positions of the integration points on the flank surface, namely:

$$\mathbf{T}_k^+ = T_k \, \mathbf{m}_k^+ \qquad \mathbf{T}_k^- = T_k \, \mathbf{m}_k^- \tag{8}$$

where T_k is the modulus of the cohesive traction computed through the cohesive law from the effective jump ℓ_k and

$$\mathbf{m}_k^+ = \frac{\mathbf{x}_k^c - \mathbf{x}_k^+}{\left\|\mathbf{x}_k^c - \mathbf{x}_k^+\right\|} \qquad \mathbf{m}_k^- = \frac{\mathbf{x}_k^c - \mathbf{x}_k^-}{\left\|\mathbf{x}_k^c - \mathbf{x}_k^-\right\|}. \tag{9}$$

In these expressions, it is assumed that all the positions $\mathbf{x}_k$ are given, since they derive from the integration of the equations of motion.

In the absence of contact with a blade, the cohesive traction is simply:

$$\mathbf{T}_k = \mathbf{T}_k^+ = -\mathbf{T}_k^- = -\frac{\|\mathbf{T}_k\|}{\|\boldsymbol{\delta}_k\|} \boldsymbol{\delta}_k. \tag{10}$$

A linear softening law, as depicted in figure 3, rules the magnitude of the cohesive force transmitted by the string element at the two opening faces as a function of the cable length. The area beneath the curve represents the fracture energy $\mathcal{G}_c = \frac{1}{2} T_0 \ell_c$, being ℓ_c the critical length corresponding to the complete decohesion of the two flanks and T_0 being determined from (6).

3 NUMERICAL EXAMPLES

3.1 Three point bending specimen

Let us consider a three point bending specimen [5, 2], whose dimensions are shown in figure 4. Several discretizations based on solid-shell elements, with out-of-plane thickness B =5

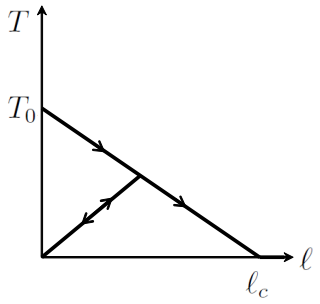


Figure 3: linear softening law adopted for the DCE.

mm, are explored, each one identified by a vertical dimension of the finite element in the set $h_e = \{7.5, 10, 15, 25, 30\}$ mm constant along the beam height (in figure 4 the case $h_e = 30$ mm is depicted); the finite element width is instead always fixed to 10 mm. The material properties are taken from [5] and are the following: Young's modulus $E = 36,500$ MPa, Poisson's ratio $\nu = 0.1$, cohesive strength $T_0 = 3.19$ MPa and mode I fracture energy $\mathcal{G}_c = 0.05$ N/mm. With these parameters the characteristic length of the cohesive process zone is $\ell_{pz} = E\,\mathcal{G}_c/T_0^2 = 179$ mm, therefore larger than the beam height. Consequently, every case discussed in this section largely satisfies the condition on the minimum number of integration points in the cohesive process zone, e.g. a minimum of 3-5 linear finite elements. The analyses are carried out using an explicit dynamics approach in displacement control, with the top middle point of the beam moving downwards with a constant velocity of 2 mm/s. A Rayleigh damping has been added to limit the oscillations due to inertial effects in the bulk material and to mimic a quasi-static response.

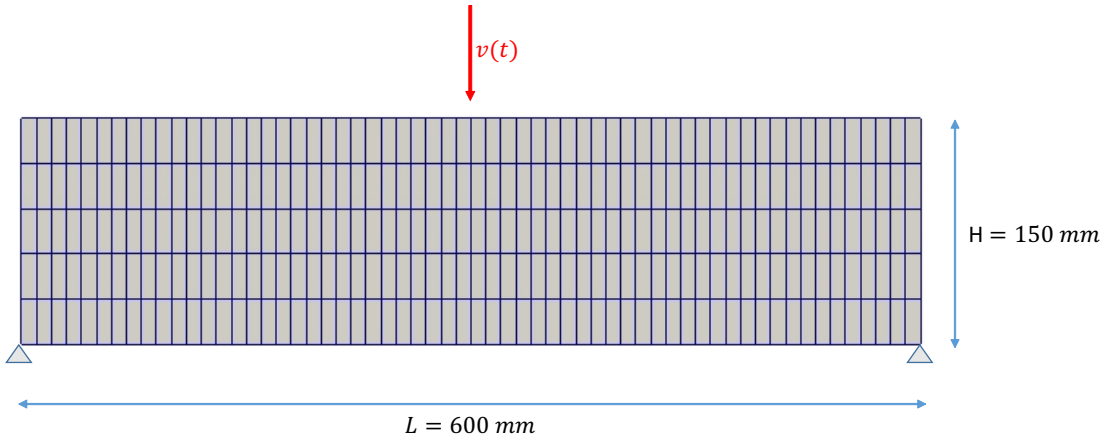


Figure 4: three point bending specimen, (coarser) mesh size h_e=30 mm.

To evaluate the influence of the mesh size, the results are shown in figure 5 in terms of dimensionless load-displacement curves and accelerations vs the dimensionless vertical, imposed

displacement. In these examples, only one cable is inserted between the opening faces. It can be seen that in all the cases a sudden decrease of the load corresponds to the snap-back experimental behaviour. The mesh sizes $h_e = 25, 30$ mm show a pathological behaviour in the post-peak regime, partially explained by the damping factor adopted. The accelerations reported in the graph at the bottom in figure 5, are strongly influenced by the discretizations in space and time. However, the steep peaks registered have different interpretation: before the maximum load value they indicate the opening of the first crack faces and cable insertion during loading; then there are acceleration peaks, despite damping, in correspondence of the large load drop because of the structural snap-back, whose disturbance decreases in a not negligible time interval; in the final, descending part of the curves, the accelerations of the node at the crack mouth are related, instead, to the opening of the elements along the ligament, before the crack reaches the last element at the specimen top (in the model, this element, where the displacement is imposed, cannot be broken). When a DCE reaches its maximum elongation ℓ_c and disappears, there are no appreciable disturbances, since the exchanged cohesive forces in that case are already almost zero.

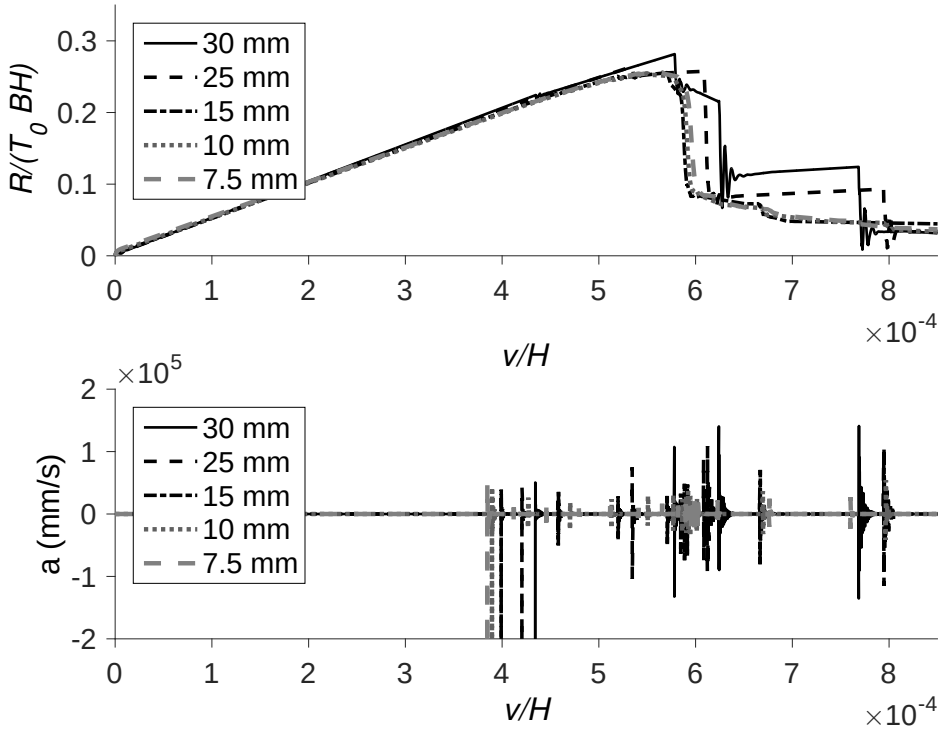


Figure 5: three point bending specimen. Comparison in terms of load displacement response and acceleration at the crack mouth between different meshes (one cable per element case).

To evaluate the influence of the number of the DCEs inserted for each solid-shell element opening face, we present in figure 6 the cases $h_e = 15, 30$ mm. Since the test is carried out with displacement control, the sudden load drop can be well reproduced even with the coarse discretization if two cables per opening face are inserted. While not reported here, the addition of more cables does not improve the response, since the element size is sufficiently smaller than the cohesive length as in this case. Since the opening face can change its configuration

according to the displacements at its nodes and the cables are inserted at the face middle plane, the crack opening profile is linear and the transmitted cohesive traction is already accurately described by two cables.

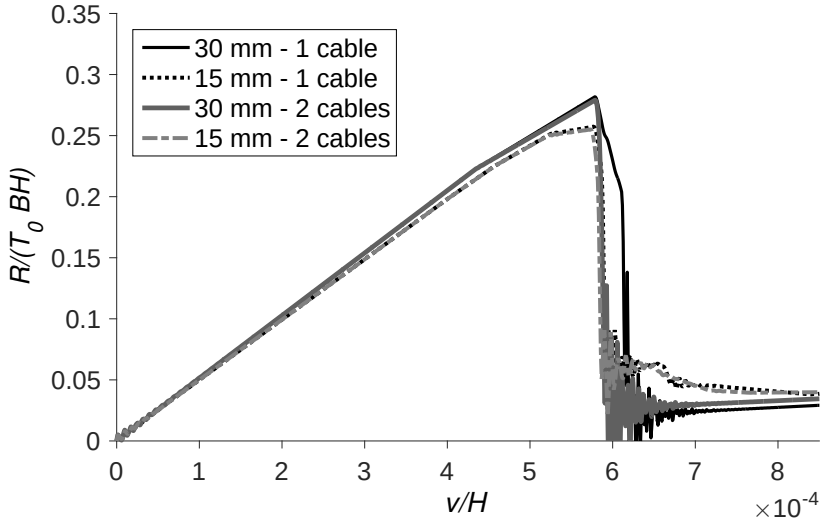


Figure 6: Three point bending specimen. Comparison of load displacement curves for different number of cables, $h_e = 15$ mm.

All the previous results have been obtained using the correction approach defined in (6). To point out its role, in figure 7 we compare the outcomes with and without correction for the case $h_e = 15$ mm. It can be noticed that the introduction of the correction procedure allows to obtain a smooth response, avoiding the unphysical jumps that can be observed in the curve without correction. The entity of the correction is slightly different for each DCE along the crack propagation direction: the maximum correction value for T_0 is an increasing of about 40% for the one cable case; the entity of the correction, however, is smaller when the number of the cables is increased.

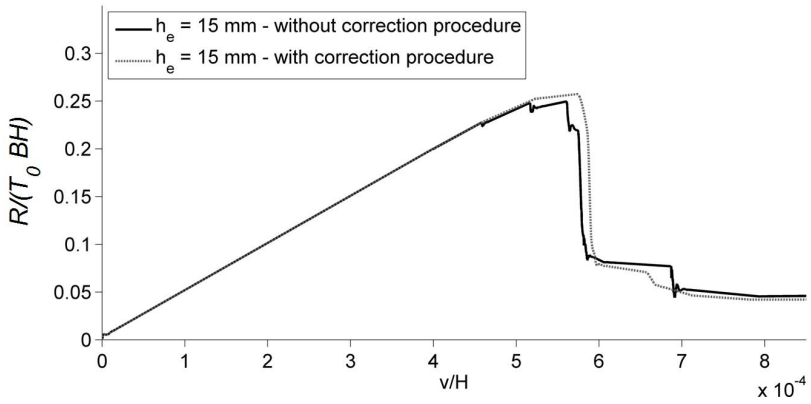
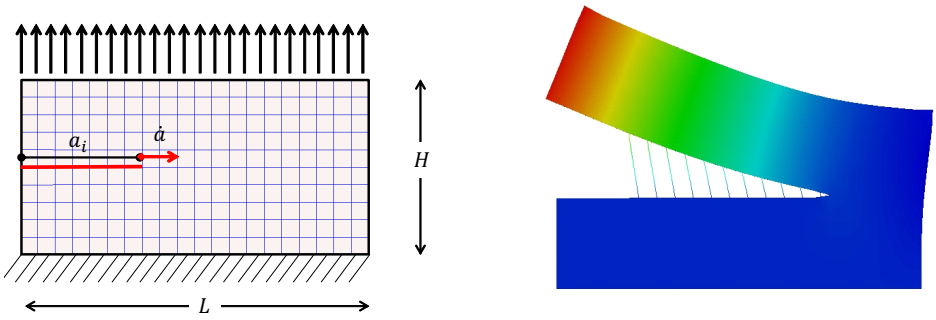


Figure 7: Three point bending specimen. Comparison of load displacement curves for different number of cables, $h_e = 15$ mm.

3.2 Coarse regular mesh with a rectilinear crack

In this section we adapt the example from [10], where brittle fracture was considered (no cohesion), to the present DCE approach with solid-shell elements in explicit dynamics, with the proposed correction of the cohesive strength, according to equation (6). A rectilinear crack is propagated in a coarse and regular mesh with an imposed advancing crack tip velocity of 1 m/s (figure 8a). The bottom nodes of the model are restrained, while at the top edge a constant stress $\sigma = 10$ kPa is imposed. Dimensions are the following: length of the specimen $L = 0.2$ m, width $H = 0.1$ m, initial crack length $a_i = 0.04$ m. The material properties for the bulk are linear elastic, with Young's modulus $E = 1$ MPa, Poisson's ratio $\nu = 0.3$, and mass density $\rho = 1$ kg/m^3. The cohesive strength is $T_0 = 10$ kPa, and $\mathcal{G}_c = 0.05$ N/mm. The damping parameter is set to 2%.



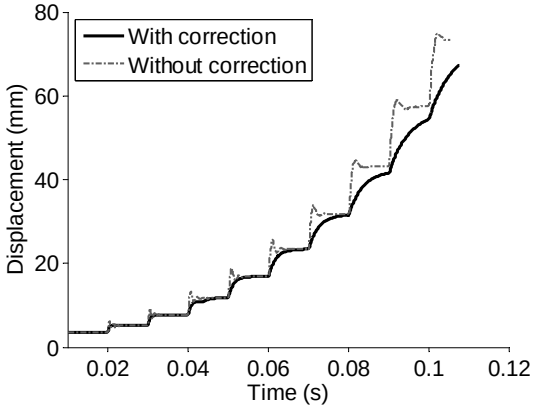
(a) (b)

Figure 8: rectilinear crack propagation. (a) geometry, load and boundary conditions and (b) deformed shape.

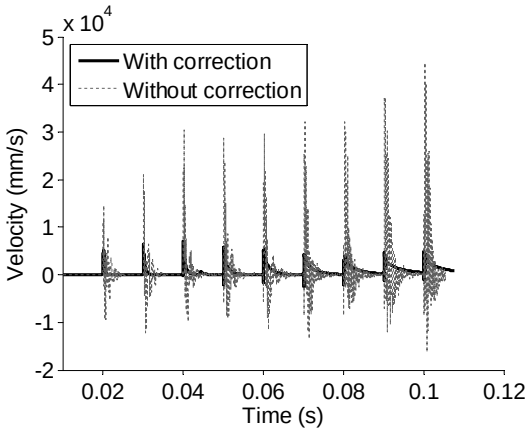
Spurious oscillations at the top left corner node, induced by the sudden insertion of the cohesive elements at the nodes of the element near the crack tip, are monitored. In figure 9a the displacement vs time response appears smoothed when the correction is activated; it is instead strongly discontinuous without correction. The effect of the correction becomes more evident as the time increases, i.e. when the crack is far from the top left vertex node during the propagation. It should be emphasized the the different response is obtained using in the two analyses cohesive models with the same fracture energy. The only difference is the initial value of the traction T_0 that in the case without correction is taken equal to the normal component of Cauchy stress at the time t_{nd} of node duplication, while in the case of the correction is determined according to (6). In figure 9b the velocity at the same point is depicted: the correction decreases the peaks also in this case. It is worthwhile to emphasize that, without correction, the induced velocities appear completely not physical; when we include the correction, the values turn out to be much smaller and acceptable.

4 CONCLUSIONS

In this paper some progresses on the development of the directional cohesive finite element approach have been presented, including: i) the possibility to increase the number of inserted cables for each opening face of solid-shell finite elements, and ii) corrected definition of maximum cohesive tractions based on nodal equilibrium to avoid the spurious oscillations in the local stress field arising from the insertion of cohesive finite elements at the crack tip. These numerical improvements can be helpful during the simulation of sharp blades cutting through



(a)



(b)

Figure 9: rectilinear crack propagation. Comparisons for corrected and not corrected responses in terms of (a) vertical displacement and (b) velocity at the top left corner vertex.

layered shells, where the cohesive zone length is of the order of the curvature radius of the cutting blade. We showed, through numerical examples taken from the literature, how different choices influence the numerical results; in particular, the beneficial role of corrected nodal cohesive forces to smooth numerical oscillations at the insertion of cohesive elements in explicit dynamics problems is emphasized.

ACKNOWLEDGEMENTS

The financial support by Tetra Pak Packaging Solutions is kindly acknowledged.

REFERENCES

[1] G. Alfano and M. A. Crisfield. Finite element interface models for the delamination analysis of laminate composites: mechancial and computational issues. *International Journal for Numerical Methods in Engineering*, 50:1701–1736, 2001.

[2] A. Carpinteri and G. Colombo. Numerical analysis of catastrophic softening behaviour (snap-back instability). *Computers & Structures*, 31:607 – 636, 1989.

[3] G. Cocchetti, M. Pagani, and U. Perego. Selective mass scaling for distorted solid-shell elements in explicit dynamics: optimal scaling factor and stable time step estimate. *International Journal for Numerical Methods in Engineering*, 101(9):700–731, 2015. nme.4829.

[4] F. Confalonieri, A. Ghisi, and U. Perego. 8-node solid-shell elements selective mass scaling for explicit dynamic analysis of layered thin-walled structures. *Computational Mechanics*, 56(4):585–599, 2015.

[5] B.C. Do, W. Liu, Q.D. Yang, and X.Y. Su. Improved cohesive stress integration schemes for cohesive zone elements. *Engineering Fracture Mechanics*, 107:14 – 28, 2013.

[6] A. Frangi, M. Pagani, U. Perego, and R. Borsari. Directional Cohesive Elements for the Simulation of Blade Cutting of Thin Shells. *Computer Modeling in Engineering & Sciences*, 57:205–224, 2010.

[7] A Giampieri and U Perego. An interface finite element for the simulation of localized membrane-bending deformation in shells. *Computer Methods in Applied Mechanics and Engineering*, 200:2378–2396, 2011.

[8] P. Gilormini and J. Diani. Testing some implementations of a cohesive-zone model at finite strain. *Engineering Fracture Mechanics*, 148:97 – 109, 2015.

[9] T. Menouillard and T. Belytschko. Correction force for releasing crack tip element with xfem and only discontinuous enrichment. *European Journal of Computational Mechanics*, 18(5-6):465–483, 2009.

[10] T. Menouillard and T. Belytschko. Smoothed nodal forces for improved dynamic crack propagation modeling in XFEM. *International Journal for Numerical Methods in Engineering*, 84:47–72, 2010.

[11] M. Ortiz and A. Pandolfi. Finite-defomation irreversible cohesive elements for three-dimensional crack propagtion analysis. *International Journal for Numerical Methods in Engineering*, 44:1267–1282, 1999.

[12] M. Pagani and U. Perego. Explicit dynamics simulation of blade cutting of thin elasto-plastic shells using a "directional" cohesive elements in solid-shell finite element models. *Computer Methods in Applied Mechanics and Engineering*, 285:515 – 541, 2015.

[13] H. Qiao, W. Q. Chen, Q.D. Yang, and J. Lua. Augmented Cohesive Elements for Efficient Delamination Analyses of Composite Laminates. *Journal of Engineering Materials and Technology*, 133:041010–1, 2011.

[14] J. C. J. Schellekens and R. De Borst. On the numerical integration of interface elements. *International Journal for Numerical Methods in Engineering*, 36:43–66, 1993.

[15] M. Schwarze and S. Reese. A reduced integration solid-shell finite element based on the eas and the ans concept: large deformation problems. *International Journal for Numerical Methods in Engineering*, 85(3):289–329, 2011.

[16] A. Turon, P. P. Camanho, J. Costa, and J. Renart. Accurate simulation of delamination growth under mixed-mode loading using cohesive elements: Definition of interlaminar strengths and elastic stiffness. *Composite Structures*, 92(8):1857–1864, 2010.

[17] A. Turon, C. G. Dávila, P. P. Camanho, and J. Costa. An engineering solution for mesh size effects in the simulation of delamination using cohesive zone models. *Engineering Fracture Mechanics*, 74(10):1665–1682, 2007.

[18] Q. D. Yang, X. J. Fang, J. X. Shi, and J. Lua. An improved cohesive element for shell delamination analyses. *International Journal for Numerical Methods in Engineering*, 83(5):611–641, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EVOLUTIONARY LEVEL SET METHOD FOR CRASHWORTHINESS TOPOLOGY OPTIMIZATION

Mariusz Bujny[1], **Nikola Aulig**[2], **Markus Olhofer**[2], **Fabian Duddeck**[1]

[1]Technische Universität München
Arcisstraße 21, 80333 Munich, Germany
e-mail: {mariusz.bujny, duddeck}@tum.de

[2] Honda Research Institute Europe GmbH
Carl-Legien-Straße 30, 63073 Offenbach/Main, Germany
e-mail: {nikola.aulig, markus.olhofer}@honda-ri.de

Keywords: Level Set Method, Evolutionary Algorithms, Topology Optimization, Crashworthiness.

Abstract. *Vehicle crashworthiness design belongs to one of the most complex problems considered in the design optimization. Physical phenomena that are taken into account in crash simulations range from complex contact modeling to mechanical failure of materials. This results in high nonlinearity of the optimization problem and involves remarkable amount of numerical noise and discontinuities of the objective functions that are optimized. Consequently, the sensitivity information, which is necessary for the majority of Topology Optimization approaches, can be obtained analytically only for considerably simplified problems, which, in most cases, excludes the use of the gradient-based optimization methods. As a result, in the state-of-the-art methods for crashworthiness Topology Optimization, strong and thus arguable assumptions about the properties of the optimization problem are made and heuristic approaches are used. This problem can be solved with use of Evolutionary Algorithms, where no assumptions about the optimization problem have to be made and which perform well even for highly nonlinear and discontinuous problems. We propose a novel approach using evolutionary optimization techniques together with a geometric Level-Set Method in crashworthiness Topology Optimization. Both standard Evolution Strategies and the state-of-the-art Covariance Matrix Adaptation Evolution Strategy are used. In order to evaluate the proposed method, an energy maximization problem for a rectangular beam, fixed at both ends and impacted in the middle by a cylindrical pole, is considered. The results show that the evolutionary optimization methods can be efficiently used for an optimization of crash-loaded structures, while defining the objective function explicitly.*

1 INTRODUCTION

In the recent years, a lot of effort has been put in the automotive industry to shorten product cycles as much as possible. This is one of the most crucial demands of the car market that is also pushing car companies to broaden the offer of their products and increase the efficiency of development and production. Also the complexity of the vehicles produced today is increasing rapidly, which, together with rising material costs and strict CO_2 emission reduction targets for new cars, forces car companies to use numerical simulation and optimization in the vehicle design. The number of factors to be taken into account and fields to analyze is rising constantly, as well. However, crashworthiness optimization, due to its complexity, poses the main difficulty and affects the character of the whole vehicle design process [11].

The wide spectrum of physical phenomena incorporated in crash simulations, ranging from complex contact modeling to mechanical failure of materials, results in high nonlinearity of the optimization problem, numerical noise and discontinuities of the optimized objective functions. As a result, in general case, analytical gradients are not available and the gradient-based optimization methods cannot be applied. Therefore, development of alternative approaches for crashworthiness optimization is necessary.

The main focus of this paper is crashworthiness Topology Optimization, which is used to determine the best structural concept at early stages of the vehicle development process. Due to the complexity of crash phenomena mentioned above, in the state-of-the-art methods for crashworthiness Topology Optimization, very strong assumptions about the properties of the optimization problem are made and heuristic approaches are used to optimize structures. In the Equivalent Static Loads Methods [6, 7, 12, 29], equivalence of static and dynamic loads is assumed. In the Ground Structure Approaches [14, 31], simplified crash models, involving considerable calibration effort, are used. Graph and Heuristic approaches use rules derived from the expert knowledge [28]. Finally, both in the Hybrid Cellular Automata [30, 25] and the Hybrid Cellular Automata for Thin-Walled Structures [13, 22] homogeneity of energy density either all over the structure or in larger macro-structures is required. As a result, in each of those methods, the assumptions are arguable and convergence to the true optima cannot be guaranteed.

An alternative approach is to use evolutionary optimization methods, which are gradient-free and work very well even for highly nonlinear, noisy, discontinuous problems. In evolutionary methods, the optimization process is carried out solely through evaluating values of the objective function in different points of the design space, thus no additional assumptions have to be made. Therefore, the objective function can be precisely defined and solutions are exclusively judged by their objective values.

In this paper, we propose a novel approach for crashworthiness Topology Optimization, which uses an implicit parametrization of mechanical structures with geometric level-set functions [15]. In the Level-Set Methods [10], the material interface is precisely defined by the iso-contours of the level-set function, which is a crucial property from the point of view of manufacturability and the accuracy of crash simulations [5], where the presence of intermediate densities in the finite element model might lead to severe deviations from the real crash behavior. Moreover, by introducing the parameterization with the geometric level-set functions [15], the number of design variables can be reduced significantly, while allowing for a relatively high flexibility of topological changes. As a result, Evolutionary Algorithms, whose numerical costs highly depend on the dimensionality of the optimization problem, can be used efficiently for this type of parametrization.

Evaluation and validation of the proposed method is realized for a simple, 2D crash case, where a cylindrical pole impacts a clamped structure defined within a rectangular design domain. An energy maximization problem with mass constraint is considered. The optimization is carried out with use of standard Evolution Strategies (ES) [4] and the state-of-the-art Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [17]. Finally, performance of the structures optimized with use of the above mentioned methods and the Hybrid Cellular Automata technique is compared.

The paper has the following structure. Section 2 formulates the optimization problem and presents an overview of the optimization methods used in this research. In Section 3 the implicit parametrization with the geometric level-set functions is described. The experimental setup, optimization results and a corresponding discussion is presented in Section 4. The final conclusions are described in Section 5.

2 OPTIMIZATION PROBLEM

2.1 Problem formulation

In this paper, the problem of energy absorption E_{abs} maximization[1] at a given point of a crash event and under mass constraint is considered. Similar optimization problems were investigated by Hunkeler [22] and Aulig et al. [3]. In the most general form, the optimization problem can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{z}} \left(-E_{abs}(\mathbf{z})\right) \\ \text{s.t.} : \mathbf{r}(t) = \mathbf{0}, \\ m(\mathbf{z}) \leq m_{req}, \end{aligned} \tag{1}$$

where the residual $\mathbf{r}(t) = \mathbf{0}$ expresses the dynamic equilibrium at time t, m is mass of the structure and m_{req} is the required mass of the final design. The vector of design variables is denoted by $\mathbf{z}$.

Since this is a constrained optimization problem, an appropriate constraint-handling technique has to be chosen. In the field of evolutionary optimization, the most commonly used approaches for constrained optimization are the exterior penalty methods [9, 32]. For simplicity, static penalties [24] were used in this work, for both standard ES and the CMA-ES[2]. In the context of evolutionary computation a minimum of so-called fitness function has to be found. For the given problem, after introducing the penalty term, it takes the following form:

$$f(\mathbf{z}) = -E_{abs}(\mathbf{z}) + c \cdot \max\left(0, m(\mathbf{z}) - m_{req}\right), \tag{2}$$

where c is a weighting coefficient for the mass constraint.

2.2 Optimization methods

Biologically-inspired evolutionary optimization methods turned out to be very successful in many engineering applications [8, 27, 37, 21]. In particular, they showed very good capabilities

[1] Equivalent to minimization of negative energy absorption.

[2] Very little research on constraint handling for the CMA-ES has been done so far [23, 2] and behavior of the algorithm for more sophisticated constraint-handling techniques is not known. Therefore, the use of simpler approaches is favored.

for solving multimodal, noisy and discontinuous optimization problems. This makes their application to the crashworthiness Topology Optimization very promising. Below, a short overview of the evolutionary methods used in this paper is presented.

2.2.1 Evolution Strategy

Evolution Strategies were developed in Germany by the research group of Ingo Rechenberg and Hans-Paul Schwefel [33, 34]. Compared to genetic algorithms, the primary emphasis in Evolution Strategies lies on mutation instead of recombination. Most of evolution strategies are based on the following core structure:

t:=0;
initialize: $P(0) := \{\mathbf{a}_1(0), ..., \mathbf{a}_\mu(0)\} \in I^\mu$;
evaluate $P(0) : \{f(\mathbf{a}_1(0)), ..., f(\mathbf{a}_\mu(0))\}$;
while $(\iota(P(t)) \neq true)$ **do**
  recombine: $P'(t) := r_{\Theta_r}(P(t))$;
  mutate: $P''(t) := m_{\Theta_m}(P'(t))$;
  evaluate $P''(t) : \{f(\mathbf{a}''_1(t)), ..., f(\mathbf{a}''_\lambda(t))\}$;
  select: $P(t+1) := s_{\Theta_s}(P''(t))$;
  t:=t+1;
end

Algorithm 1: Standard Evolutionary Algorithm [4].

where $f : I \to \mathbb{R}$ denotes the fitness function to be minimized (I is the genotype space) and $\mathbf{a} \in I$ is an individual. The size of the parent and offspring population are denoted by $\mu \geq 1$ and $\lambda \geq \mu$, respectively. A population at generation t comprises all the parent individuals, i.e. $P(t) := \{\mathbf{a}_1(t), ..., \mathbf{a}_\mu(t)\}$. The recombination operator is a mapping $r_{\Theta_r} : I^\mu \to I^\lambda$, whereas mutation operator: $m_{\Theta_m} : I^\lambda \to I^\lambda$. Both recombination and mutation are controlled by sets of operator parameters: Θ_r and Θ_m. The selection operator defined as $s_{\Theta_s} : I^\lambda \to I^\mu$ is used to choose the individuals, which compose the parent population in the next generation.

The functioning of the algorithm can be characterized as follows. After the initialization and evaluation of the parent population, the main optimization loop begins. First of all, the new offspring is generated from the parent population in the recombination step. Recombination is usually performed both on design variables and strategy parameters. Depending on the type of recombination, it can be performed according to one of the following rules [4]:

$$z'_i = \begin{cases} z_{S,i} \text{ without recombination,} \\ z_{S,i} \text{ or } z_{T,i} \text{ discrete recombination,} \\ z_{S,i} + \chi \cdot (z_{T,i} - z_{S,i}) \text{ intermediate recombination,} \\ z_{S_i,i} \text{ or } z_{T_i,i} \text{ global, discrete recombination,} \\ z_{S_i,i} + \chi_i \cdot (z_{T_i,i} - z_{S_i,i}) \text{ global, intermediate recombination,} \end{cases} \tag{3}$$

where z_i is the i-th component of the vector of object variables, S and T denote two randomly selected parent individuals, whereas $\chi \in [0, 1]$ is a random variable from the uniform distribution. In case of the global recombination, parents S and T as well as the χ factor are chosen independently for each component of the $\mathbf{z}$ vector.

According to Bäck et al. [4], the best results were observed for discrete recombination on design variables and intermediate recombination on strategy parameters. Also historically, recombination in its intermediate and global form were applied for a constant value of $\chi = \frac{1}{2}$.

In the second step of the main optimization loop, the mutation operator is used. In the most general form, the mutation operator produces mutated individuals by deviating first the strategy parameters and then the design variables:

$$\begin{aligned} \sigma_i' &= \sigma_i \cdot \exp\left(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)\right) \\ \mathbf{z}' &= \mathbf{z} + \mathbf{N}\left(\mathbf{0}, \boldsymbol{\sigma}'\right), \end{aligned} \tag{4}$$

where τ' and τ are global and local learning rates, respectively.

For a single standard deviation for all object variables, (4) can be reduced to:

$$\begin{aligned} \sigma' &= \sigma \cdot \exp\left(\tau' \cdot N(0,1)\right) \\ \mathbf{z}' &= \mathbf{z} + \mathbf{N}\left(\mathbf{0}, \sigma'\right) \end{aligned} \tag{5}$$

According to Schwefel [36] a good choice of those parameters is the following:

$$\begin{aligned} \tau &= \frac{1}{\sqrt{2\sqrt{n}}} \\ \tau' &= \frac{1}{\sqrt{2n}} \end{aligned} \tag{6}$$

The mutation mechanism presented above allows the algorithm to evolve its own strategy parameters during the optimization process. As a result, it is often referred to as the "self-adaptation" mechanism and was first formulated by Schwefel [35].

Finally, after the evaluation step, the best individuals are chosen to form a new parent population. There are two main methods to do that. Either μ parents are selected from λ offspring individuals ((μ, λ)-ES) or the new parent population is selected out of parent and offspring populations combined ($(\mu + \lambda)$-ES). In this research, we use the (μ, λ)-ES since it can better deal with noisy quality evaluations [4].

2.2.2 Covariance Matrix Adaptation Evolution Strategy

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [19] is a derandomized Evolution Strategy, where the covariance matrix of the normal mutation distribution is adapted on the basis of the previous search steps [17]. This is a similar concept to the gradient-based quasi-Newton methods, where the Hessian matrix is estimated iteratively as the optimization process progresses. Initially, the method was designed for small populations and has proven to be a robust and efficient local search strategy [16]. Particularly, the CMA-ES can minimize efficiently unimodal functions [17]. The superiority of the method on non-separable and ill-conditioned problems and its applicability to real world problems has been also demonstrated [19]. An extension of the CMA-ES by the rank-μ-update [26, 20] allowed to use more effectively the information from large populations without influencing the performance when small populations are considered. As it is one of the state-of-the-art methods, its Python implementation [18] was used in this research, as well.

3 PARAMETRIZATION

Let us first define a global level-set function as follows:

$$\begin{cases} \Phi(\mathbf{x}) > 0, \mathbf{x} \in \Omega, \\ \Phi(\mathbf{x}) = 0, \mathbf{x} \in \partial\Omega, \\ \Phi(\mathbf{x}) < 0, \mathbf{x} \in D \backslash \Omega. \end{cases} \tag{7}$$

Analogically, we introduce a local level-set function of the i$^{\text{th}}$ elementary structural component having the following property:

$$\begin{cases} \phi_i(\mathbf{x}) > 0, \mathbf{x} \in \Omega_i \\ \phi_i(\mathbf{x}) = 0, \mathbf{x} \in \partial\Omega_i \\ \phi_i(\mathbf{x}) < 0, \mathbf{x} \in D \backslash \Omega_i \end{cases} \tag{8}$$

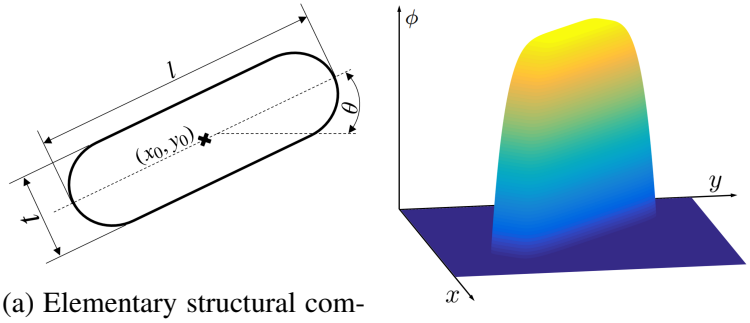
where Ω_i is a part of the design domain D occupied by an elementary component. As a result, the material phase is defined as:

$$\Omega = \bigcup_{i=1}^{e} \Omega_i, \tag{9}$$

where e denotes the number of elementary components. We introduce a level-set function (defining an elementary component) after Guo et al. [15], which for $D = \mathbb{R}^2$, $\mathbf{x} = (x, y)^T$, has the following form:

$$\begin{aligned} \phi_i(\mathbf{x}) = -\Bigg(\left(\frac{\cos\theta_i \left(x - x_{0i}\right) + \sin\theta_i \left(y - y_{0i}\right)}{l_i/2} \right)^q \\ + \left(\frac{-\sin\theta_i \left(x - x_{0i}\right) + \cos\theta_i \left(y - y_{0i}\right)}{t_i/2} \right)^q - 1 \Bigg), \end{aligned} \tag{10}$$

where (x_0, y_0) is the position of the center of a component (see Figure 1a) with length l and thickness t. The rotation angle of a component is denoted by θ. Similarly to the approach presented by Guo et al. [15], we take the exponent $q = 6$. The plot of the level-set function defined by (10) is presented in the Figure 1b.



(a) Elementary structural component.

(b) Level-set function.

Figure 1: Parametrization of the elementary structural component and the corresponding level-set function (where negative values are set to zero) [15].

As in the other Topology Optimization approaches [5], we use a standard density-based geometry mapping [10], where the dependency between the level-set field and the material density[3] $\rho(\mathbf{x})$ at position $\mathbf{x} \in D$ is given by:

$$\rho(\mathbf{x}) = H(\Phi(\mathbf{x})), \tag{11}$$

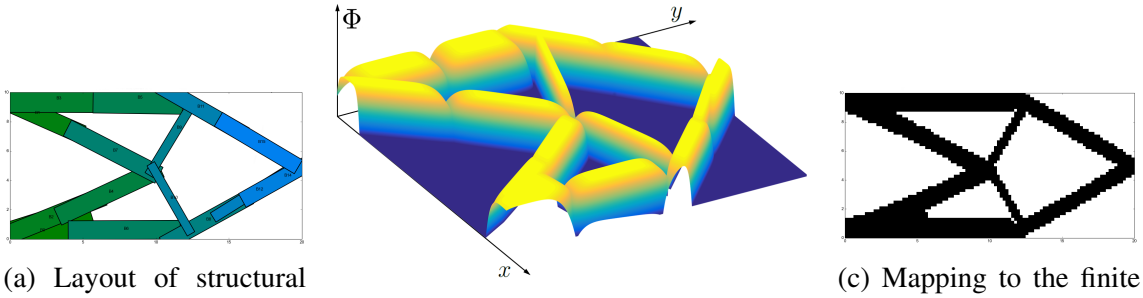
where:

$$\Phi(\mathbf{x}) = \max(\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), ..., \phi_m(\mathbf{x})), \tag{12}$$

and $H(x)$ denotes the Heaviside function:

$$H(x) = \begin{cases} 0, \text{ if } x < 0 \\ 1, \text{ if } x \geq 0. \end{cases} \tag{13}$$

Figure 2 presents a possible layout of structural components for the compliance minimization of the standard cantilever beam benchmark case [5]. Plots of the corresponding (global) level-set function and its mapping to the finite element mesh are shown, as well.



(a) Layout of structural components.

(b) Corresponding level-set field.

(c) Mapping to the finite element mesh.

Figure 2: Possible layout of structural components for the compliance minimization for the cantilever beam problem [5].

4 NUMERICAL EXPERIMENTS

4.1 Setup

As a test problem, topology optimization of a rectangular aluminum beam is considered. The beam is fixed at both ends and impacted in the middle by a cylindrical pole. Dimensions of the beam, as well as initial and boundary conditions are shown in Figure 3.

The LS-Dyna FEM mesh is composed of 1600 eight-node solid elements and the problem is considered as a 2D crash case (displacements of all nodes in the direction perpendicular to the cross-section shown in Figure 3 are set to 0). A piecewise linear elastic-plastic material model is used. Exact setup of the test case and material properties are given in Table 1.

The initial layout of structural components and the mapping of the corresponding level-set field to the LS-Dyna FEM mesh are shown in Figure 4.

[3] In the finite element discretization, $\rho(\mathbf{x}) = 0$ results in deletion of the finite element from the mesh.

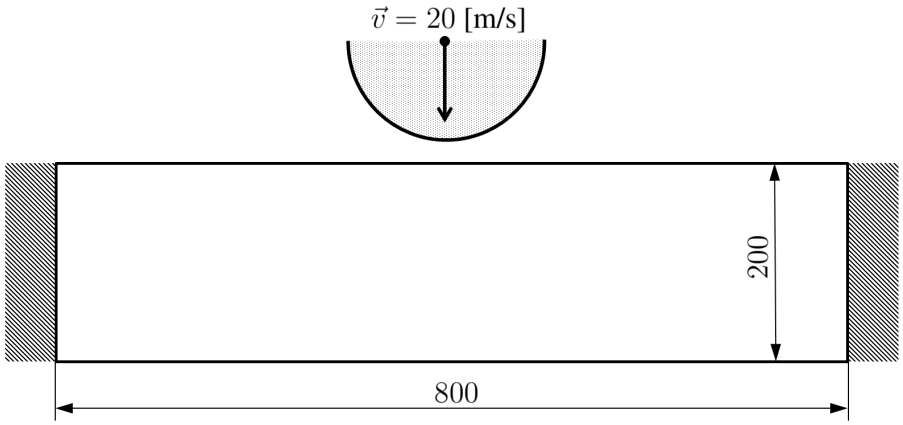
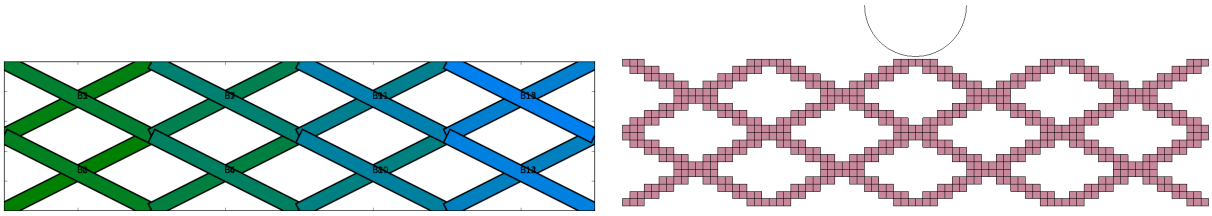


Figure 3: Design domain, initial and boundary conditions for the investigated crash case.

Property	Symbol	Value	Unit
Beam material density	ρ	$2.7 \cdot 10^3$	kg/m^3
Young's modulus	E	$7.0 \cdot 10^4$	MPa
Poisson's ratio	ν	0.33	-
Yield stress	R_e	241.0	MPa
Tangent modulus	E_{tan}	70.0	MPa
Friction coefficient	μ	0.1	-
Pole velocity	v	20	m/s
Pole mass	m_p	11.815	kg
Required structure mass	m_{req}	2.16	kg
LS-Dyna termination time	t_{end}	1.5	ms
LS-Dyna mesh resolution	-	80 x 20	-

Table 1: Configuration of the test case.



(a) Layout of structural components.

(b) Corresponding finite element mesh.

Figure 4: Initial layout of structural components and the corresponding LS-Dyna finite element mesh.

4.2 Results

For evaluation of the proposed topology optimization approach, 10 optimization runs of both the standard Evolution Strategy (ES) and the state-of-the-art Covariance Matrix Adaptation Strategy (CMA-ES) were completed. In each case, the optimizations were stopped after 1000 generations. For both algorithms an offspring population of the size of 17 and a parent

population consisting of 8 individuals[4] were used. In both cases, the initial value of σ was set to 0.1.

The results show that both ES and the CMA-ES can be successfully used for optimization of energy-absorbing structures. In Figure 5, averaged convergence of the fitness function for both ES and CMA-ES is presented. Figure 6 shows the statistical evaluation of both algorithms.

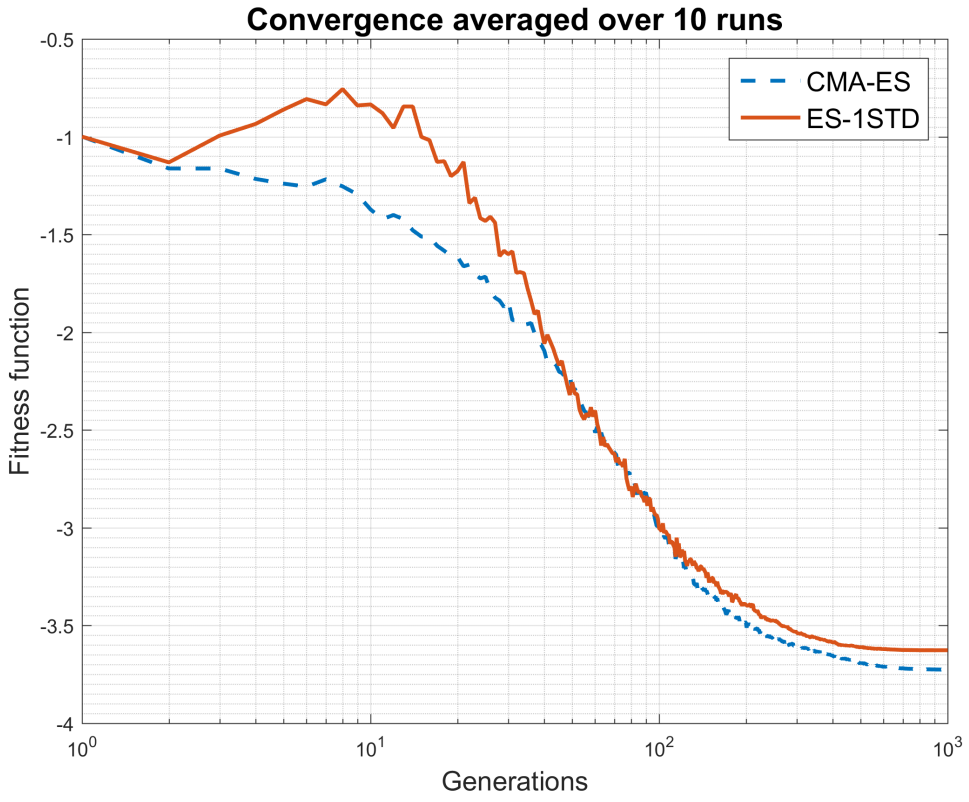


Figure 5: Convergence of the fitness function averaged over 10 runs of ES and CMA-ES.

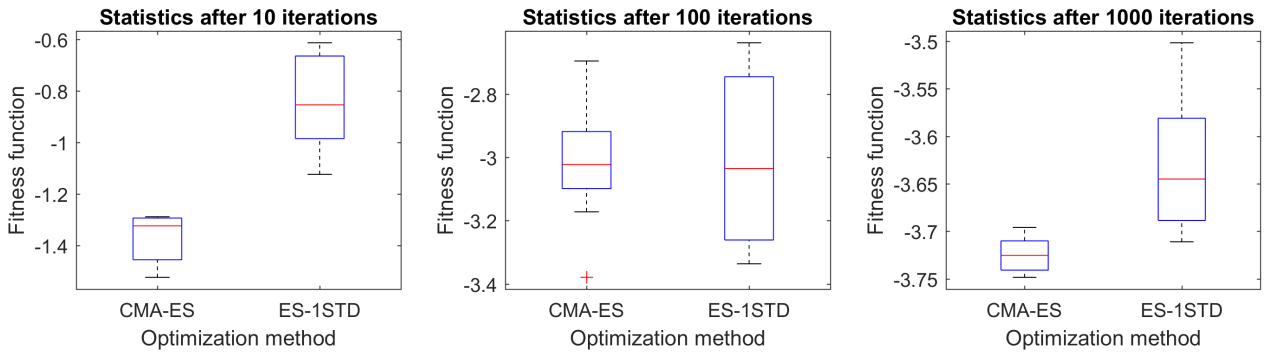


Figure 6: Box plots for 10 optimization runs of ES and CMA-ES after 10, 100 and 1000 generations, respectively.

[4]A default population size for a problem with 80 design variables, estimated and used by the Python implementation of the CMA-ES [18].

The CMA-ES exhibits considerably better performance at the beginning and the end of the optimization process. However, in the range especially interesting for practical applications (ca. 100 generations - equivalent to 1700 evaluations), the performance of both methods is similar. Nevertheless, in general, ES much more frequently results in inferior designs and is characterized by a relatively high variance of the fitness function.

The best designs out of 10 optimizations carried out with ES and CMA-ES are shown in Table 2. As a reference, a design obtained with the Hybrid Cellular Automata Technique is shown, as well.

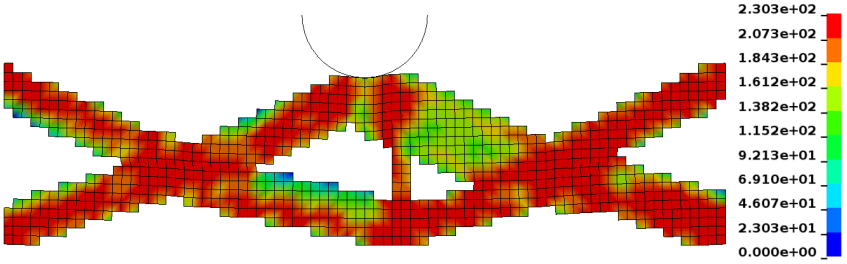
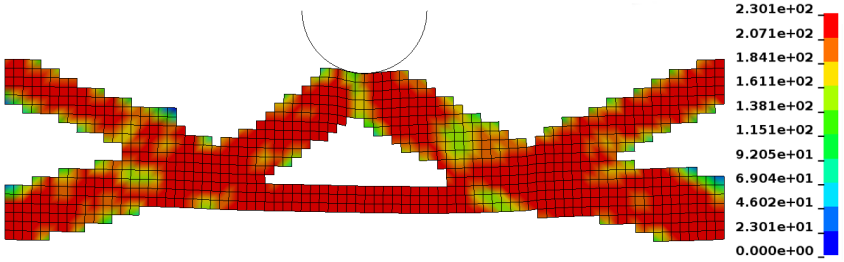
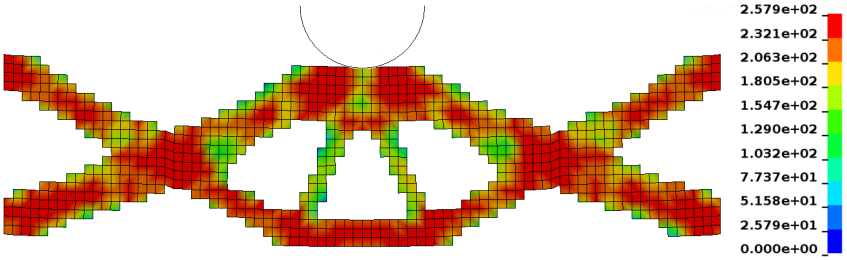
Method	**Effective stress field (von Mises) at the final time step** [MPa]	**Energy** [kJ]
ES-1STD		2.23
CMA-ES		2.25
HCA		2.24

Table 2: Best designs obtained within 10 optimization runs of ES and CMA-ES and a design optimized with the HCA method (with use of the LS-TaSC [1]). Von Mises stress field and energy absorption.

All of the designs exhibit similar performance with respect to energy absorption, but the design obtained with CMA-ES performs slightly better than the design obtained with the HCA method. This justifies the use of the proposed method, as a good alternative for the heuristic methods. Of course, the computational cost in case of evolutionary methods is considerably higher, but it may result in much better designs, especially when highly nonlinear cases are

considered and the assumptions used in the HCA method are no longer correct. Another advantage of Evolutionary Algorithms is their good scalability on parallel machines, which can significantly reduce the optimization time and enable their use in industrial applications.

5 CONCLUSIONS

In this paper, a novel approach for crashworthiness Topology Optimization was presented. This technique uses Evolutionary Algorithms for optimization of crash structures parameterized implicitly with geometric level-set functions. For evaluation of the proposed method, optimization of the topology of a rectangular, clamped beam, impacted in the middle by a cylindrical pole, was considered. Performance of both standard Evolution Strategy and the Covariance Matrix Adaptation Evolution Strategy was compared.

The results show that the proposed approach can be successfully used as an optimization tool in the initial phase of development of crash structures. It can be considered as an attractive alternative for the state-of-the-art, often heuristic methods, such as the Hybrid Cellular Automata technique, since in evolutionary-based methods no additional assumptions have to be made and the optima can be precisely identified. Furthermore, unlike in the other approaches, handling of different objectives and constraints is straightforward, what broadens considerably the scope of possible applications. However, evolutionary optimization methods require much higher number of fitness function evaluations, which, especially in case of costly crash simulations, might pose a difficulty for industrial applications. Nevertheless, the excellent scalability properties of evolutionary methods make their use on parallel machines very efficient and thus, make them relevant also for the industry.

All in all, the future research in this field is promising, since the proposed methodology might be successfully used also in case of alternative objective functions even in case of highly nonlinear and noisy crash events, where the applicability of the standard crashworthiness optimization methods is limited. Further parallelization of simulations can help to overcome the problems associated with high computational costs and make the proposed approach an efficient method for crashworthiness optimization in early stages of the product development process.

REFERENCES

[1] LS-TaSC - Topology and Shape Computations for LS-DYNA, v2.0, 2011.

[2] D. V. Arnold and N. Hansen. A (1+1)-CMA-ES for Constrained Optimisation. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pages 297–304, New York, NY, USA, 2012.

[3] N. Aulig, S. Menzel, E. Nutwell, and D. Detwiler. Towards Multi-objective Topology Optimization of Structures subject to Crash and Static Load Cases. In *Engineering Optimization 2014*, pages 847–852. CRC Press, 2014.

[4] T. Bäck and H.-P. Schwefel. An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation*, 1(1):1–23, 1993.

[5] M. P. Bendsøe and O. Sigmund. *Topology Optimization*. Springer Berlin Heidelberg, Germany, 2004.

[6] M. Cavazzuti, A. Baldini, E. Bertocchi, D. Costi, E. Torricelli, and P. Moruzzi. High performance automotive chassis design: a topology optimization based approach. *Structural and Multidisciplinary Optimization*, 44(1):45–56, 2010.

[7] J. Christensen, C. Bastien, and M. V. Blundell. Effects of roof crush loading scenario upon body in white using topology optimisation. *International Journal of Crashworthiness*, 17(1):29–38, 2012.

[8] C. A. Coello Coello. and G. B. Lamont. *Applications of Multi-Objective Evolutionary Algorithms*. World Scientific, 2004.

[9] C. A. Coello Coello. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering*, 191(1112):1245–1287, 2002.

[10] N. P. v. Dijk, K. Maute, M. Langelaar, and F. v. Keulen. Level-set methods for structural topology optimization: a review. *Structural and Multidisciplinary Optimization*, 48(3):437–472, 2013.

[11] F. Duddeck. Multidisciplinary optimization of car bodies. *Structural and Multidisciplinary Optimization*, 35(4):375–389, 2007.

[12] F. Duddeck and K. Volz. A new topology optimization approach for crashworthiness of passenger vehicles based on physically defined equivalent static loads. In *Proceedings ICRASH conference, Milano*, Italy, 2012.

[13] F. Duddeck, S. Hunkeler, P. Lozano, E. Wehrle, and D. Zeng. Topology Optimization for Crashworthiness of Thin-walled Structures under Axial Impact Using Hybrid Cellular Automata. *Struct. Multidisc. Optim.*, 2015.

[14] H. Fredricson, T. Johansen, A. Klarbring, and J. Petersson. Topology optimization of frame structures with flexible joints. *Structural and Multidisciplinary Optimization*, 25(3):199–214, 2003.

[15] X. Guo. Doing topology optimization explicitly and geometrically: a new moving morphable components based framework. In *Frontiers in Applied Mechanics*, pages 31–32. Imperial College Press, London, UK, 2014.

[16] N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation*, pages 312–317, 1996.

[17] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[18] N. Hansen. The CMA Evolution Strategy: A Comparing Review. In J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea, editors, *Towards a New Evolutionary Computation*, number 192 in Studies in Fuzziness and Soft Computing, pages 75–102. Springer Berlin Heidelberg, Germany, 2006.

[19] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. In X. Yao, E. K. Burke, J. A. Lozano, J. Smith, J. J. Merelo-Guervós, J. A. Bullinaria, J. E. Rowe, P. Tiňo, A. Kabán, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature - PPSN VIII*, number 3242 in Lecture Notes in Computer Science, pages 282–291. Springer Berlin Heidelberg, Germany, 2004.

[20] N. Hansen, S. Müller, and P. Koumoutsakos. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[21] M. Hasenjäger, B. Sendhoff, T. Sonoda, and T. Arima. Three Dimensional Evolutionary Aerodynamic Design Optimization with CMA-ES. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, GECCO '05, pages 2173–2180, New York, NY, USA, 2005.

[22] S. Hunkeler. *Topology Optimisation in Crashworthiness Design via Hybrid Cellular Automata for Thin-walled Structures*. PhD thesis, Queen Mary University of London, UK, 2013.

[23] O. Kramer, G. Rudolph, and A. Barthelmes. *Constraint Handling for Covariance Matrix Adaption with Meta-Models*. Technische Universität Dortmund, Germany, 2009.

[24] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.

[25] C. K. Mozumder. *Topometry optimization of sheet metal structures for crashworthiness design using hybrid cellular automata*. PhD thesis, University of Notre Dame, USA, 2010.

[26] S. D. Müller, N. Hansen, and P. Koumoutsakos. Increasing the Serial and the Parallel Performance of the CMA-Evolution Strategy with Large Populations. In J. J. M. Guervós, P. Adamidis, H.-G. Beyer, H.-P. Schwefel, and J.-L. Fernández-Villacañas, editors, *Parallel Problem Solving from Nature — PPSN VII*, number 2439 in Lecture Notes in Computer Science, pages 422–431. Springer Berlin Heidelberg, Germany, 2002.

[27] M. Olhofer, T. Arima, T. Sonoda, and B. Sendhoff. Optimisation of a stator blade used in a transonic compressor cascade with evolution strategies. In I. C. Parmee, editor, *Adaptive Computing in Design and Manufacture (ACDM)*, pages 45–54. Springer Verlag, 2000.

[28] C. Ortmann and A. Schumacher. Graph and heuristic based topology optimization of crash loaded structures. *Structural and Multidisciplinary Optimization*, 47(6):839–854, 2013.

[29] G.-J. Park. Technical overview of the equivalent static loads method for non-linear static response structural optimization. *Structural and Multidisciplinary Optimization*, 43(3):319–337, 2010.

[30] N. M. Patel. *Crashworthiness Design Using Topology Optimization.* PhD thesis, University of Notre Dame, USA, 2007.

[31] C. B. W. Pedersen. Topology optimization design of crushed 2d-frames for desired energy absorption history. *Structural and Multidisciplinary Optimization*, 25(5-6):368–382, 2003.

[32] S. S. Rao *Engineering Optimization: Theory and Practice*. John Wiley & Sons, 2009.

[33] I. Rechenberg. Evolutionsstrategie: Optimierung Technischer Systeme Nach Prinzipien Der Biologischen Evolution. *PhD Thesis, Technical University of Berlin, Germany*, 15, 1971.

[34] H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Birkhäuser Basel, Basel, Germany, 1977.

[35] H.-P. Schwefel. *Collective phenomena in evolutionary systems*. Universität Dortmund. Abteilung Informatik, Germany, 1987.

[36] H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley VCH, New York, har/dskt edition, 1995.

[37] T. Sonoda, Y. Yamaguchi, T. Arima, M. Olhofer, B. Sendhoff, and H.-A. Schreiber. Advanced High Turning Compressor Airfoils for Low Reynolds Number Condition, Part 1: Design and Optimization. *Journal of Turbomachinery*, 126:350–359, 2004.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL SIMULATION OF HIGH-VELOCITY INTERACTION ELEMENTS WITH THIN-WALLED STRUCTURES

Alexander V. Gerasimov[1], **Sergey V. Pashkov**[1]

[1] Tomsk State University
36, Lenin st., Tomsk, Russia
ger@niipmm.tsu.ru

Keywords: Numerical Simulation, Structure, Projectile, Fragmentation, Probability.

Abstract. *The process of high-velocity collision of the elements of thin-walled structures with spherical and elongated projectiles, simulating fragments of cosmic bodies, artificial objects, ice particles and damaging elements is considered in three-dimensional Lagrangian formulation. The problem is solved taking into account probabilistic nature of crushing the material of the interacting bodies. Collision proceeds as normal and at an angle to a barrier surface. The approach proposed to solve fragmentation problems enables to reproduce the processes of barriers breaking with high-velocity elements in three-dimensional formulation which is the most complete from the physical point of view. The processes of interaction of the streams of high-velocity elements with thin-walled enclosures at different angles of approach to the latter, as well as the behavior of enclosure fragments and element debris behind the first shell of the construction, are of great practical interest to ensure reliable operation of spacecrafts, aircraft and missile technology. The streams of high-velocity elements may have different physical nature. They may be man-made and natural meteoric particles in space, ice particles in the atmosphere and damaging elements in all media. In this paper we consider processes of plates breaking by clusters of spherical and elongated projectiles and fragmentation pattern, as well as the formation of debris streams in behind-the barrier space.*

1 INTRODUCTION

To solve problem high – velocity interaction a group of elements with thin-walled structures it is necessary to have a reliable and sufficiently universal method to enable adequate reproduction of the processes occurring in solids under high- velocity collision. The natural heterogeneity of the structures of barrier and projectile materials affects the distribution of physical-mechanical characteristics (PMC) of the material, and is one of the most important factors determining the fracture behavior of real materials. One can account for it in the equations of deformable solid mechanics using a random distribution of the initial deviations of the strength properties from a nominal value (simulating the initial heterogeneity of the material). The relations of deformable solid mechanics, used in major recent works on dynamic fracture of structures and materials, ignore this factor. It can distort a real pattern of impact and explosive fractures of the bodies under consideration. The latter is particularly evident in the solution of axisymmetric problems, where the properties at all points on the circumferential coordinate of a calculated element are initially equal due to the use of standard equations of continuum mechanics in the numerical simulation. However in practice there are many problems where fragmentation is mainly a probabilistic process, for example the explosive fracture of axisymmetric shells, where the fragmentation pattern is unknown beforehand. The introduction of a random distribution of the initial deviations of the strength properties from a nominal value in the PMC of the body leads to the fact that, in these cases, the process of fracture becomes probabilistic in nature, and more consistent with the experimental data. To calculate elastoplastic flows we used a procedure implemented on tetrahedral cells and based on the combined application of Wilkins method intended to calculate the internal points of the body and Johnson method to calculate contact interactions. The initial heterogeneities of the structure were simulated by imposing a distribution of ultimate equivalent plastic strain in the cells of the calculated domain by means of a modified random number generator issuing a random variable obeying the selected distribution law. The probability densities of the random variables used were in the form of a normal Gaussian distribution with arithmetical mean being equal to the tabulated value and variable dispersion.

2 THE EQUATIONS DESCRIBING THE MOTION OF A COMPRESSIBLE ELASTIC-PLASTIC BODY TAKING INTO ACCOUNT PROBABILISTIC NATURE OF FRACTURE

The equations describing spatial adiabatic motion of a solid compressible medium are differential consequences of the fundamental laws of conservation mass, pulse and energy. In general they have the following forms [1-3]:

continuity equation

$$\frac{1}{\rho}\frac{d\rho}{dt}+\frac{\partial v_i}{\partial x_i}=0; \tag{1}$$

equation of motion

$$\rho\frac{dv_i}{dt}=\rho F_i-\frac{\partial P}{\partial x_i}+\frac{\partial S_{ij}}{\partial x_j}; \tag{2}$$

energy equation

$$\rho\frac{dE}{dt}=S_{ij}\varepsilon_{ij}+\frac{P}{\rho}\frac{d\rho}{dt}, \tag{3}$$

where x_i - the coordinates; t - time; ρ_0 - the initial density of the medium; ρ - the current density of the medium; v_i - components of the velocity vector; F_i - components of mass forces vector; S_{ij} - components of stress tensor deviator; E - specific internal energy; ε_{ij} - components of deviator of strain rate tensor; P - pressure.

To equations (1) - (3) we must add the equations taking into account relevant thermodynamic effects associated with adiabatic compression and strength of the medium. In general case, under the influence of the forces on the solid-deformable body, both volume (density) and the shape of the body are changed by different dependencies. Therefore, stress tensor is the sum of spherical tensor and the stress tensor deviator.

$$\sigma_{ij} = S_{ij} - P\delta_{ij}, \; i, j = 1,2,3,$$

$$\delta_{ij} = 1, \quad i = j, \qquad \text{where } \delta_{ij} \text{ - the Kronecker delta.}$$

$$\delta_{ij} = 0, \quad i \neq j,$$

To describe shear strength of the body the following relations are used:

$$2\mu(e_{ij} - \frac{1}{3}e_{kk}\delta_{ij}) = \frac{DS_{ij}}{Dt} + \lambda S_{ij}; \tag{4}$$

$$\frac{DS_{ij}}{Dt} = \frac{dS_{ij}}{dt} - S_{ik}\omega_{jk} - S_{jk}\omega_{ik}; \tag{5}$$

$$2\omega_{ij} = \frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i}; \tag{6}$$

$$2e_{ij} = \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}, \tag{7}$$

as well as the condition of plasticity

$$J_2 = \frac{1}{2}S_{ij}S_{ij} = \frac{1}{3}\sigma^2, \tag{8}$$

where e_{ij} - the components of the strain rate tensor; μ - shear modulus; σ - dynamic yield stress; D/Dt - Yauman derivative.

The equation of a solid state was chosen in the form of Mi-Gruneisen

$$P = \frac{K(1 - \Gamma_0\xi/2)}{(1 - c\xi)^2}\xi + \rho_0\Gamma_0 E, \tag{9}$$

where Γ_0 - Gruneisen coefficient; c, K - constant of the material; ρ_0 - the initial density of the medium; $\xi = 1 - \rho_0/\rho$.

As a shear failure criterion we used the criterion of limiting equivalent plastic strain [4] $\varepsilon^p = \varepsilon_*^p$. In this case, when ε^p reaches the limit value ε_*^p the calculated cell is considered to be destroyed. The system of equations (1) - (9) is written in a general form for the special motion of a deformable body.

The process of destructing the real materials is largely determined by the internal structure of the medium, the presence of heterogeneities, usually caused by a different orientation of the grains in the polycrystalline material or heterogeneities in the composition of composite materials, the difference in the micro-strength inside the grain and on the intergrain or interface boundaries. Therefore, to improve compliance of numerically simulated process with the experimental data it is necessary to generate disturbances in physical-mechanical characteristics of the medium being destroyed, i.e. to set a random distribution of the factors determining strength properties of the material. The introduction of information about polycrystalline structure of the material into calculation technique requires a large amount of experimental data and increased requirements for computer power that limits the ability of the implementation and apply this approach. In view of this, we used a simplified version of probabilistic modeling of fracture mechanism. Physical and mechanical characteristics of the medium responsible for strength are assumed to be randomly distributed over the material volume. The distribution probability density of these parameters is taken as various distribution laws, which are generally dependent on table (average) value of the parameter being distributed, dispersion of the parameter distribution being varied, and other characteristics of the medium. Such parameters as yield strength, tensile strength, maximum strains and other constants, that define the moment of destruction in various theories of strength and fracture criteria, are directly dependent on the number and size of defects and should be randomly distributed over the volume with dispersion depending on material homogeneity. Natural fragmentation of projectiles and barrier is calculated by introducing probabilistic mechanism for distribution of the initial defects of the material structure to describe tear and shear cracks. As a criterion of failure under intense shear strains we used the achievement of limiting values by equivalent plastic deformation. The initial heterogeneities were simulated so that the maximum equivalent plastic strain was distributed into membrane cells using a modified random number generator, which gave out a random variable obeying to the distribution law selected.

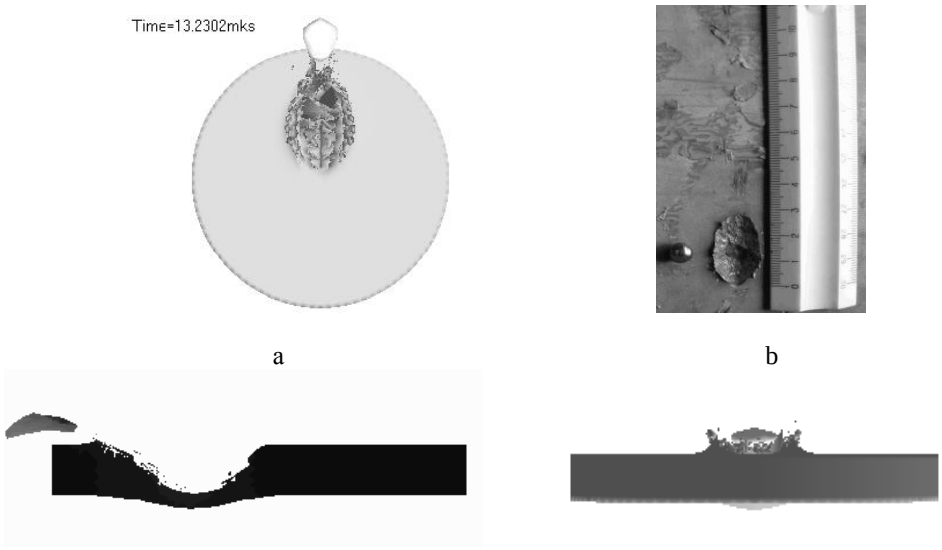
The system of basic equations is added with necessary initial and boundary conditions. At the initial moment of time all points of the projectile have axial velocity V_0 in view of its sign and the barrier state is assumed to be unperturbed. The boundary conditions are as follows, namely, the conditions $\sigma_n = \tau_n = 0$ are satisfied on borders free from stress. Conditions of ideal sliding of one material relative to another along the tangent and impermeability along the normal are set on contact sites between the bodies: $\sigma_{n1} = \sigma_{n2}, \nu_{n1} = \nu_{n2}, \tau_{n1} = \tau_{n2} = 0,$ where , σ_n, τ_n are the normal and tangent components of the stress vector; $n\,v$ is the normal component of the velocity vector at the point of contact; subscripts 1 and 2 refer to the bodies being in contact.

To calculate the elastic-plastic flows, use is made of a technique implemented on tetrahedral cells and based on joint application of the Wilkins method for calculating interior points of the body and the Johnson method for calculating the contact interactions [2, 5-6]. Three-dimensional domain was successively partitioned into tetrahedrals with subroutines of automatic meshing. The ideology and methodology of applying a probabilistic approach to solids fracture are completely described in [7].

3 TEST CALCULATIONS AND EXPERIMENTS

Figure 1 presents the results [8] of numerical simulation of ricocheting a steel ball-projectile with a diameter of 0.8 cm at interaction with a 0.95 cm-in-thick titanium barrier with a diameter of 8 cm for a time moment t=13.23 μs. The projectile initial velocity was 3600 m/s, angle of impact from the normal to the barrier was 75°.

In numerical study we obtained the following values of crater parameters, namely, a major crater axis was 30.5 mm, a minor axis - 15 mm and crater depth – 8.2 mm. The experimental data were characterized by the following values, namely, the major crater axis was - 28 mm, the minor axis was 16 mm and the crater depth was 7 mm.



a b

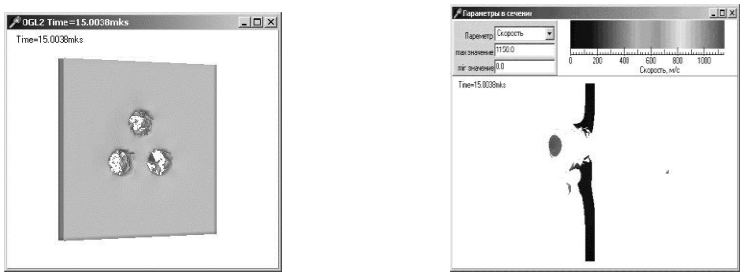
c d

Figure1: Numerical simulation of a steel ball ricocheting at interaction with titanium barrier (t = 13.23 μs ; V = 3600 m/s; $\alpha = 75$ °): a - top view; c, d - a side and end views; b - experiment (photos of the steel ball in the initial state and the trace in the titanium plate after their interaction).

4 COMPUTATIONAL RESULTS

Previously, the interaction of cylindrical elements with barriers at lower angles of impact and erosion fracture mechanism making it impossible to consider the interaction of the fragments with each other and with the barrier was considered in [9, 10]. In [11], the authors numerically investigated the interaction of ice projectiles with the barriers made of aluminum alloy and asbotextolite. The problem was solved in a two-dimensional formulation for the case of axial symmetry.

In this paper, the interaction of a group of three steel balls with an aluminum plate is considered in three-dimensional formulation. A ball radius was 0.28 cm. The thickness of each plate - 0.35 cm sized 5x5 cm. The impact velocity is 1500 m/s. The balls were arranged in a circle of radius 0.75 cm.



a b

Figure 2: Configuration of balls - barrier system at instant of time $t = 15\,\mu s$ for impact angle of 0^0 from a normal: a - general view; b - plane section.

Figure 2 shows the results of interaction of an aluminum plate with a cluster of steel balls. As is seen from Figure 2 after the collision the area of undestroyed material between the holes was quite large and the velocity of projectiles after the barrier breaking fell to 1130 m/s (Figure 2,b). The increase in the impact angle to 45^0 (Figure 3) caused the decrease in the area of undestroyed material between the punched holes and the formation of a significant fragmentation flow behind the barrier.

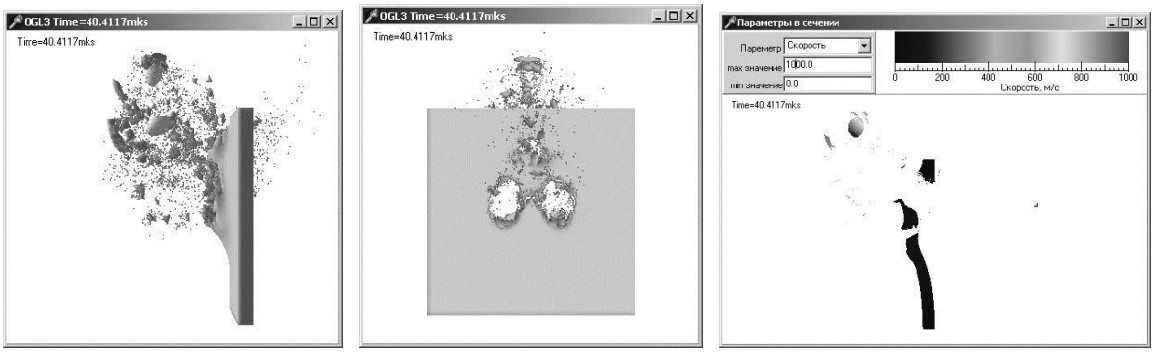


a b c

Figure 3: Configuration of balls - barrier system at instant of time $t = 38{,}4\ \mu s$ for impact angle of 45^0 from the normal: a, b - general views; c - plane section.

Two left Figures 3a and 3b show different views of the punched holes and the fragmentation field. Plane section of balls - barrier system and velocity distribution in the section is shown in Figure 3c. Such presentation of the results obtained is used for the subsequent figures. In the case under consideration there are two leading balls, followed by the third one. The velocity of the balls after breaking fell to 980 m/s.

Changing the direction of the impact by 180^0 (Figure 4), when there is one leading ball, one can observe cracks in the bridges between the holes and the velocity of the projectiles also falls to 980 m/s.

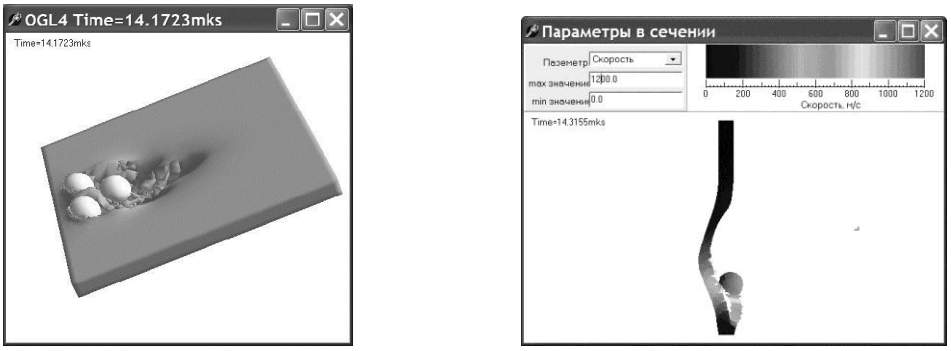


a b c

Figure 4: Configuration of balls - barrier system at instant of time $t = 40{,}4\ \mu s$ for impact angle of 45^0 from the normal: a, b - general views; c - plane section.

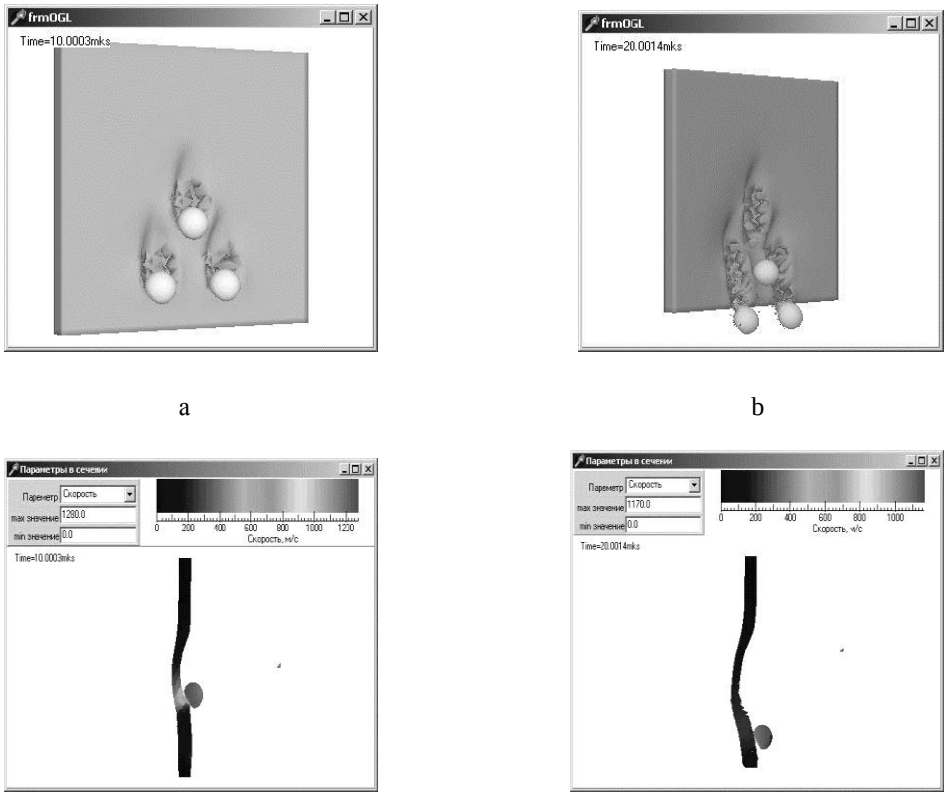
The next group of calculations was related to the study of projectile ricocheting at the following angles of impact: 70^0 (Figure 5), 75^0 (Figure 6) and 78^0 (Figure7). With the increase in the angle of impact the area of the deformed material and the values of aluminum plate deflections decreased. If at the angle of impact 70^0 (Figure 5b) we see the formation of cracks on the barrier face in the contact with the projectile, then for the angle of impact 75^0 (Figure

6d) the number and size of cracks decrease. At an angle of impact 78^0 (Figure 7d) one observes no cracks and reduced area of dents caused by balls-projectiles interactions with the plate.



a b

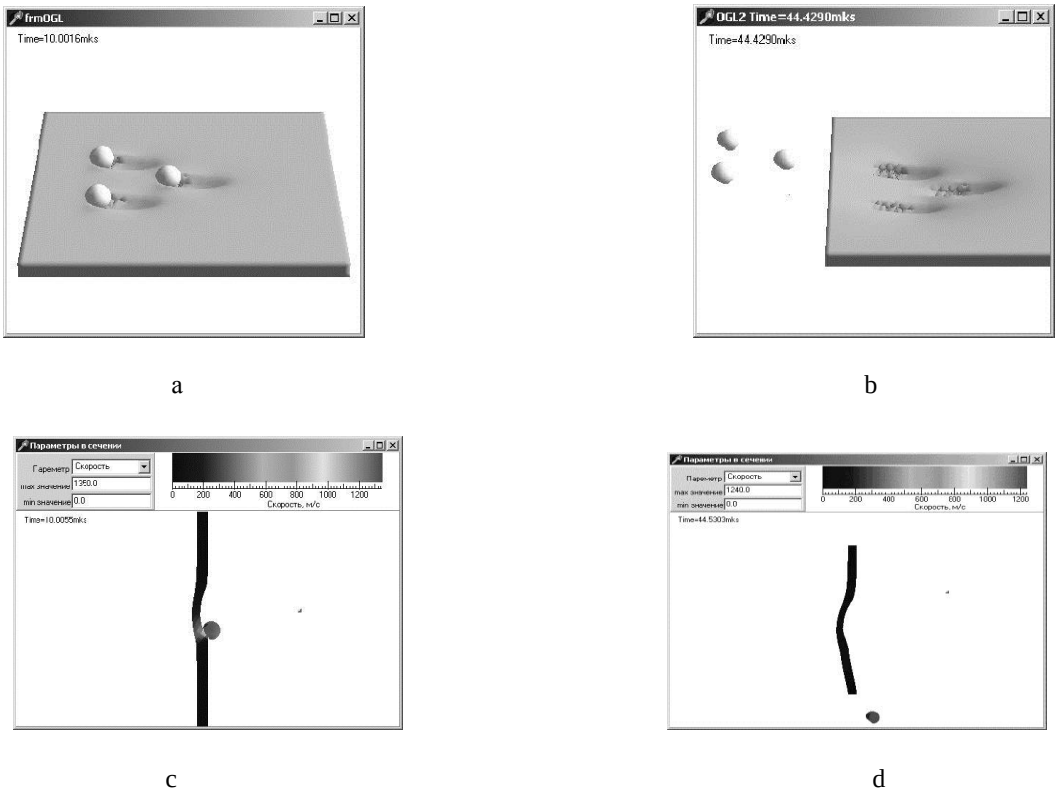
Figure: 5. Configuration of balls - barrier system at instant of time $t = 14,3$ μs for impact angle of 70^0 from the normal: a - general view; b - plane section.



a b

c d

Figure: 6. Configuration of balls - barrier system at instant of time $t = 10$ μs (a, c) and $t = 20$ μs (b, d) for the impact angle of 75^0 from the normal: a, b - general views; c, d - plane sections.

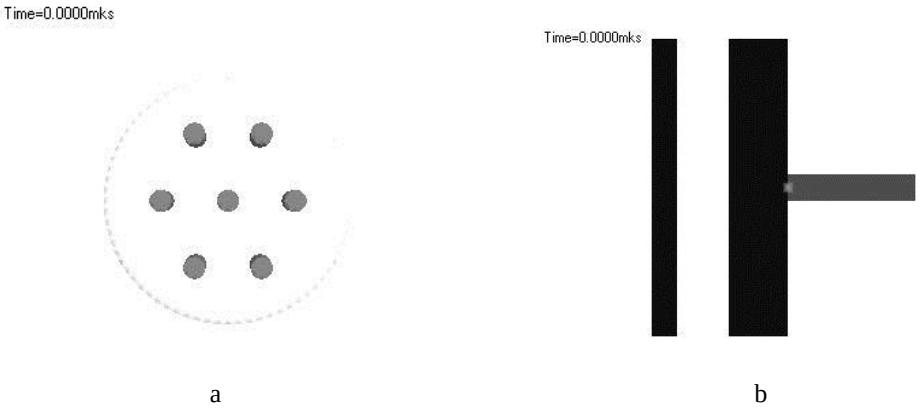


a b

c d

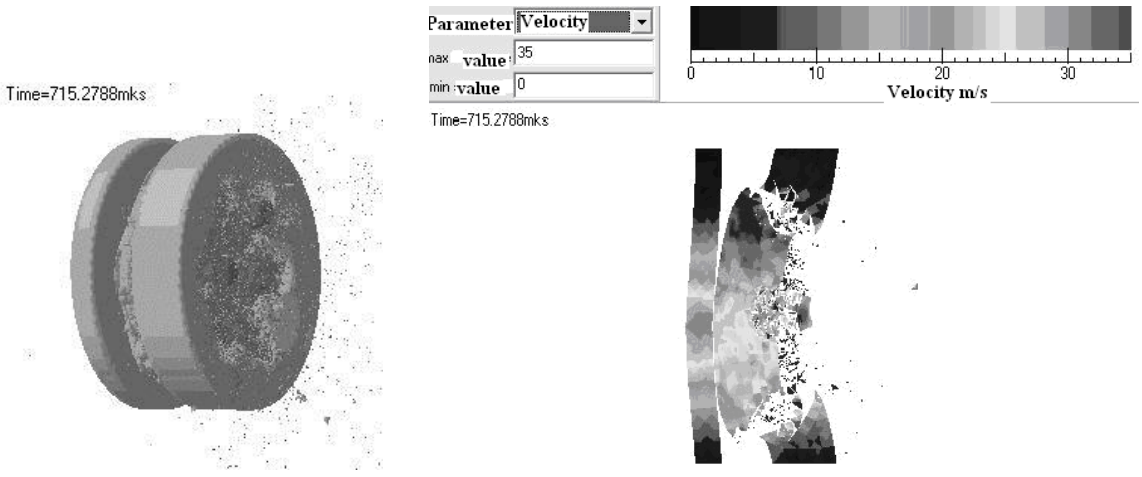
Figure 7: Configuration of balls - barrier system at instant of time $t = 10\,\mu s$ (a, c) and $t = 44{,}5\,\mu s$ (b, d) for the impact angle of 78^0 from the normal: a, b -general views; c, d – plane sections.

In this paper we considered the interaction of a group of seven rods of tungsten alloy with the system of steel plates. The radius of the rod was 1, 5 cm, length was 15 cm. Thickness of the first plate was 7 cm, the second was 3 cm, the distance between them was 6 cm, diameter was 35 cm. Collision speed was 1000 m / s. Rods arranged in a circle with a variable radius R. One projectile was located in the center, the other six circumferentially uniformly. The distance between the center of the first projectile and the remaining centers R during calculations varied.



a b

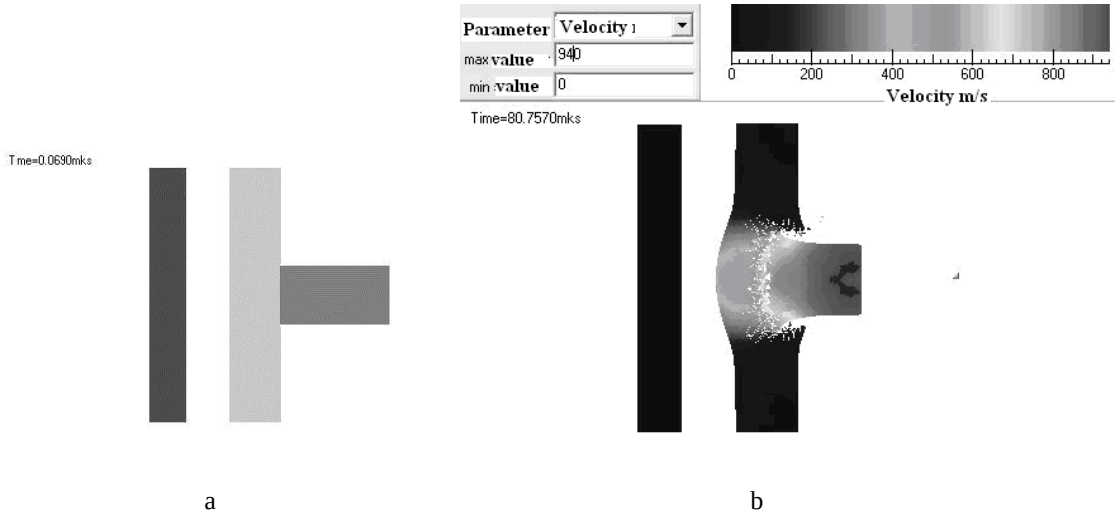
Figure 8: Initial configuration of system "obstacles - projectiles":
a - three-dimensional pattern; b-2D cross-section of a three-dimensional computational domain.



a b

Figure 9: Interaction of a projectiles with the barrier at R = 9 cm
a - three-dimensional pattern; b -2-D cross-section of a three-dimensional computational domain.

Figure 10 shows the results of calculating the penetration of a single projectile with a mass equal to mass of seven projectiles. It can be seen that the size of the mass which is knocked out (more light area in Figure10, b) much less of mass, which is knock out seven projectiles (Figures 8-9).



a b

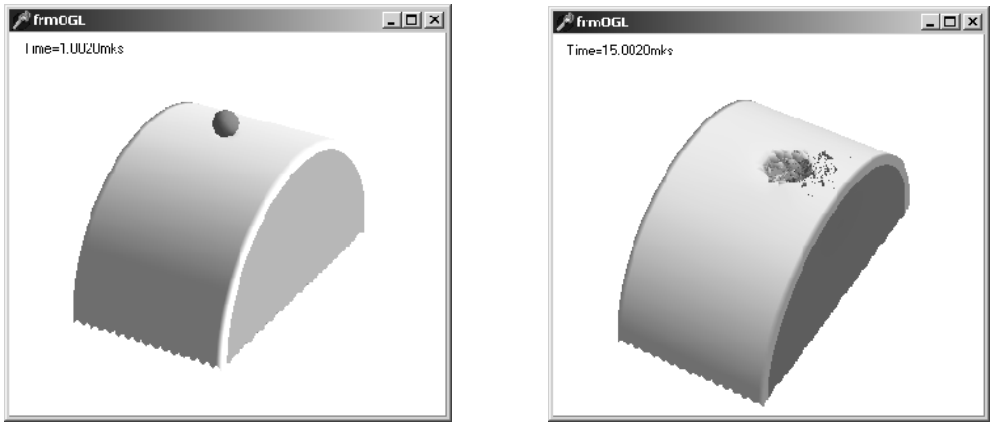
Figure 10: The interaction of a single projectile with obstacles: a - initial configuration of the system "barriers - projectiles "; b-2-D cross-section of a three-dimensional computational domain.

In this paper, a number of problems of penetration and destruction of shells with solid or liquid filler by projectiles interacting not only normally but also at an angle with the surface of the shell in three-dimensional formulation is considered. Problems are solved in a Lagrangian formulation for a wide range of velocities (up to 7 km / s) with the crushing of the material interacting bodies.

To reduce the calculation time only the upper shell part constituting half of the whole structure was considered. This technique can be used in the case when the double wave travel time to the boundary is greater than the time of the impact process. This condition is well satisfied for shells whose radius is much larger than the penetration depth of the projectile. On the right and left ends was used developed by the authors a technique that allows you to avoid

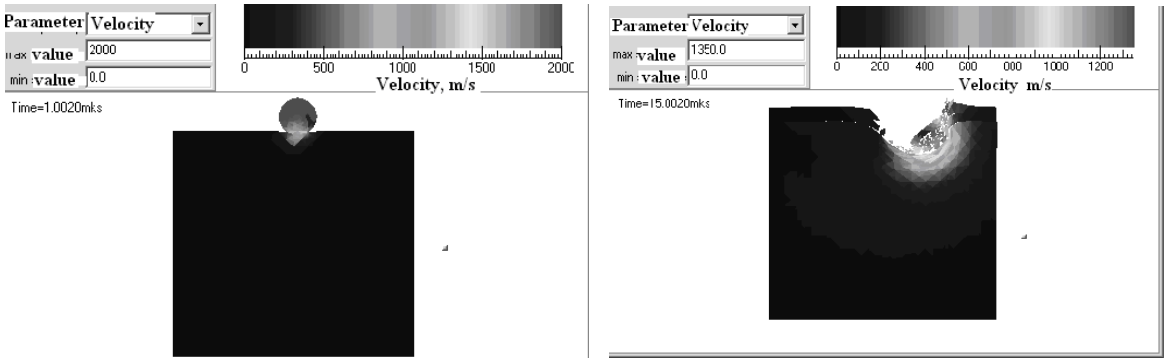
the effect of reflected waves and calculate the loading of long objects only on a limited part of them.

Collision shell and filler with a ball of the tungsten alloy is considered. The impact was normal and at an angle 45 to the generatrix of the shell. Velocity of projectile was 2 km / s. Dimensions of interacting objects are as follows: length of the steel shell was 8 cm; outer radius was 7.5 cm; inner radius was 7 cm; radius of the ball was 0.635 cm. The shell material - steel having the following physical-mechanical characteristics: density = 7.7 g / cm3, the shear modulus = 860 kbar, yield limit = 9.4 kbar; projectile - a tungsten alloy: density = 17.1 g / cm3; filler has the following characteristics: density = 1.75 g / cm3 = 34.7 kbar, yield limit = 1 kbar.



a b

Figure 11: The collision of the ball with a shell at an angle of 45^0: a-1μs, b-15μs.



a b

Figure 12: The velocity distribution in the structure upon impact at an angle of 45^0: a-1μs, b-15μs.

The process of interaction occurs with intense crushing of the ball and shell and further penetration of the fragments in the filler.

5 CONCLUSION

The probabilistic approach proposed and numerical technique developed on its basis enable us to simulate the processes of barrier breaking in a wide range of angles of impact. The decrease in the angle of collision causes projectile ricocheting and change in the nature of the aluminum plate deformation. It is possible to explore projectile and barrier fragmentations as well as the nature of the forming fields of fragmentation behind the barrier.

The paper compares the effectiveness of the impact of the group of projectiles and we made an assessment of their impact on the degree of damage obstacles. The calculation results showed greater risk of exposure to the group of rods for the protecting shell of the spacecraft relative to the impact of a single projectile with a mass equal mass of seven percussionists, and the same rate. The developed numerical method allows to simulate the interactions of shells spacecraft with high-speed long rods in a wide range of speeds and angles of impact, and also to investigate the processes of fragmentation barriers and a nature of fields of fragmentation .

This work was financially supported by RF Ministry of Education in the framework of state task № 2014/223 (project code 1567).

REFERENCES

[1] *Physics of explosion.* K. Stanjukovich ed. Nauka, 1975.

[2] M.L. Wilkins, *Computer simulation of dynamic phenomena*. Springer,1999.

[3] A.V. Gerasimov, S.V. Pashkov, Numerical simulation of the perforation of layered barriers. *Composites: Mechanics, Computations, and Applications. An International Journal.* **4,** 97-111, 2013.

[4] K.N. Kreyenhagen, M.H. Wagner, J. J. Piechocki, R. L. Bjork, Finding the ballistic limit in a collision with multilayer targets. *AIAA J.*, **8**, 42-47, 1970.

[5] G.R. Johnson, D.D. Colby, D.J. Vavrick, Tree-dimensional computer code for dynamic response of solids to intense impulsive loads. *Int. J. Numer. Methods Engng.*, **14**, 1865-1871, 1979.

[6] G.R. Johnson, Dynamic analysis of explosive-metal interaction in three dimensions. *Trans. ASME. J. of Appl. Mech*., **48,** 30-34, 1981.

[7] *Theoretical and experimental studies of high-velocity interaction of bodies,* A.V. Gerasimov ed. Tomsk University Press, 2007.

[8] A.V. Gerasimov, A.A. Konyaev, S.V. Pashkov, Modeling rebound someone compact shock when you hit the plate of finite thickness . *Proc. of All-Russian conference "Up-to-date Ballistics and Related Matters of Mechanics."* Tomsk, Tomsk University Press, 2009.

[9] I.E. Khorev, S.A. Zelepugin, A.A. Konyaev, V.N. Sidorov, V.E. Fortov, Breaking obstacles by a group of high-velocity bodies. *Doklady Physics,* **44,** 818-821, 1999.

[10] S.A. Zelepugin, A.A. Konyaev, V.N. Sidorov, I.E. Khorev, V.K. Yakushev, Experimental and theoretical investigation of encounter of a group of particles with protecting elements of spacecraft. *Space Research,* **46**, 529-539, 2008.

[11] V.P. Glazyrin, Yu.N. Orlov, Yu.N. Orlova, Analysis of barriers breaking with ice projectiles. *Russian Physics Journal,* **56**, 41-43, 2013

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL SIMULATION OF MULTILAYER COMPOSITES FAILURE UNDER DYNAMIC LOADING

Sergey A. Zelepugin[1,2], Aleksey S. Zelepugin[1,2], and Vladimir F. Tolkachev[1]

[1] Tomsk State University
36 Lenin Avenue, Tomsk, 634050, Russia
szel@yandex.ru

[2] Tomsk Scientific Centre, Siberian Branch of the Russian Academy of Sciences
10/3 Akademicheskii Avenue, Tomsk, 634055, Russia
szel@dsm.tsc.ru

Keywords: Failure, Multilayer Composites, Numerical Simulation.

Abstract. *The processes of multilayer composites failure under dynamic loading were investigated. In the experiments the Al_3Ti – Ti samples were tested on a ballistic stand and the features of the sample failure were investigated. The processes of high-velocity interaction of a projectile with a multilayer composite target were numerically investigated in axisymmetric geometry using the finite element method. To simulate numerically the failure of the material under dynamic loading, we applied the active-type kinetic model determining the growth of microdamages, which continuously changes the properties of the material and induce the relaxation of stresses. The strength characteristics of the medium (shear modulus and dynamic yield strength) depended on temperature and the current level of damage. The critical specific energy of shear deformations was used as a criterion of the erosion failure of the material that occurs in the region of intense interaction and deformation of contacting bodies. To simulate the brittle-like failure of the intermetallic material under high velocity impact, we modified the kinetic model of failure and included the possibility of failure above Hugoniot elastic limit (HEL) in the shock wave and sharp drop in strength characteristics if the failure begins. In the computations, the target consisting from intermetallic Al_3Ti - titanium alloy Ti-6-4 layers has been used. The results show that the depth of penetration depends on the thicknesses of intermetallic and titanium alloy layers. The composite target withstands the impact loading in the case of the ratio about 4/1 (Al_3Ti / Ti-6-4).*

1 INTRODUCTION

The field of material microstructure design targeted for a specific set of structural and functional properties is now a recognized field of focus in materials science and engineering [1]. A new class of structural materials called metal-intermetallic laminate (MIL) composites can have micro-, meso- and macrostructure [2-7]. The superior specific properties of this class of composites makes them attractive for high-performance aerospace applications, and the fabrication method for creating MIL composites allows new embedded technologies to be incorporated into the materials, enhancing their functionality and utility.

In this work experimental and numerical investigations have been done. In the experiments the Al_3Ti – Ti samples were tested on a ballistic stand and the features of the sample failure were investigated. The processes of high-velocity interaction of a projectile with a multilayer composite target were numerically investigated in axisymmetric geometry using the finite element method. The set of equations for describing unsteady adiabatic motion of an elasto-plastic medium, including nucleation and accumulation of microdamages and temperature effects, consists of the equations of continuity, motion, and energy [8-10].

To simulate numerically the failure of the material under high velocity impact, we applied the active-type kinetic model determining the growth of microdamages, which continuously changes the properties of the material and induce the relaxation of stresses. The strength characteristics of the medium (shear modulus and dynamic yield strength) depended on temperature and the current level of damage taking into account probabilistic approach to numerical simulation of fracture [11-13]. The critical specific energy of shear deformations was used as a criterion of the erosion failure of the material that occurs in the region of intense interaction and deformation of contacting bodies. To simulate the brittle-like failure of the intermetallic material under high velocity impact, we modified the kinetic model of failure and included the possibility of failure above Hugoniot elastic limit (HEL) in the shock wave and sharp drop in strength characteristics if the failure begins.

In the computations, the target consisting from 17 composite intermetallic Al_3Ti - titanium alloy Ti-6-4 layers has been used. Total thickness of the target was 19.89 mm. Thicknesses of intermetallic layer and a layer of titanium alloy were varied. The penetrator used was a tungsten heavy alloy rod with initial diameter of 6.15 mm and length of 23 mm. Initial impact velocity was of 900 m/s.

Results of computations demonstrate that destruction of the intermetallic layer is brittle as against to plastic failure of the metal layer. It was shown that the optimal composite target has higher ballistic resistance in comparison with a uniform target either Al_3Ti or Ti-6-4. Optimum construction of the composite target should include metal layer of sufficient thickness, which should provide the termination of propagation of brittle damage. The results show that the depth of penetration depends on the thicknesses of intermetallic and titanium alloy layers. The composite target withstands the impact loading in the case of the ratio about 4/1 (Al_3Ti / Ti-6-4).

2 EXPERIMENTAL RESULTS

In the experiments the Al_3Ti – Ti samples were tested on a ballistic stand and the features of the samples failure were investigated. Test conditions: SKS carbine, armour-piercing cartridge, cal. 7.62, the distance to the target was ~ 0.5 m, $\upsilon_0 = 760 \pm 3$ m/s. Target was a massive aluminium plate. The test sample was applied to the massive aluminium target. The measured density of the samples was 3.604 – 3.655 g/cm^3. Results of the microanalysis (AT CON,%) for selected points of intermetallic layer: Al = 73.029 - 73.868, Ti = 25.971 - 26.132, which corresponds to Al_3Ti (the density of Al_3Ti is 3.29 g/cm^3).

In Fig. 1, the Al_3Ti – Ti laminate composite after ballistic test is shown. Figs. 1a, 1b: sample after the test. There is delamination, but not full. Fig. 1c: sample after the cut. The head part of the core is stuck in the lower layers of the sample and protrudes at a distance of about 9.5 mm. Before cutting, this part of the core was separated, since the sectional plane of the sample and the core was different.

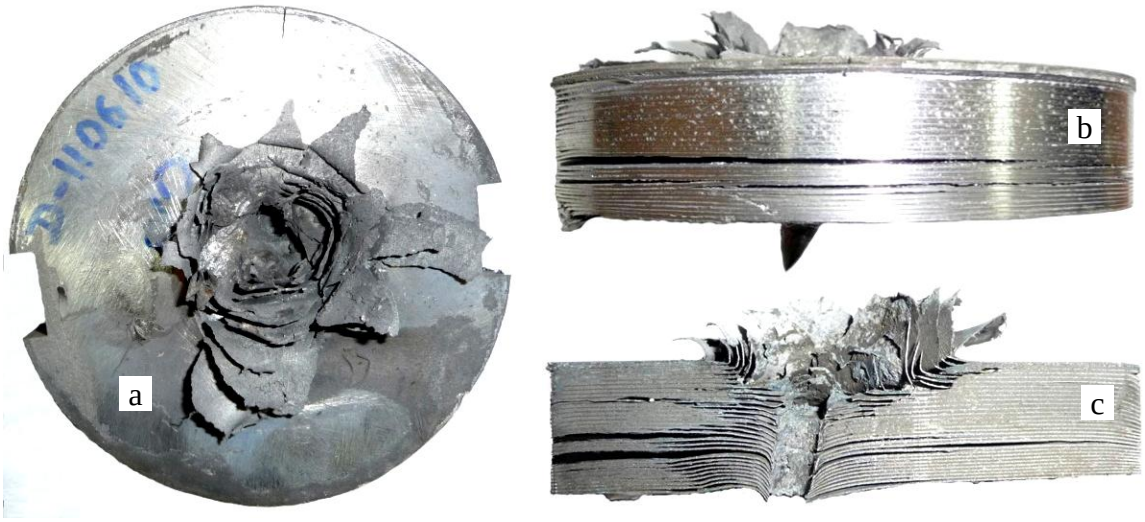


Figure 1: Failure of Al_3Ti – Ti sample in ballistic test.

3 NUMERICAL SIMULATION

3.1 Formulation of the problem

To simulate numerically the processes of high-rate shock loading, we use the model of an elastic-plastic damaged medium characterized by the presence of microcavities (pores, cracks). In the model the total volume of the medium W comprises the undamaged part of the medium of density ρ_c which occupies volume W_c, and microcavities of zero density which occupy volume W_f. The average density of the damaged medium is connected with the above-introduced parameters by the relationship $\rho = \rho_c(W_c/W)$. The degree of damage of the medium is characterized by the specific volume of microcavities $V_f = W_c/(W_*\rho)$.

A mathematical model employed in the numerical code for solving high velocity impact problems is based upon a set of differential equations of continuum mechanics that govern a material flow. The system of equations governing the nonstationary, adiabatic (for both elastic and plastic deformation) motion of a compressible medium with allowance for the evolution of microdamages comprises the continuity equation, the equation of motion, the energy equation [9, 10].

To simulate numerically the failure of the material at high velocity impact, we applied the active-type kinetic model determining the growth of microdamages, which continuously change the properties of the material and induce the relaxation of stresses:

$$\frac{dV_f}{dt} = \begin{cases} 0, \text{ if } |P_c| \le P^* \text{ or if } (P_c > P^* \text{ and } V_f = 0), \\ -\text{sign}(P_c) K_f (|P_c| - P^*)(V_2 + V_f), \text{ if } P_c < -P^* \text{ or if } (P_c > P^* \text{ and } V_f > 0) \end{cases} \quad (1)$$

Here $P^* = P_k V_1/(V_f + V_1)$, and V_1, V_2, P_k, and K_f are material constants determined experimentally. The form of condition (1) was chosen on based on the experimental data. It was assumed that there are fracture nuclei in the material, of identical initial sizes with the effective

specific volume V_1. Cracks or pores are formed and grow on these nuclei when the stretching pressure exceeds a certain critical value P^* which decreases as formed microdamages grow. The constants in (1) were adjusted by comparing the results of computations and experiments concerning the recording of a free surface velocity when a specimen is loaded by planar impulses of compression. One and the same set of constants is used when calculating both build-up and collapse of cracks and pores (depending on the sign of P_c).

The material model includes the equation of state of the Mie-Grüneisen type that represents pressure as a function of specific volume and specific internal energy, the deviatoric elastic constitutive relationships, the von Mises yield criterion taking into account temperature effects. The strength characteristics of the medium (shear modulus and dynamic yield strength) depend on temperature and the current level of damages.

To simulate the brittle-like failure of the intermetallic material under high velocity impact, we included in the model the possibility of failure above HEL in the shock wave and sharp drop in strength characteristics if the failure begins [14]:

$$\sigma = \begin{cases} \sigma_0 P_f K_T \left(1 + \frac{cP}{(1+\mu)^{1/3}}\right)\left(1 - \frac{V_f}{V_4}\right), \text{ if } V_f < V_f^k \\ \sigma_f K_T, \text{ if } V_f^k \le V_f < V_4 \\ 0, \text{ if } V_f \ge V_4 \end{cases}, \tag{2}$$

$$P_f = \begin{cases} 1, & \text{if } \sigma_{sh} < \sigma_{HEL} \\ P_f^k, & \text{if } \sigma_{sh} \ge \sigma_{HEL} \end{cases}, \tag{3}$$

where σ_{sh} is the stress in the shock wave ($\sigma_{sh} < 0$ for compression), P_f^k, V_f^k, σ_f, σ_{HEL} are the constants.

The critical specific energy of shear deformations is used as a criterion of the erosion failure of the material that occurs in the region of intense interaction and deformation of contacting bodies. The current value of the specific energy of shear deformations is defined from relationship

$$\rho \frac{dE_{sh}}{dt} = S_{ij}\varepsilon_{ij} \tag{4}$$

The critical value of the specific energy of shear deformations depends on the conditions of interactions and is a function of the initial impact velocity

$$E_{sh}^c = a_{sh} + b_{sh}\upsilon_0 \tag{5}$$

where a_{sh} and b_{sh} are constants.

When $E_{sh} > E_{sh}^c$ in the computational cell near the contact boundaries, the cell is assumed damaged and the parameters in neighbouring cells are corrected with regard for the principles of conservation laws.

3.2 Results and discussion

The problem of interaction of a projectile with a finite thickness target is under consideration. The problem is formulated in the Cartesian coordinate system with initial (at t = 0) and boundary conditions. The initial conditions are characterized by the absence of internal stresses, and motion of the projectile toward the target with a velocity υ_0 is a cause of interaction. There are no external loads on free surfaces of the interacting bodies throughout the entire process, while the conditions of sliding are realized on contact surfaces between the projectile and the target. The finite-element relations used to solve the formulated problem are given in [15].

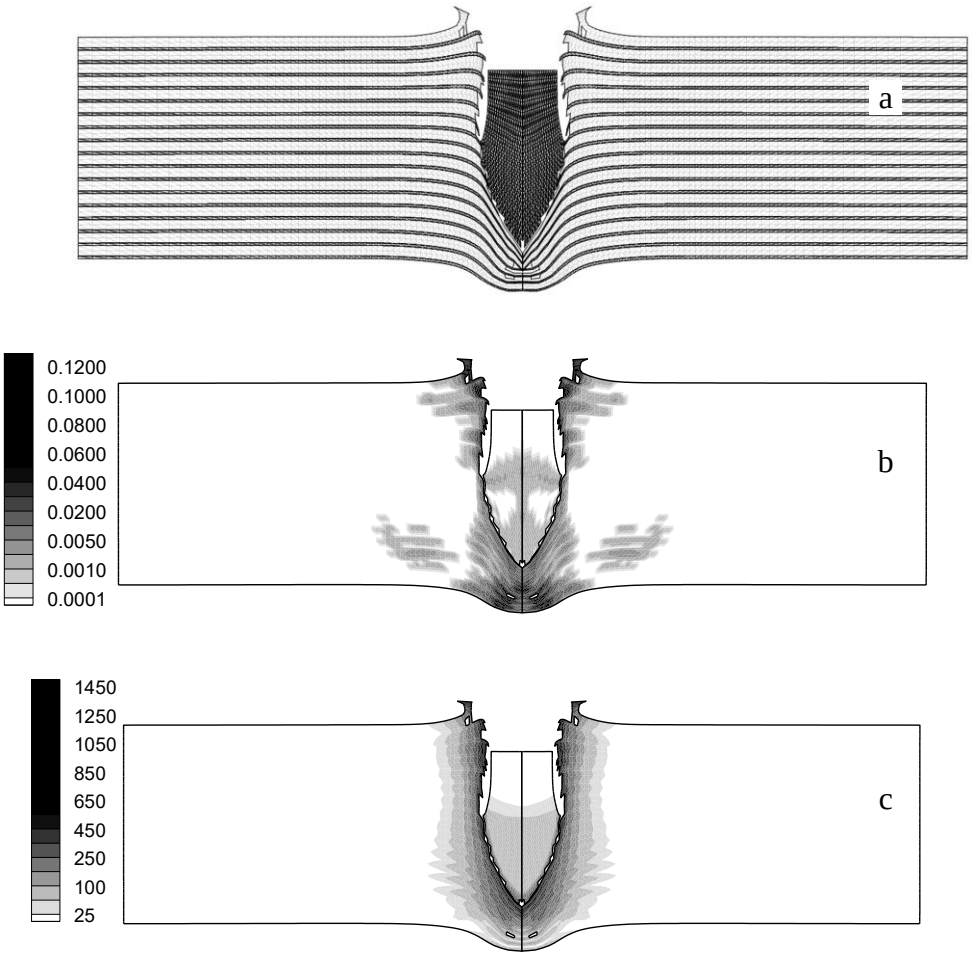


Figure 2: Computer images of the projectile/target assembly at 60 μs. Contours of a section of the projectile and composite target and finite elements in the target (a); specific volume of microdamage, cm^3/g (b); specific shear deformation energy, KJ/kg (c).

Fig. 2 shows computer images of a section of the projectile and composite target at the moment of time of 60 μs. Thickness of intermetallic Al_3Ti layer in this case was of 0.94 mm, thickness of a layer of Ti-6Al-4V titanium alloy was of 0.23 mm. The computations demonstrate the fact that in this case the MIL composite target withstands the impact loading.

The distribution of the damage and deformation patterns are illustrated in Figs. 2b and 2c, which show the contours of a section of the projectile and composite target and contours and fields of the specific volume of microdamage (Fig. 2b) and of the specific shear deformation

energy (Fig. 2c). The low level of microdamage in layers of titanium alloy shows the termination of propagation of brittle damage taking place in intermetallic layers.

	Al_3Ti [mm]	Ti-6-4 [mm]	Areal density [g/cm^2]	Depth of penetration [mm]		Average velocity [m/s]	
				40 μs	60 μs	40 μs	60 μs
1	0.94	0.23	7.02	17.00	18.49	150	30
2	1.17	-	6.54	21.12	25.41	350	150
3	-	1.17	8.97	18.73	20.90	200	50
4	0.47	0.70	7.99	22.95	28.71	430	250
5	0.23	0.94	8.49	24.15	-	470	-
6	0.70	0.47	7.52	21.86	26.44	410	220
7	1.04	0.13	6.81	22.85	-	440	-

Table 1: Results of simulations for different thicknesses of target layers.

The Table 1 represents results of simulations for different thicknesses of target layers. The results show that the depth of penetration depends on the thicknesses of intermetallic and titanium alloy layers. The MIL composite target withstands the impact loading in the case of 0.94 mm Al_3Ti / 0.23 mm Ti-6-4 (the ratio is about 4/1). In this case the intermetallic layer provides the failure of the projectile and the metal layer terminates the propagation of damage. In the other cases the perforation of the MIL composite target takes place. There is the same result for the uniform target made of either Al_3Ti (line 2 in the Table 1) or Ti-6-4 (3).

4 CONCLUSIONS

- Results obtained demonstrate that destruction of the intermetallic layer is brittle as against to plastic failure of the metal layer.
- It was shown in the computations that the optimal composite target has higher ballistic resistance in comparison with a uniform target either Al_3Ti or Ti-6-4. Optimum construction of the MIL composite should include metal layer of sufficient thickness, which should provide the termination of propagation of brittle damage.
- The results show that the depth of penetration depends on the thicknesses of intermetallic and titanium alloy layers. The composite target withstands the impact loading in the case of the ratio about 4/1 (Al_3Ti / Ti-6-4).

5 ACKNOWLEDGEMENTS

We thank A.M. Patselov for help in conducting the experiments. The reported study was partially supported by the Russian Foundation for Basic Research (project no. 16-08-00037).

REFERENCES

[1] O.N. Kryukova, A.G. Knyazeva, Simulating the formation of composite surface structure during electron beam treatment. *Advanced Materials Research*, **880**, 242-247, 2014.

[2] K.S. Vecchio, Synthetic multifunctional metallic-intermetallic laminate composites. *JOM*, **March**, 25-31, 2005.

[3] T. Li, F. Jiang, E.A. Olevsky, K.S. Vecchio, M.A. Meyers, Damage evolution in Ti6Al4V–Al_3Ti metal-intermetallic laminate composites. *Materials Science and Engineering A*, **443**, 1–15, 2007.

[4] R.D. Price, F. Jiang, R.M. Kulin, K.S. Vecchio, Effects of ductile phase volume fraction on the mechanical properties of Ti-Al_3Ti metal-intermetallic laminate (MIL). *Materials Science and Engineering A*, **528**, 3134-3146, 2011.

[5] Y. Cao, Sh. Zhu, Ch. Guo, K.S. Vecchio, F. Jiang, Numerical investigation of the ballistic performance of metal-intermetallic composites. *Appl. Compos. Mater.*, **22**, 437-456, 2015.

[6] A. Patselov, B. Greenberg, S. Gladkovskii, R. Lavrikov, E. Borodin, Layered metal-intermetallic composites in Ti-Al system: strength under static and dynamic load. *AASRI Procedia*, 3, 107-112, 2012.

[7] S.A. Zelepugin, V.I. Mali, A.S. Zelepugin, E.V. Ilina, Failure of metallic-intermetallic laminate composites under dynamic loading. *Shock Compression of Condensed Matter. USA: American Institute of Physics,* **1426**, 1101-1104, 2012.

[8] S.A. Zelepugin, A.S. Zelepugin, Numerical simulation of multilayer composites failure under dynamic loading. *Applied Mechanics and Materials,* **756**, 408-413, 2015.

[9] S.A. Zelepugin, A.S. Zelepugin, Failure of multilayer composites under dynamic loading. E. Oñate, X. Oliver, A. Huerta eds. *Proceedings of the jointly organized 11th World Congress on Computational Mechanics, the 5th European Conference on Computational Mechanics and the 6th European Conference on Computational Fluid Dynamics (WCCM XI – ECCM V – ECCM VI 2014)*, Barcelona, Spain, July 20-25, 2014, **4**, 3926-3933, 2014.

[10] S.A. Zelepugin, S.S. Shpakov, Failure of metallic-intermetallic multilayered composite under high-velocity impact. *Mekhanika kompozitsionnykh materialov i konstruktsii*, **15**(3), 369-382, 2009.

[11] A.V. Gerasimov, S.V. Pashkov, Numerical simulation of the perforation of layered barriers. *Composites: Mechanics, Computations, and Applications. An International Journal*, 4 (2), 97–111, 2013.

[12] A.V. Gerasimov, D.B. Dobritsa, S.V. Pashkov, Yu.F. Khristenko, Theoretical and experimental study of a method for the protection of spacecraft from high-speed particles. *Cosmic Research*, **54**(2), 118–126, 2016.

[13] A.V. Gerasimov, S.V. Pashkov, Numerical simulation of fracture in thin-walled structures by a group of high-velocity elements. *Russian Aeronautics,* **58**(2), 137-144, 2015.

[14] V.A. Gorelskii, S.A. Zelepugin, V.F. Tolkachev, Experimental and numerical study of ceramics destruction by high-velocity impact. *Chemical Physics Reports*, **18**(10-11), 2211-2217, 2000.

[15] V.A. Gorelski, S.A. Zelepugin, A.Yu. Smolin, Effect of discretization in calculating three-dimensional problems of high-velocity impact by the finite-element method. *Computational Mathematics and Mathematical Physics*, **37**(6), 722-730, 1997.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SIMULATION OF MECHANICAL PROPERTIES OF CERAMIC PARTS PRODUCED BY ADDITIVE TECHNOLOGIES IN WIDE RANGE OF LOADING RATES

Vladimir A. Skripnyak [1,2], Evgeniya G. Skripnyak [1], Vladimir V. Skripnyak [1], and

Irina K. Vaganova[1]

[1] National Research Tomsk State University
36 Lenin Avenue, 634050 Tomsk, Russia
e-mail: skrp@ftf.tsu.ru

[2] Institute of Strength Physics and Materials Sciences SB RAS
2/4, pr. Akademicheskii, Tomsk, 634021, Russia
e-mail: skrp2006@yandex.ru

Keywords: Instructions, ECCOMAS Congress, Computational Methods, Engineering Sciences, Proceedings.

Abstract. *The multiscale simulation approach was used to study the mechanical behavior of ceramic parts created by the selective laser sintering (SLS) technology. The goal of the research was studding the difference in the elastic moduli and dynamic strength of ceramic parts, manufactured by the high-temperature sintering and the selective laser sintering. The mechanical properties of ceramic parts were determined by statistic averaging of the results of multi scale computer simulations of representative volumes of materials under loadings. It was shown that the probability of the fracture of such parts under dynamic loading depends on the time history evolution of the distribution of local values of specific internal energy and the damage parameter on the mesoscale level. It was found that the stress threshold of the dynamic fracture beginning of ceramic bodies depends on the distribution of the pore sizes and not just on the average porosity.*

1 INTRODUCTION

Additive manufacturing (AM) of ceramics is intensively expanding in recent years, an AM technologies are moving to the class of economically beneficial industrial technologies [1]. These AM technologies can all be applied to creation of complex shape ceramic parts from nano-sized ceramic powders. In accordance with the ISO/ASTM AM technology is usually classified in two groups. The group of "single-step" technology is considered as a group in which the geometry and material are created simultaneously.

Great interest is manifested by the improvement of "single-step" AM technologies of manufacturing of ceramic parts [1-8]. In recent years, the selective laser sintering (SLS) manufacturing processes have been used for the production of ceramic parts [2-8]. SLS is an additive

manufacturing process for creating 3D parts with complex geometries because the use of powder bed SPS makes it possible to densify the UHTC composites at a lower temperature and in a shorter time compared with conventional techniques of high temperature pressing (HIP). The list of UHTC materials with melting temperatures above ~3000 K is limited to perhaps 16 phases or compounds (Table I).

Phase / /crystalline lattice	Mass density, 10^3 kg/m^3	Young's modulus, GPa	Poisson's ratio	Linear expansion coefficient 10^{-6} K^{-1}) in a range (293 – 1300 K)	Melting temperature, T_m, K	Debye temperature, Θ_D, K
HfB_2 / Hexagonal	11.19	530	0.12	6.3	3653	690
HfC / FCC	12.76	352 - 462	0.18	6.6	4173	680
HfN / FCC	13.9	380	0.25	6.5	3658	684
ZrB_2 / Hexagonal	6.10	450-500	0.128	5.9 -6.5	3518	942
ZrC / FCC	6.56	348	0.18	6.7	3673	744
ZrN / FCC	7.29	510	0.25	6.7	3223	684
TiB_2 / Hexagonal	4.52	575	0.14	4.6 -5.2	3498	820-1140
TiC / Cubic	4.94	460-497	0.19-0.25	7.95-8.58	3373	916
TiN / FCC	5.39	440-615	0.12	9.35	3223	809
TaB_2 / Hexagonal	12.54	551		8.2	3313	1309
TaC/ Cubic	14.50	472	0.1719	6.3	4073	776
WC / Hexagonal	15.77	668-714	0.24	3.85-3.9	3143	779
TaN / Cubic	14.30	466	0.246	8.31	2973	
SiC / Polymorphic	3.21	454	0.16	4.7	3093	1200
B_4C / Rhombohedral	2.52	432-463	0.15-0.18	4.5	2718	1016
ZrO_2 / Polymorphic	6.05	360	0.303 - 0.312	8.0-10.6	2973	590
Si_3N_4 / Polymorphic	3.29	166 -310	0.23 - 0.276	2.64-3.3	2661 -2769	1178

Table 1. Mechanical and thermodynamic properties of UHTC [1-18]

The use of SLS allows creating new composite materials simultaneously with the manufacture of structural elements. Leu and co-authors (2008), Sun, and Gupta (2008), Pattnaik and co-authors (2012) showed the possibility of freeform parts fabrication of ultra-high tempera-

ture ceramics (UHTC) based on the zirconium diboride by laser sintering [5, 6]. Liu and co-authors (2015) showed the possibility of laser sintering of ZrO_2 parts [7]. Parts that are made of UHTC have high oxidation resistance and are used for nozzles, the leading edges of space vehicles re-entering the earth's atmosphere, cutting tools, bearings, igniters, and others.

The main problem in creating high strength UHTC is connected with difficulties of consolidation due to high melting point and low self-diffusion coefficient of borides, carbides, and nitrides phases.

Computer designing of ceramic parts produced by SLS method, we need to resolve two key tasks. The first task is ensuring the required dimensions of parts. The second task is to provide the required strength of ceramics for normative loading conditions. Various methods and models are applied for the solution of these problems at different scale levels.

On the macro-level, the architecture of sintered parts is designed by CAD/CAM technology and the laser scanning patterns. Complex shapes of ceramic parts with hollow structures can be fabricated by SLS. To predict the strength of ceramic parts, it is required to know the effective values of mechanical properties of material include the strength under tension and compression, crack resistance, and other characteristics.

Mechanical properties of sintered ceramics depend on the parameters of laser scanning and the properties of ceramic's powder [7,9]. For theoretical prediction of mechanical properties of sintered ceramics it is need models that must take into account certain determining factors, which include: residual stresses, grain size distributions, interface boundaries, cracks and pores distributions, formed during sintering.

In present study, this approach was used to develop a two-scale model. The mechanical characteristics of the ceramic material (elastic moduli, the Hugoniot elastic limit, the spall strength, etc.) are predicted using the results of computer simulation of response of model representative volume element (RVE) over a wide range of strain rates. Statistical–numerical RVE determination techniques discussed in [10,11].

Using a modified smoothed particle hydrodynamics (SPH) method made it possible to separately take into account the kinetics of damage accumulation at the micro - and mesoscale levels [12]. Damage and fracture of condensed phases at the micro level are described implicitly using a damage parameter. At the mesoscale level the formation of mesoscopic cracks was simulated.

This approach will enable to adequately consider the impact of a wide range of structural defects on mechanical properties of materials obtained by SLS method [13,14].

2 MODEL OF MECHANICAL PROPERTIES OF CERAMICS PRODUCED BY ADDITIVE TECHNOLOGIES

Mechanical properties of sintered ceramics under dynamic loads can be defined using simulation of mechanical response structured RVE of ceramics. Model representative volumes are proposed to create using the evaluations of residual stress, the parameters of the interface boundaries, the parameters of the distributions of cracks and pores. The dimensions of the representative volume should exceed a multiple of the largest size structures of pores, cracks, volumes of melted particles. Model RVE of sintered ceramics with porosity ~ 0.05 shown in Figure 1. A continuum approach was used for describing response of structured RVE at the mesoscale level. Data on the mechanical properties of ceramic phases, the sintering temperature are required for prediction of mechanical properties of sintered materials. Properties of material particles in this case are defined for volume of matter ~ 1 μm^3. In ceramics, fabricated by SLS of submicron powders, the size of pores and cracks can range from tens of nanometers to tens of micrometers [5,6,15,and 16].

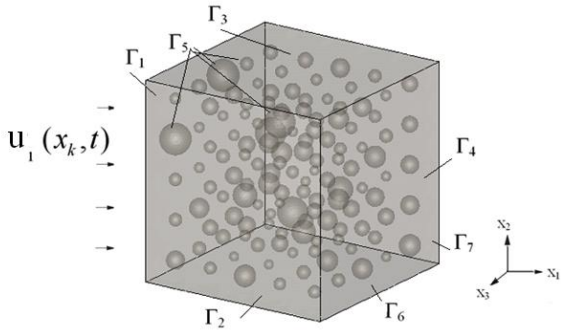


Figure 1. Model RVE and scheme of boundary conditions.

Pores and cracks with dimensions exceeding ~1 μm are set explicitly in the RVE. Their distribution in the RVE can be set randomly or taking into account the special distribution (for gradient or layered porous structures).The closed pores and penny shaped cracks of submicron size take into account implicitly by means of correction parameters of mechanical properties of material particles. In this case volumetric mass density and elastic moduli are defined by relations [17]:

$$\rho = \rho_{th} \ \ (1-\alpha_0),\ M \approx M_{th}(1-\alpha_0), \tag{1}$$

where ρ is mass density, ρ_{th} is theoretical mass density of condensed phase, α_0 is the specific volume of micro-pores, M is the effective elastic modulus (the shear modulus, or the Young's modulus, or the bulk modulus) and M_{th} is the elastic modulus of the condensed phase.

Dynamics of structured media in RVE is described in Lagrangian coordinate system by mass conservation equation, momentum conservation equation energy and conservation equation [18,19]:

$$\nabla_i \sigma_{ij} = \rho \dot{u}_i \ , \ \dot{\rho} = \rho \nabla_i u_i \ , \ \rho \dot{E}_{\text{int}} = \sigma_{ij} \dot{\varepsilon}_{ij} \ , \tag{2}$$

where σ_{ij} is components of the Cauchy stress tensor, ρ is mass density, u_i are components of particle velocity vector, $\dot{\varepsilon}_{ij}$ are components of the strain rate tensor, E_{int} is the local specific internal energy per unit mas, the substantial derivative are market by dot notation.

Boundary conditions were applied to the cube-shaped RVE shown in Figure1.

$$\begin{gathered} u_1(x_k,t) = F(x_k,t) \ \ x_k \in \Gamma_1, \\ u_2(x_k,t)=0, \ \ x_k \in \Gamma_2, \ x_k \in \Gamma_3, \ u_3(x_k,t)=0, \ x_k \in \Gamma_6, x_k \in \Gamma_7, \\ p = u_n [\rho C]_{\Gamma_4} \ for\ compression,\ or\ u_1(x_k) = -F(x_k),\ for\ tension, x_k \in \Gamma_4, \end{gathered} \tag{3}$$

where C is decreasing from longitudinal sound velocity in the elastic precursor to the bulk sound velocity C_B, $F(x_k,t)$ is a function for assigning a pulse shape and duration.

The initial conditions were used in the form:

$$\rho(x_k) = \rho^{(0)}(x_k), \ \ \varepsilon_{ij}(x_k) = \varepsilon_{ij}^{(0)}(x_k), \ E(x_k) = E^{(0)}(x_k), \ D(x_k) = D^{(0)}(x_k) \ , \tag{4}$$

where D is the damage parameter.

Residual stress is included in the model using eigenstrain $\varepsilon_{ij}^{(0)}(x_k)$ [20]. Eigenstrain is an incompatible strains set up during sintering of a material. Residual stresses caused by thermal

expansion mismatch between the randomly distributed microparticles of ceramic phases were introduces in the initial state of material particles.

The initial value of the damage parameter $D^{(0}$ was determined by the method of homogenization to damages of a material particle containing a distribution of flaws [21].

Components of local strain rate tensor are defined by the relation:

$$\dot{\varepsilon}_{ij} = \frac{1}{2}(\nabla_i u_j + \nabla_j u_i) \ , \tag{5}$$

where u_i are components of velocity of material particles, $i=1,2,3$, ∇_i is the Hamilton operator.

The components of strain rates tensor is expressed by sum of the elastic and the inelastic parts:

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}^e_{ij} + \dot{\varepsilon}^n_{ij} \ . \tag{6}$$

The constitutive equation of material particles is written in the form:

$$\sigma_{ij} = \sigma_{ij}{}^{(c)} \varphi(D), \ \sigma^{(c)}{}_{ij} = -p^{(c)} \delta_{ij} + S_{ij}{}^{(c)}, \tag{7}$$

where p is the pressure, S_{ij} is the stress deviator tensor, the superscript c indicates the condensed phase of the damaged material, $\varphi(D) \approx 1 - D$, D is the local damage parameter, *and* δ_{ij} is the Kronecker delta.

The local damage parameter D is introduced in the form:

$$D = \int_0^{t_f} \frac{\dot{\varepsilon}^n_{eq}}{\varepsilon^n_f} dt \ , \qquad . \tag{8}$$

where $\dot{\varepsilon}^n_{eq} = [(2/3)\dot{\varepsilon}^n_{ij}\dot{\varepsilon}^n_{ij}]^{1/2}$ denote an intensity of inelastic strain rate tensor, $\varepsilon_f{}^n$ is the threshold of inelastic deformations of material particles (the volumes of the phases at the micro level) at fracture.

The growth of the damage parameter of the material particles implicitly reflects the degradation of the strength properties of phases due to nucleation and growth of microcracks. The criterion of local failure corresponds to D=1.

Threshold of effective shear strain ε_f is approximated by relation:

$$\varepsilon^n_f = D_1 (P^* + T^*)^{D_2} \ , \tag{9}$$

where $T^* = \sigma_{sp}/P_{HEL}$, $P^* = p/P_{HEL}$, P_{HEL} is pressure correspond to the Hugoniot elastic limit, $\sigma_{sp} = (3\rho C_l \dot{\varepsilon}_{eq} K^2_{1C})^{1/3}$, C_l is the longitudinal sound velocity, D_1, D_2 are material constants [19, 22].

Using the SPH-method, the formation of mesocrack described as the creation of spatial clusters of particles with broken bonds with neighboring particles [23].

In this case, the distance between neighboring particles should exceed the critical opening displacements at the crack tip δ^*.

The critical opening smooth particle displacement was computed by *Eqs* (10) [24]:

$$\delta = r_{kl}\big|_t - r_{lk}\big|_{t=0} \approx \delta^*, \ \ \delta^* = 2G_f / \sigma_f \ , \ 2G_f \approx 2K_{1C}{}^2 / E \ , \tag{10}$$

where r_{kl} is the current (at time t) distance between k-th and l-th smoothed particles and δ^* is the critical distance of the crack formation (corresponds to loss of interaction between particles), G_f is the fracture energy, σ_f is the thresholds stress of local brittle fracture.

Macroscale fracture of RVE is characterized by complex temporal and spatial organizations of mesocracks. The thresholds stress of local brittle fracture for ceramics under compression σ_f in wide range of strain rates can be described by *Eqs*. (10):

$$\sigma_f / \sigma_0 = 1 + (\dot{\varepsilon}_{eq} / \dot{\varepsilon}_0)^{2/3},\ \sigma_0 = 4.25 K_{1C} \eta^{1/4} / [\alpha]^{1/2},\ \dot{\varepsilon}_0 = 4.25 K_{1C} \eta^{3/4} / [\alpha \rho E]^{1/2}\ , \quad (11)$$

where α is a specific volume of voids, ρ is the mass density, K_{1C} is the fracture toughness, E is the Young's modulus, and η is a square density of unstable microcracks [25,26].

The thresholds stress of local brittle fracture for ceramic under tension at high strain rates described by *Eq*. (11):

$$\sigma_{f\,sp} = \sigma_t + (3\rho C_l\, \dot{\varepsilon}_{eq}\, K_{1C}^2)^{1/3}, \quad (12)$$

where σ_t is the quasi-static tensile strength.

Pressure calculated by *Eqs* (13):

$$p = p_x + \Gamma \rho_0 E_{\text{int}}^T,$$

$$p_x = \frac{3}{2} \cdot B_0 \cdot (\xi^{-7/3} - \xi^{-5/3}) \cdot (1 - \frac{3}{4}(4 - B_1) \cdot (\xi^{-2/3} - 1)) \text{ for compression } \xi^{-1} - 1 > 0\,, \quad (13)$$

$$p_x = B_0\, \xi^{-1} - 1, \text{ for tension } \xi^{-1} - 1 \le 0\,,$$

where B_0, B_1 are material constants, $\xi = \rho_0 / \rho$, Γ is a Gruneisen coefficient, E_{int}^T is thermal part of a local specific internal energy per unit mass [27].

Shear micro-cracks nucleation in a condensed ceramic phase was accounted

$$\frac{D[S_{ij}]^{(m)}}{Dt} = 2\mu(\dot{\varepsilon}_{ij}^e - \frac{1}{3}\dot{\varepsilon}_{kk}^e\ \ \delta_{ij}),\ \frac{D[S_{ij}]^{(m)}}{Dt} = [\dot{S}_{ij}]^{(m)}\ \ - [S_{ik}]^{(m)} \dot{\omega}_{jk} - [S_{jk}]^{(m)} \dot{\omega}_{ik}\,, \quad (14)$$

where μ is the shear modulus.

$$\dot{\varepsilon}_{ij}^e = \dot{\varepsilon}_{ij} - \dot{\varepsilon}_{ij}^n\ ,\ \dot{\varepsilon}_{ij}^n = \dot{\lambda}\ \frac{\partial g}{\partial \sigma_{ij}},\ g = \left[(3/2) S_{ij} S_{ij}\right]^{1/2} - \sigma_f (1 - D)\,, \quad (15)$$

where $\dot{\lambda}$ is a model parameter.

The following parameters of Eq (11) for ZrB_2 were used in the calculations: ZrB_2: E =495 GPa, ρ =6.11 103 kg/m^3, K_{1C} = 3.5 MPa m$^{1/2}$, σ_{HEL} =7.11 GPa, p_{HEL}=3.07 GPa, D_1= 0.1, D_2 =1.0, Cl = 9.233 km/s, σ_{fsp}=0.5 GPa, σ_{sd}=0.5 GPa.

The values of the eigenstrain $\varepsilon_{ij}^{(0)}(x_k)$were assessed according to the estimated residual pressures that occur after cooling of sintering of ceramics. SLS of UHTC uses a sintering temperature from 1900 K to 2200 K [1-8].

The residual pressure was calculated by *Eq*.(16) [30]:

$$p_{res} = \frac{2(\alpha_m - \alpha_{inc})(T - T_r) E_m E_{inc}}{E_{inc}(1 + \nu_m) + 2E_m(1 - 2\nu_{inc})}\,, \quad (16)$$

where p_{res} is the residual pressure, α_m , α_{inc} are the linear thermal expansion coefficients of matrix and inclusion, respectively , T, T_r are the sintering temperature, and the room tempera-

ture, E_m, E_{inc}, ν_m, ν_{inc} is Young's modulus and Poisson's ratio; subscripts *m* and *inc* correspond to matrix and inclusions, respectively.

The tensile stress decreases with the distance from the interface between grains of the matrix and the inclusion [30]:

$$p_{res} = \frac{(\alpha_m - \alpha_{inc})(T - T_r) E_m E_{inc}}{E_{inc}(1-2\nu_m) + 2E_m(1-2\nu_{inc})} \left(\frac{d_{inc}}{r + d_{inc}}\right)^3 , \tag{17}$$

where d_{inc} is the size of the inclusion, and r is the radial distance from the inclusion surface, respectively.

The increment of the fracture toughness K_{IC} due to residual pressure can be estimated by the *Eq.* (18) [30]:

$$\Delta K_{IC} = \frac{-2\, C_{inc}\, p_{res}}{1 - C_{inc}} \sqrt{\frac{2(L - d_{inc})}{\pi}} , \tag{18}$$

where C_{inc} is the volume fractions of inclusions, L is the average distance between inclusions, d_{inc} is the average size of inclusions, p_{res} is the residual pressure in the matrix.

The fracture toughness K_{1C} calculated using the residual pressure was used to determine the parameters σ_f ,σ_0 , $\dot{\varepsilon}_0$ in the *Eq.* (11).

The mechanical properties of ceramics were determined by statistic averaging of the results of multi scale computer simulations of representative volumes of materials by method [19].

3 RESULT AND DISCUSSION

Figure 2 shows the thresholds stress of local brittle fracture for ceramics under compression versus normalized strain rates.

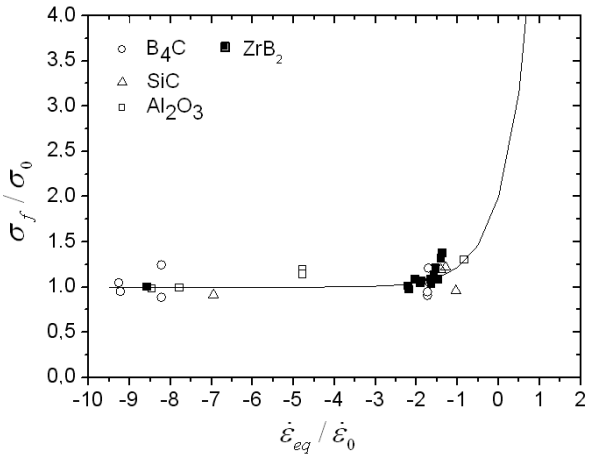


Figure 2. The thresholds stress of local brittle fracture for ceramics under compression versus normalized strain rates.

The characteristic stress and strain rate parameters for ceramics fabricated by hot isostatic pressing shown in Table 2 [25, 28, 33]. To predict the thresholds stress of quasi-brittle fracture under dynamic loading of ceramic bodies fabricated by additive technologies, including SLS, it is necessary to consider the impact the residual pressure, density and the size distribution of pores and flaws on the characteristic stress σ_0 and strain rate $\dot{\varepsilon}_0$ parameters.

Material	σ_0 (GPa)	$\dot{\varepsilon}_0\,(s^{-1})$
B_4C	3.4	10^4
SiC	5.4	$5\ 10^3$
ZrB_2	2.1	$0.95\ 10^5$
Al_2O_3	3.05	$9.0\ 10^3$

Table 2. Characteristic stress and strain rate parameters for ceramics showed in Figure. 2.

Characteristic stress and strain rate parameters depend on the fracture toughness K_{IC}. The increase in the fracture toughness K_{IC} may be due to residual pressure (see *Eq.*18).

Figure 3 shows the calculated by the *Eq.* (17) residual pressure in nanocomposite ZrB_2 ceramics versus difference of the maximum sintering temperature and the room temperature.

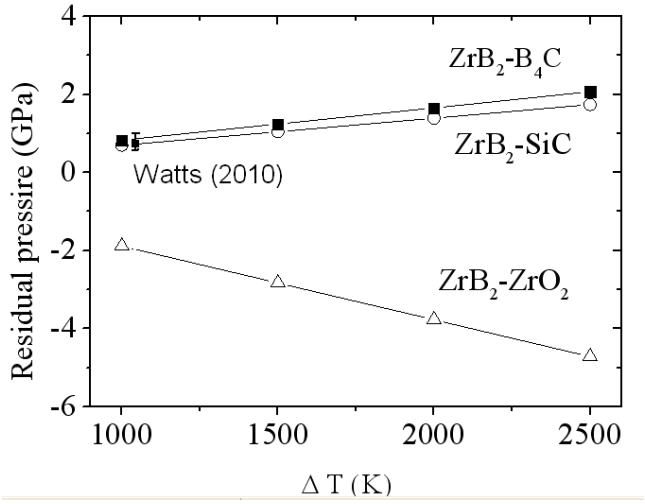


Figure 3. Residual pressure versus the cooling temperature.

Thermo-mechanical properties of matrix and potential inclusions are given in Table I. In multiphase system the residual pressures can be positive and negative. Recently Watts et al. have determined tensile residual stresses of ~800 MPa in the ZrB_2 matrix of (ZrB_2 –30 vol. SiC) ceramics using Raman spectroscopy and neutron diffraction methods [29]. Calculated tensile stresses of ~800 MPa in the ZrB_2 matrix near SiC inclusions is in good agree with experimental data ~810 MPa for ZrB_2 –30 vol. SiC [29].

Residual pressure varies linearly with difference of cooling temperature. Note, that the temperature rise caused by the energy density increasing during sintering and repeating laser scans. But as a result of increasing energy density in the sintering process not only increases the level of residual stresses, but reduces the pore size and porosity. Leu and co-authors showed that specimens of ZrB_2 have lower porosity and smaller pores after SLS at the energy density of 0.115 J/mm^2 as compared to 0.103 J/mm^2 [5].

Katcoff and Graham-Brady presented the model for compressive dynamic failure of brittle materials with circular pore flaws, which incorporates both the number density and the size distribution of flaws [26].

For predicting the mechanical properties of ceramics fabricated by SLS under dynamic loading results of multi-scale computer simulation were used. Figure 4a shows SEM image of Al_2O_3 specimen fabricated by SLS [15]. Figure 4b shows the distribution of damage parameter *D* behind the front of a shock wave with amplitude of 7.2 GPa.

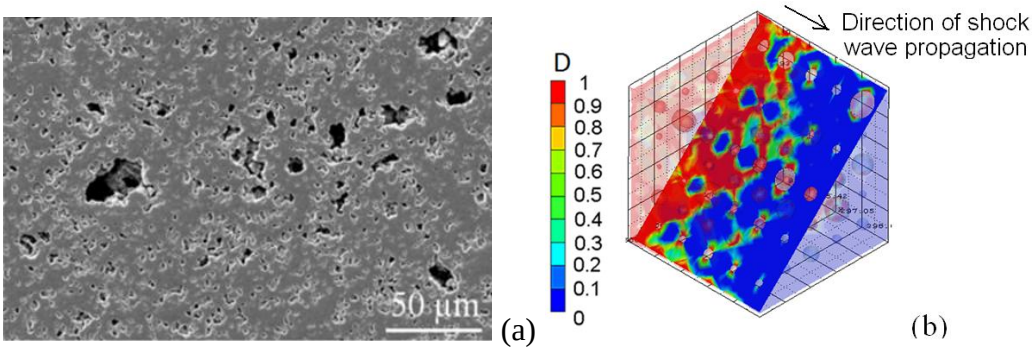


Figure 4. (a) SEM image of Al_2O_3 specimen fabricated by SLS [15]. (b) The distribution of damage parameter D behind the front of a shock wave with amplitude of 7.2 GPa.

The calculated average damage parameter versus time of shock compression of the samples of Al_2O_3 ceramics with a porosity of 5 % is shown in Figure 5a. Dashed curve shows the time dependence of damage parameter for models with random pore size distribution (see Figure 1, 4b), and the solid curve is for a fixed pore size.

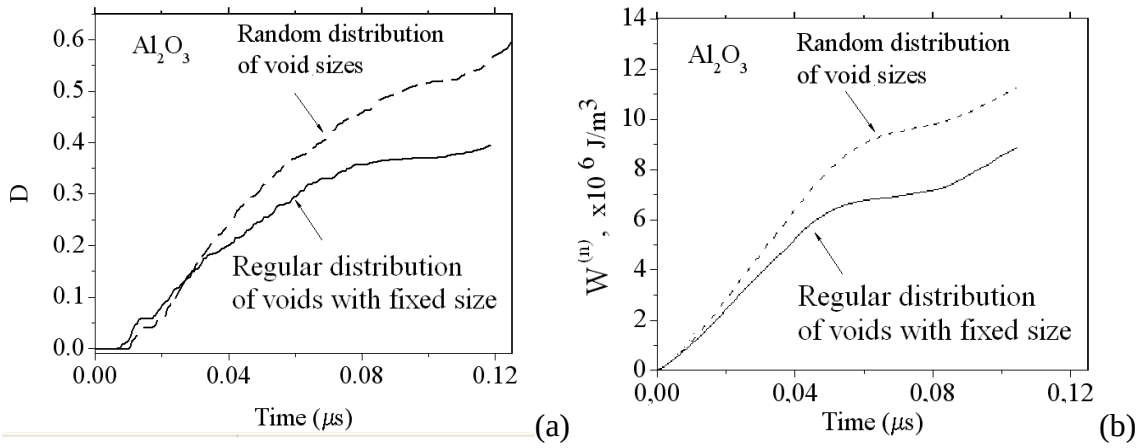


Figure 5. (a) Damage parameter of Al_2O_3 specimens behind the front of a shock wave with amplitude of 7.2 GPa versus time. (b) Specific work of stress on the inelastic deformation under shock compression versus time .

The distance between centers of neighboring pores in the models was equal to 100 μm. Therefore parts fabricated by SLS and having the same initial porosity will demonstrate a wide statistical distribution of the dynamic strength under the same loading conditions. Figure 4b shows the specific work of stress on the inelastic deformation under shock compression versus time. The dissipative properties and also the specific internal energy will have wide distribution for ceramic parts fabricated by SLS due to distribution of pore sizes. The probability of the fracture of such parts under a dynamic loading depends on the time history evolution of a distribution of local values of specific internal energy and the damage parameter on the mesoscale level.

Figure 6 shows SEM image of ZrB_2 specimen fabricated by SLS and calculated pressure and damage parameter behind the shock wave with an amplitude of 6.9 GPa. Collapse of pores is accompanied by pressure relaxation in the area surrounding the pores and the formation of cracks in the inter-porous space. Thus, the stress threshold of the dynamic fracture beginning of ceramic bodies depends on the distribution of the pore sizes and not just the average porosity. Wu and co-authors showed that the final microstructure and quality of the laser-sintered alumina powder beds are depended strongly on the laser parameters, such as laser power, laser scanning speed and laser beam size [31].

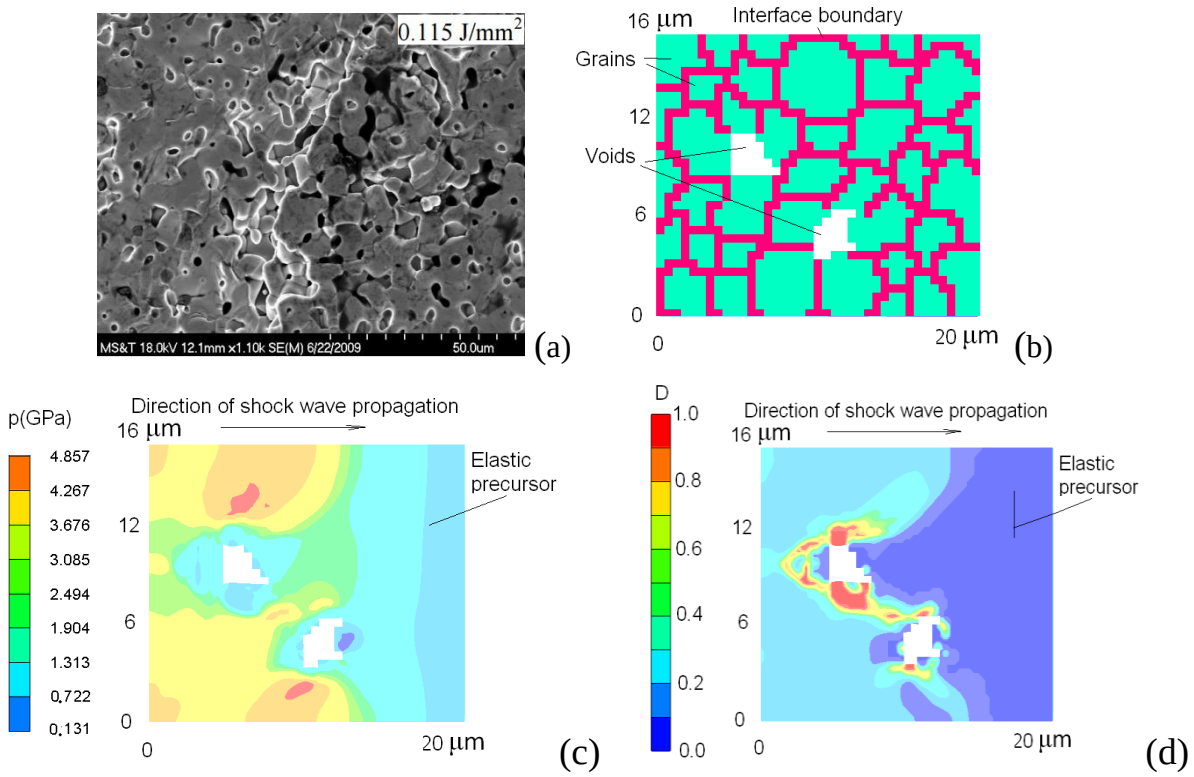


Figure 6. (a) SEM image of ZrB_2specimen fabricated by SLS [6]. (b) The cross section of the model RVE. (c) The pressure behind the shock wave with amplitude of 6.9 GPa. (d) The distribution of damage parameter behind the front of a shock wave with amplitude of 6.9 GPa.

Under short time laser-irradiation on the powder bed, the bonding particles can be ensured by solid-phase sintering. In this case the intergranular phase has a reduced strength compared with the phase of the particles. The multiscale simulation approach was used to study the mechanical behavior of ceramic parts created by the low energy SLS. Figure 7a shows the cross section of the model RVE of laser sintered ZrB_2 ceramics.

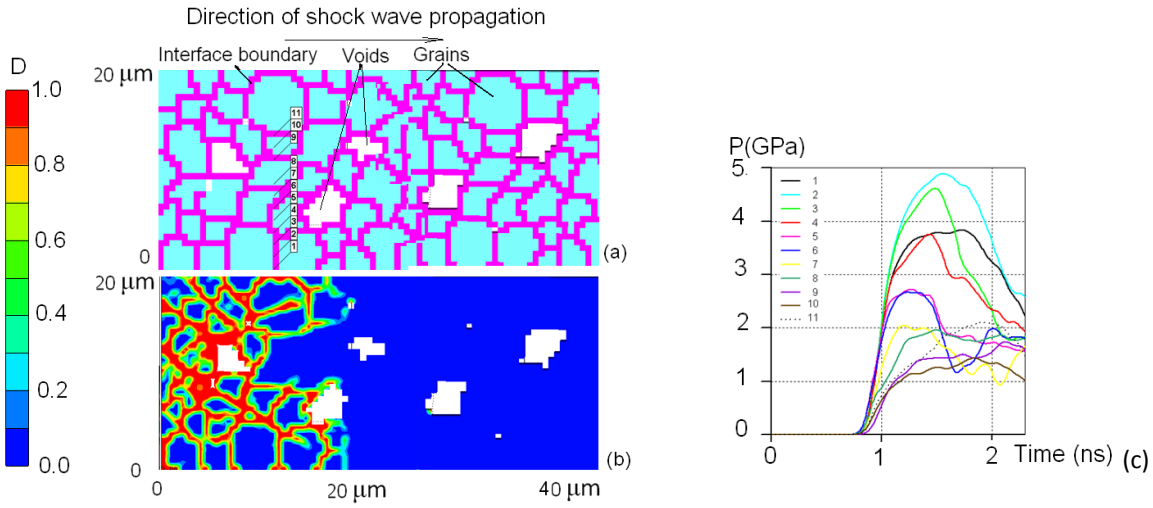


Figure 7. (a) The cross section of the model RVE. (b) The distribution of damage parameter behind the front of an elastic precursor. (c) Local pressure versus time in the points 1-11 (see Figure 6a).

Figure 7b shows the distribution of damage parameter behind the front of an elastic precursor. Figure 7(c) shows a temporal change of pressure in points behind front of the shock wave with amplitude of 7 GPa. Reviewed ceramics under dynamic compression damaged primarily at the interface boundaries. The damage is accompanied by energy dissipation and formation of fragments having dimensions comparable with the largest grain size of the powder. During compression the local pressure and the equivalent shear stress are described by distribution functions, whose parameters vary in time. Figure 8a,b show the probability densities of the normalized pressure in ZrB_2 ceramics (see Figure 7c) at 1.5 ns and 2.2 ns, respectively.

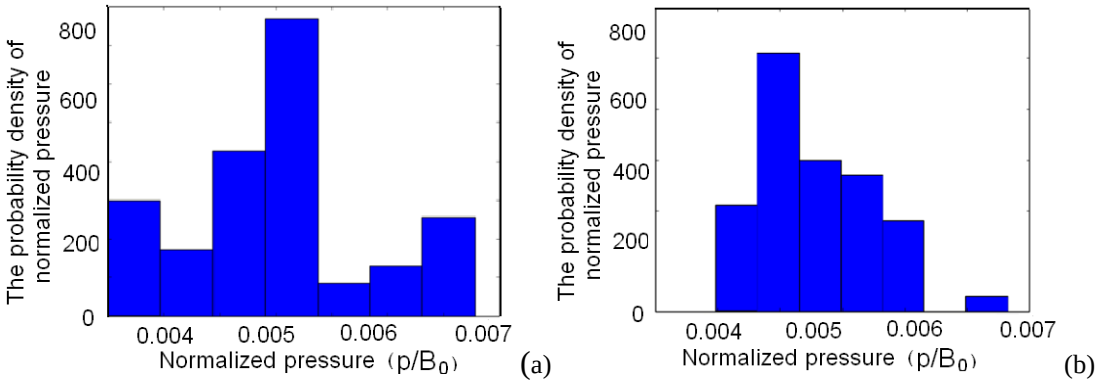


Figure 8. (a) The probability density of normalized pressure in ZrB_2 ceramics. (b) The probability density of normalized pressure in ZrB_2 ceramics.

The simulation results show the local and the average macroscopic values of stress can differ significantly under dynamic loading of laser sintered ceramics.

Since the damage and fracture always start at the micro-scale level, the prediction of the dynamic strength of parts fabricated by laser sintering requires additional data on the distribution of stresses in the materials on the mesoscale level. These data can be obtained using multiscale simulation.

The presented model can be used for examining the distribution of the state parameters and prediction of strength in the process of dynamic loading on parts and ceramic materials fabricated by additive technologies.

4 CONCLUSIONS

The multiscale simulation approach was used to study the mechanical behavior of ceramic parts created by the selective laser sintering (SLS) technology.

It was shown that the probability of the fracture of such parts under a dynamic loading depends on the time history evolution of a distribution of local values of specific internal energy and the damage parameter on the mesoscale level.

It was found that the stress threshold of the dynamic fracture beginning of ceramic bodies depends on the distribution of the pore sizes and not just on the average porosity.

The formation of frame structure by means of SLS allows increasing the dynamic limit of elasticity of the parts compared with ceramic materials with equivalent porosity.

ACKNOWLEDGMENT

This work was supported partially by the Grant from the President of Russian Federation, and by a grant from the Foundation of D. I. Mendeleev's National Tomsk State University

within the Program of increasing the competitiveness of TSU. The authors are grateful for the support of this research.

REFERENCES

[1] J. Deckers, J. Vleugels, J.-P. Kruth, Additive manufacturing of ceramics: A review. *Journal of Ceramic Science and Technology*, **5**, (4), 245–260, 2014.

[2] B. Qian, Z. Shen, Laser sintering of ceramics. *Journal of Asian Ceramic Societies*, **1**, 315–321, 2013.

[3] N. Travitzky, A. Bonet, B. Dermeik, et al., Additive manufacturing of ceramic-based materials. *Advanced engineering materials*, **16,** 729–754, 2014.

[4] M. Schwentenwein, and J. Homa, Additive manufacturing of dense alumina ceramics. *International Journal Applied Ceramic Technology*, **12**, 1–7, 2014.

[5] S. Pattnaik, M. C. Leu, and G. E. Hilmas, Investigation of laser sintering for freeform fabrication of zirconium diboride parts. *Virtual and Physical Prototyping*, **7**, 25–36, 2012.

[6] N.C. Sun, and M.C. Gupta, Laser sintering of ZrB_2. *Journal of American Ceramic Society*, **91** (5), 1729–1731, 2008.

[7] Q. Liu, Y. Danlos, B. Song, et al., Effect of high-temperature preheating on the selective laser melting of yttria-stabilized zirconia ceramic. Journal of Materials Processing Technology**, 222**, 61–74, 2015.

[8] G. N. Levy, R. Schindel, J.P. Kruth, Rapid manufacturing and rapid tooling with layer manufacturing (lm) technologies, state of the art and future perspectives. *CIRP Annals - Manufacturing Technology*, 52(2), 589–609, 2003.

[9] J. Xing, W. Sun, R.S. Rana, 3D modeling and testing of transient temperature in selective laser sintering (SLS) process. *Optik*, **124**, 301–304, 2013.

[10] C. Pelissou, J. Baccou, Y. Monerie, F. Perales, Determination of the size of the representative volume element for random quasi-brittle composites. *International Journal of Solids and Structures*, **46,** 2842–2855, 2009.

[11] T. D.P. Tran, S. D. Pang, S. T. Quek, Does representative volume element exist for quasi-brittle composites? *Materials Science and Engineering A* **528**, 7757–7767, 2011.

[12] Parshikov A.N., Medin S.A., Loukashenko I.I., and Milekhin V.A., Improvements in SPH Method by means of interparticle contact algorithm and analysis of perforation tests at moderate projectile velocities. *International Journal of Impact Engineering*, **24**, 779–796. 2000.

[13] R. Ganeriwala, T. I. Zohdi, Multiphysics modeling and simulation of selective laser sintering manufacturing processes. *Procedia CIRP*, **14**, 299–304, 2014.

[14] S. A. Khairallah, A. Anderson, Mesoscopic simulation model of selective laser melting of stainless steel powder. *Journal of Materials Processing Technology*, **214**, 2627–2636, 2014.

[15] K. Shahzad, J. Deckers, J. P. Kruth, J. Vleugels, Additive manufacturing of alumina parts by indirect selective laser sintering and post processing. *Journal Materials Proc*. Technology, **213**, 1484–1494, 2013.

[16] Y Wu, J Du, KL Choy, LL Hench, Laser densification of alumina powder beds generated using aerosol assisted spray deposition. *Journal of the European Ceramic Society*, **27** (16), 4727–4735, 2007.

[17] W. Pabst, E. Gregorova, G. Ticha, Elasticity of porous ceramics. A critical study of modulus−porosity relations. *Journal of the European Ceramic Society***, 26,** 1085–1097, 2006.

[18] E. G. Skripnyak, V.A. Skripnyak, and V. V. Skripnyak, Fracture of nanoceramics with porous structure at shock wave loadings. M. D. Furnish et al. eds. *Shock Compression of Condensed Matter -2011*. AIP Conference Proceedings 1426, Melville, NY, 2012, 965 –970.

[19] V.V. Skripnyak, E. G. Skripnyak, V. A.Skripnyak, and I.K. Vaganova, Computer simulation of fracture quasi-brittle ceramic nanocomposites under pulse loading. *11th. World Congress on Computational Mechanics (WCCM XI),* Barcelona, Spain, June 25–29, 2014, Vol. IV, 3904-3914.

[20] T. S.Jun, A. M. Korsunsky, Evaluation of residual stresses and strains using the eigenstrain reconstruction method. *International Journal of Solids and Structures*, **47** (13), 1678–1686, 2010.

[21] A. L. Tonge, K.T. Ramesh, Multi-scale defect interactions in high-rate brittle material failure. PartI: Model formulation and application to ALON. *Journal of the mechanics and physics of solids*, **86**, 117–149, 2016.

[22] T. J. Holmquist, G. R. Johnson, Response of silicon carbide to high velocity impact. Journal Applied Physics, **91** (9), 5858– 5866, 2002.

[23] S. Chakraborty, A. Shaw, A pseudo- spring based fracture model for SPH simulation of impact dynamics. *International Journal of Impact Engineering* , **58**, 84–95, 2013.

[24] Y.V. Petrov, B.L. Karihaloo, V.V. Bratov, A.M. Bragov, Multi-scale dynamic fracture model for quasi-brittle materials. *International Journal of Engineering Science*, **61**, 3**–**9, 2012.

[25] J. Kimberley, K.T. Ramesh, N.P. Daphalapurkar, A scaling law for the dynamic strength of brittle solids. *Acta Materialia*, **61**, 3509–3521, 2013.

[26] C.Z. Katcoff, L.L. Graham-Brady, Modeling dynamic brittle behavior of materials with circular flaws or pores International. *Journal of Solids and Structures*, **51**, 754–766, 2014.

[27] A. L. Tonge, K.T. Ramesh Multi-scale defect interactions in high-rate brittle material failure. PartI: Model formulation and application to ALON. *Journal of the mechanics and physics of solids*, 86, 117–149, 2016.

[28] Vaganova I.K., Skripnyak V.A., Skripnyak V.V., Skripnyak E.G., Modeling of mechanical behavior of ceramic nanocomposites. *Applied Mechanics and Materials*, **756** , 187–195, 2015.

[29] J. Watts, G.E. Hilmas, W.G. Fahrenholtz, Stress measurements in ZrB_2–SiC composites using Raman spectroscopy and neutron diffraction. *Journal of European Ceramics Society*, **30**, 2165–2171, 2010.

[30] J. W. Zimmermann, G.E. Hilmas, W.G. Fahrenholtz, Thermal shock resistance of ZrB_2 and ZrB_2–30% SiC. *Material Chemistry and Physics*, **112**, 140–145, 2008.

[31] Y. Wu, J Du, KL Choy, LL Hench, Laser densification of alumina powder beds generated using aerosol assisted spray deposition. *Journal of the European Ceramic Society*, **27** (16), 4727–4735, 2007.

[32] Q Liu, Y Danlos, B Song, et al., Effect of high-temperature preheating on the selective laser melting of yttria-stabilized zirconia ceramic. *Journal of Materials Processing Technology*, **222**, 61–74, 2015.

[33] Skripnyak V.V., Skripnyak V.A., Vaganova I.K. et al. Multiscale simulation of porous quasi-brittle ceramics fracture. *Applied Mechanics and Materials*, **756**, 196–204, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

FRACTURE OF THIN METAL SHEETS WITH DISTRIBUTION OF GRAIN SIZES IN THE LAYERS

Natalia V. Skripnyak[1,2], Vladimir V. Skripnyak[1], Vladimir A. Skripnyak[1,3]

[1] National Research Tomsk State University
36 Lenin Avenue, 634050 Tomsk, Russia
e-mail: skrp@ftf.tsu.ru

[2]Linköping University
Linköping, SE-581 83 Sweden.
e-mail: natali.skrp@mail.ru

[3]Institute of Strength Physics and Materials Sciences SB RAS
2/4, pr. Akademicheskii, Tomsk, 634021, Russia
e-mail: skrp2006@yandex.ru

Keywords: Multi-scale Simulation, Shock Waves, Grain Size Distribution, Light Alloys, Mechanical Behaviour.

Abstract. *The influence of ultra-fine grained surface layers on the fracture of light alloys sheets was studied by the method of multiscale simulation. The deformation and damaging of 3D structured elementary volumes of thin metal sheets with structured surface layers under tension and compression were calculated. The modified smooth particle hydrodynamics method was used for numerical simulation. It was found that inelastic deformation and the damage are localized at the boundary between ultrafine-grained and coarse grained layers. The deflection of cracks caused by the residual stresses lead to ductility increasing of metal sheets with layered ultrafine-grained and coarse-grained structure. Fracture of thin sheets of aluminium, magnesium and titanium alloys with nanostructured and fine grained surface layers under loading has probabilistic character and depends on parameters of layered structure.*

1 INTRODUCTION

Metallic multilayered structures can be created in thin metal sheets using a technology of surface severe plastic deformation (S^2PD) such as accumulative roll bonding (ARB) [1], ultrasonic shot peening (USSP) [2], surface severe plastic deformation under friction (SPDF) [3], surface mechanical grinding treatment (SMGT) [4], high-energy shot peening (HESP) [5,6], surface mechanical attrition treatment (SMAT) [7].

Metallic multilayers structures exhibit a very pronounced size effect where the mechanical strength depends on the layer thickness. Using S^2PD nanocrystalline layers in the surface of various materials, such as alloys with a face centered cubic (FCC) structure (Al alloy [1-3]), alloys with a hexagonal close-packed HCP structure (magnesium and alpha-titanium alloys [4-13]), and alloys with a body centered cubic (BCC) structure (stainless steel [7]) have been successfully produced.

In recent years, the microstructures and properties of surface layer were systematically investigated in various S^2PD metals and alloys, including BCC, FCC and HCP crystal structures [14].

It was found that as a result S^2PD on the surface of the bulky body are formed with surface layers of NC and UFG structures [1-14].

The fields of residual stresses are formed in the surface layers of bodies [15-17]. The magnitude and sign of residual stress change with increasing distance from the surface.

Generally, compressive stresses reach their maximum at distances ~50-60 μm from the surface. Compressive residual stresses decrease with increase in distance from the surface and change the sign.

Authors [2-10] have shown that compressive stresses in plates subjected to S^2PD operate in the near-surface layer with a thickness of 200-500 μm. Distribution of grain sizes in the layers and residual stresses in thin metal sheets influences on the strength, the fatigue strength and the strain to fracture [18-19].

Guo and co-workers [7] have shown that nanograined interface layer (NGIL) can enhance the ductility of the co-rolled surface mechanical attrition treatment (SMAT) 304 stainless steel (SS), but the ductility will not increase if the NGIL thickness goes beyond 60 μm.

Liu and co-workers [5] have shown that the nanostructured surface layers can be created by using a high-energy shot peening on the low carbon steel. The average grain size in the surface layer can be as small as tens of nanometers, and gradually increases with the distance from the surface.

The yield strength was found to be significantly enhanced without considerable degradation in the ductility and the toughness. Yang and coworkers [6] have shown that rolling or shot-peening of titanium alloys induces growth of the residual stress at the surface layers. Stress concentrations grow at the interface between ultrafine-grained (UFG), nano-grained (NG) and coarse-grained (CG) layers with increasing deformation.

Results of researches testify that reduction in the size of grain of aluminum, magnesium, and titanium alloys causes enhanced strength and ductility under quasi-static loading conditions [18,19]. It was revealed that grains of nanostructured (NS) and ultrafine-grained (UFG) alloys have size distributions. The yield strength and strength under compression and tension, elongation to fracture of light alloys depends on the grain size distribution.

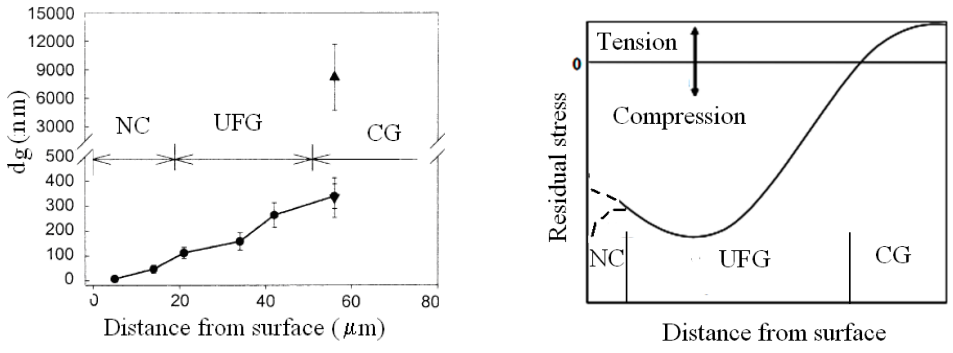
Features of deformation and fracture of thin metal sheets after treated by with S^2PD are defined by shear bands formation, nucleation and growth of damages, coalescence damages under formation of mesoscale and macroscale cracks.

Mechanisms of fracture of light metal sheets with distribution of grain sizes in the surface layers are poorly investigated. We present computational model and results of numerical simulation of damaging and fracture of aluminium and magnesium thin sheets with distribution of grain sizes in the surface layers under dynamic loading.

2 CONSTITUTIVE MODEL

The multilevel computer simulation method was used for numerical research on damage and fracture of thin sheet of aluminium, magnesium, titanium alloys with NGIL and UFG layers [20, 21]. Figure 1.a shows the variation of grain size with depth of surface layer of aluminum alloy 7075 after USSP [2].

Model layered structure thin metal sheet was created using the experimental data on grains structures.



a b

Figure 1. (a) Grain size in the surface layer of aluminum alloy 7075 after USSP. (b) Model residual stress in metal sheet after S^2PD.

Mechanical behavior of thin metal sheet with distribution of grain sizes in the surface layer was simulated under axial tension and compression at high strain rates. 3D model of the elementary volume was used for simulation of deformation and damage. The 3-D volume dimension of 1000 x 100 x 100 μm includes nanocrystalline and ultra-fine grained volumes of surface layers, coarse grained volume with a residual stress.

In the model, the second volume of the surface layer, which is formed in alloys of aluminum, magnesium and titanium as a result of S^2PD was considered. This layer is limited to the surfaces S_3 and S_4. The second layer has a thickness of ~ 100−200 μm and composed of grains having an unimodal or bimodal distribution of grain size. In this work it was assumed that the size distribution of grains is the unimodal with an average size of ~ 800 nm. The second layer is limited to the surfaces S_4 and S_5.

Figure 1.b shows the characteristic change of the residual stresses near the surface of metal alloys after SMAT [15,16]. The surface S_6 in the model correspond to the boundary on which the residual stresses close to zero. It was assumed that S_6 is at a distance of ~350 μm from the plate surface.

Mechanical behavior of structured volume is described using the approach to damaged elastic-plastic medium [15]. Smooth particles hydrodynamic (SPH) method was used for the numerical simulation [22].

The scheme of boundary conditions is shown in Fig. 2.

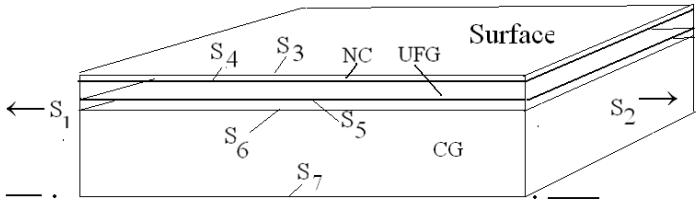


Figure 2. Boundary conditions.

The boundary conditions were written in the form:

$$u_k(x_k,t)=u_k(x_k,t),\ x_k \in S_1, x_k \in S_2,\ \sigma_{ij}(x_k,t)=0, x_k \in S_3\ . \quad (1)$$

where u_k are components of particle velocity vector, t is time, x_k are Cartesian coordinates.

It was assumed that there is no slip between the layers of the grain structure.

Initial condition describes the distribution of the residual stress and the residual strain:

$$u_k(x_k,0)=0;\ T(x_k,0)=T_0, x_k\in\Omega;\ \sigma_{kk}(x_k,0)=\sigma_R(x_k,0);\varepsilon_{kk}(x_k,0)=\varepsilon_R(x_k,0); x_k\in\Omega_{S_3\cup S_6}, \quad (2)$$

where Ω is the elementary volume, $\Omega_{S_3\cup S_6}$ is the volume in which there are residual stresses.

Kinematics of medium was described by the local strain rate tensor:

$$\dot{\varepsilon}_{ij}=\frac{1}{2}(\nabla_i u_j+\nabla_j u_i)\ , \quad (3)$$

where $\dot{\varepsilon}_{ij}$ are components of the strain rate tensor, u_i are components of particles velocity vector, ∇_i is Hamilton operator.

Components of strain rate tensor are expressed by sum of elastic and inelastic terms:

$$\dot{\varepsilon}_{ij}=\dot{\varepsilon}^e_{ij}+\dot{\varepsilon}^n_{ij}\ , \quad (4)$$

where $\dot{\varepsilon}^e_{ij}$ are components of the elastic strain rate tensor, $\dot{\varepsilon}^n_{ij}$ are components of the inelastic strain rate tensor.

Dynamics of volume is described in Lagrange coordinate system by mass conservation, momentum conservation and energy conservation equations:

$$\frac{\partial\sigma_{ij}}{\partial x_j}=\rho\frac{du_i}{dt},\ \frac{d\rho}{dt}=\rho\frac{\partial u_i}{\partial x_i},\ \rho\frac{dE}{dt}=\sigma_{ij}\dot{\varepsilon}_{ij}, \quad (5)$$

where σ_{ij} are components of stress tensor, ρ is the mass density, u_i are components of particles velocity vector, ε_{ij} is components of strain rate tensor, E is the specific internal energy per unit mass.

The bulk inelastic strain rate is described by relation:

$$\dot{\varepsilon}^n_{kk}=\frac{1}{3}\frac{\dot{D}}{(1-D)}\ , \quad (6)$$

where D is the damage parameter, the substantial time derivative is denoted via dot notation.

Damage accumulation during inelastic deformation of UFG metals leads to limited ductility of metal sheet with hardened surface layers. The local damage parameter D is introduced in the form:

$$D=\int_0^{t_f}\frac{\dot{\varepsilon}^n_{eq}}{\varepsilon^n_f}dt \quad (7)$$

where $\dot{\varepsilon}^n_{eq}=(\frac{2}{3}\dot{\varepsilon}^n_{ij}\dot{\varepsilon}^n_{ij})^{1/2}$, ε^n_f is the threshold of inelastic strain, t_f.is the time before local fracture.

For material particle the local failure criterion is written in the form:

$$D = 1. \quad (8)$$

When using SPH, numerical method criterion (8) determines the loss in strength of the damaged material in the particle [20].

Local fracture criterion (8) was used for NC and UFG light alloys at the room and elevated temperatures owing to relatively low melting temperature of light alloys. The melting temperature of aluminum and magnesium alloys is equal to~ 900 K.

Phenomenological relations (9) were used for calculation of the threshold inelastic strain ε^n_f:

$$\varepsilon_f^n \quad (d_g) / \varepsilon_f^{CG} = A_2 + (A_1 + A_2) / (1 + \exp[(x - x_0) / \bar{x}]),$$

$$\varepsilon_f^{CG} \quad = D_1 (P^* + T^*)^{D_2}, \tag{9}$$

where ε_f^{CG} is a strain to fracture for coarse grained material, x corresponds to $d_g^{-1/2}$, A_1, A_2, x_0, $\bar{x}$ are material constants, $T^*=\sigma_{sp}/P_{HEL}$, $P^*=p/P_{HEL}$, P_{HEL} is the pressure corresponding to the Hugoniot Elastic Limit, D_1, D_2 are material constants.

The threshold inelastic strain corresponds to the strain at fracture under quasistatic loading.

Figure 3 shows the dependence of normalized values of the strain at fracture versus the inverse square root of the average grain size.

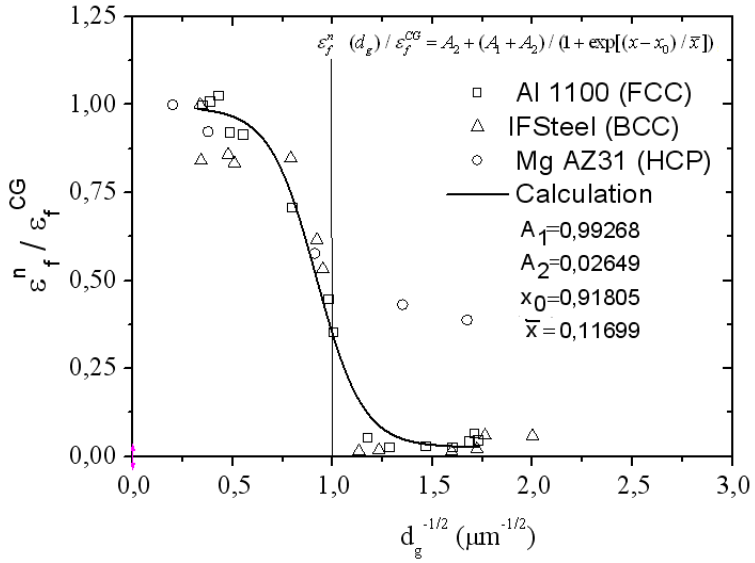


Figure 3. Normalized values of the strain at fracture vs the inverse square root of the average grain size.

The ductility of UFG light alloys increases when a relative part of coarse grains in volume is decreased. The dependence of the strain to fracture on specific volume of coarse grains in UFG Al-Mg alloy with a bimodal grain sizes distribution described by the relation [20,22]:

$$D_1 \ = 0.01 exp(\ C_{cg} / 0.363), \tag{10}$$

where D_1 is the strain to fracture under quasi-static tension, C_{cg} is the specific volume of coarse grain size of light alloys.

The criterion (8) defines a local destruction of material particles. The process of damage and fracture at a higher structural level is considered as the formation of a spatial cluster of damaged particles in an elementary volume. When using the SPH method for the numerical simulation of deformation and fracture of a model volume of plates cracking is described using the spatial formation of clusters of particles with broken bonds with neighbouring particles. The particles belonging to a given cluster is determined by the condition:

$$r_{\alpha\beta} = \sum_{i=1}^{3} (x_i^{\alpha} - x_i^{\beta})^2 > L, \tag{11}$$

where $r_{\alpha\beta}$ is the distance between neighbouring particles α and β, L is the parameter close to the particle sizes, $L > d + \varepsilon_{lim} d$, d is the size of particle, ε_{lim} is the maximum allowable strain at the crack tip.

The pressure is calculated by polynomial equation of state [22]. The stress tensor deviator is calculated by the equation:

$$d\,S^{(m)}{}_{ij} / d\,t = 2\mu(\dot{e}_{ij} - \dot{e}_{ij}{}^{p}), \tag{12}$$

where d/dt is the Jaumann derivative, μ is the shear modulus, $\dot{e}_{ij} = \varepsilon_{ij} - (1/3)\delta_{ij}\dot{\varepsilon}_{kk}$ is the deviator of the strain rate tensor, and $\dot{e}^{n}_{ij}$ is the deviator of the inelastic strain rate tensor.

The deviator of the inelastic strain rate tensor is written as:

$$\dot{e}^{p}_{ij} = (3/2)[S_{ij}\,\dot{e}^{p}_{eq} / \sigma_{eq}]. \tag{13}$$

$$(3/2)(S_{ij} - R_{ij})(S_{ij} - R_{ij}) = \sigma_s{}^{2}. \tag{14}$$

where $\dot{R}_{ij} = C_r(R_{eq}\,\dot{e}^{p}_{ij} - R_{ij}de^{p}_{eq})$, $\sigma_{eq} = [3/2\,S_{ij}\,S_{ij}]^{1/2}$, $\sigma_{ij} = -p\,\delta_{ij} + S_{ij}$, $\dot{e}^{p}_{eq} = [(2/3)\dot{e}^{n}_{ij}\,\dot{e}^{n}_{ij}]^{1/2}$, $\dot{\varepsilon}^{n}_{ij} = (1/3)\dot{\varepsilon}^{n}_{kk}\delta_{ij} + \dot{e}^{n}_{ij}$, δ_{ij} is the Kronecker delta, C_r is material parameter. C_r varies from 200 to 500 for light alloys.

For FCC, HCP and BCC alloys the yield stress was calculated by the modified Zerilli-Armstrong constitutive model [24,25]:

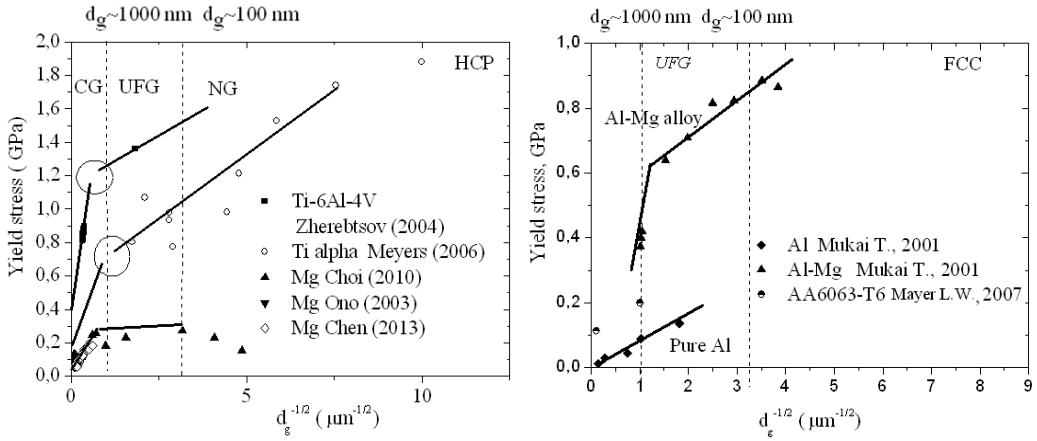
$$\begin{gathered}
\sigma_s = \sigma_{s0} + [C_6 d_g^{-1/2} + C_5(\varepsilon^{p}_{eq})^{n_1}]\exp[-T\ln(\dot{\varepsilon}_{eq}/\dot{\varepsilon}_a)] + C_1\exp[-C_3T + C_4T\ln(\dot{\varepsilon}_{eq}/\dot{\varepsilon}_b)] \text{ for HCP alloys, or} \\
\sigma_s = \sigma_{s0} + C_6 d_g^{-1/2} + C_2(\varepsilon^{p}_{eq})^{1/2}\exp[-C_3(1 - \ln\dot{\varepsilon}/\ln\dot{\varepsilon}_b)T] + \\
+ C_4[\varepsilon_r(1 - \exp(-\varepsilon_{eq}/\varepsilon_r))]^{1/2}\exp[-TC_7(1 - \ln(\dot{\varepsilon}_{eq}/\ln\dot{\varepsilon}_a))], \\
\sigma_s = \sigma_{s0} + C_6 d_g^{-1/2} + C_5(\varepsilon^{p}_{eq})^{n_1} + C_1\exp[-C_3T + C_4T\ln(\dot{\varepsilon}_{eq}/\dot{\varepsilon}_{eq0})] \text{ for BCC alloys,} \\
\sigma_s = \sigma_{s0} + C_6 d_g^{-1/2} + C_2(\varepsilon^{p}_{eq})^{1/2}\exp[-C_3T + C_4T\ln(\dot{\varepsilon}_{eq}/\dot{\varepsilon}_{eq0})] \text{ for FCC alloys.}
\end{gathered} \tag{15}$$

where C_1, C_2, C_3, C_4, C_5, C_6, C_7, n_1, ε_r , $\dot{\varepsilon}_a$, $\dot{\varepsilon}_b$ are materials parameters, T is the temperature in [K].

Parameter	Al–6% Mg–0.3% Si.		Ma2-1 (AZ31)		Ti6Al4V	
Grain size,[μm]	$1<d_g<100$	$0.1<d_g<1$	$1 < dg<100$	$0.1< dg<1$	$1< dg<100$	$0.1< dg<1$
σ_{s0} ,[MPa]	300	510	80	330	190	570
C_1,[MPa]	52		149			
C_2,[MPa]	99		160		139	
C_3, [K^{-1}]	0.0042		0.0162		0,0024	
C_4, [K^{-1}]	0.00158		0.00178		0,00043	
C_5,[MPa]			570		656	
C_6,[MPa $\mu m^{1/2}$]	326	11	303	180	1200	350
C_7, [K^{-1}]	0		0		0	
n_1	0.179		0.17		0.5	
ε_r					0.1877	
$\dot{\varepsilon}_a$, [s^{-1}]	1,0		1,0		8,25 10^5	
$\dot{\varepsilon}_b$, [s^{-1}]	1,0		1,0		1.6 10^4	

Table 1. Constitutive parameters for FCC and HCP light alloys.

The yield strength of magnesium and titanium alloys under tension versus $d_g^{-1/2}$ (d_g, average grain size) shown in Figure 4a. The yield strength of aluminium alloys under tension versus $d_g^{-1/2}$ shown in Figure 4b.



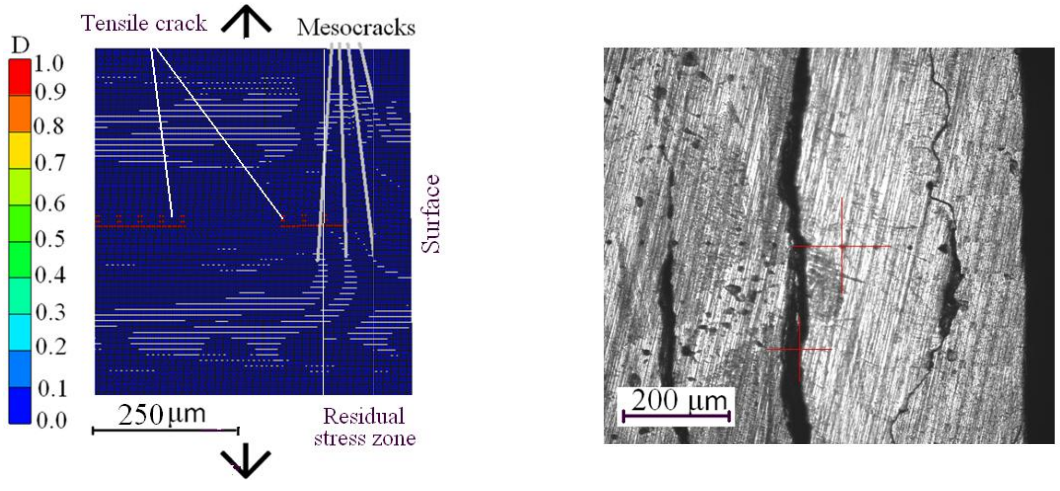
a b

Figure 4. (a) The yield strength of magnesium and titanium alloys versus $d_g^{-1/2}$ (d, average grain size). (b) The yield strength of aluminium alloys versus $d_g^{-1/2}$.

Damage accumulation in sheets of aluminum alloy Al 7075 and magnesium Ma2-1 (this is analog to AZ31 magnesium alloy) with strengthened layers under tension-compression was simulated. Numerical method was discussed in [22].

3 RESULTS AND DISCUSSION

Fig. 5a shows the damage in the section of the model layered aluminium sheet after 50 cycles of uniaxial tension - compression under stress amplitude ~1.02 σ_s. Strain rate was equal to 1 s^{-1}.



a b

Figure 5. (a) The distribution of calculated values of damage under tension-compression model the volume of a sheet of aluminum alloy with hardened surface layers. (b) Damage in the surface layer of the aluminium alloy sheet with an initial thickness of 1.5 mm after 50 cycles of axial tension-compression.

Under tension damages nucleate within the NC and UFG layers and on the boundary between layers of coarse and ultra-fine grained material. These local damages lead to meso-

cracks formation. Cracks are oriented not only across, but also along the direction of tension of a plate. Meandering cracks formed in the layers where the non-homogeneous compressive residual stresses attended. Results of fracture simulation of thin aluminium and magnesium sheets with distribution of grain sizes in the surface layers agree with experimental data. Figure 5b shows the structure of the cross-sectional specimen of Al-0.6% Mg-0.4% Si alloy after 47 cycles of tension and compression. The thickness of specimen was equal to 1.5 mm. The UFG layers have been created by the ultrasonic peening on the both surfaces of specimen. The parts of specimen have been cut into sections after testing. The cross section was done at the distance of ~ 2 mm from surface of fracture. In the section of the specimen observed microcracks oriented along the direction of the sample. Surface cracks are parallel to the plane of the sheet at different distances. The first system of cracks was formed in the interface region between UFG and CG material as predicted the results of simulation. The second system of cracks is formed in the zone of maximum gradient of residual stress. Micro-voids were created in the NC layers.

Fracture is a result of stochastic damage nucleation and growth. Structure heterogeneity may influence the random distribution of total plastic work over the model elementary volume. Therefore, fracture of fine-grained layer alloys with a bimodal grain size distribution under dynamic tension has probabilistic character.

The probability of microcracks nucleation within the material particles in layer with a grain size distribution can be described by the Eq. [18]:

$$P_r(W^p) = 1 - \exp[(\frac{W^p - W^p_{\min}}{W^p_0})^{\alpha}],\ W^p \geq W^p_{\min}, \tag{16}$$

where P_r is the Weibull probability function, $W_{p\,min}$ is the minimal possible total plastic work per unit volume at microscale level, α and W_{p0} are constant parameters of the distribution.

The damaged volume retains the resistance to plastic deformation during dynamic loading. Note that for the light alloys with bimodal grain size distributions, the ultrafine-grained volume can be considered as a quasi-brittle phase, in which the microcracks are generated during the plastic deformation while the macroscopic mechanical behavior exhibit good ductility. Thus, density of microcracks in the surface layer with grain size distribution can be written by Eq. (17) which is similar to Eq. referenced in [19]:

$$N = N_0 P_r(W^p) = N_0 \{1 - \exp[(\frac{W^p - W^p_{\min}}{W^p_0})^{\alpha}]\},\ W^p \geq W^p_{\min}, \tag{17}$$

where N is the density of microcracks, N_0 is the saturation density of microcracks at formation of mesoscopic crack.

Mesocracks nucleated within the UFG volume and propagates through the coarse grained volume. The strain to fracture of thin metal sheets with UFG layers was calculated as average strain at which cracks cross the elementary volume.

Figure 6 shows the calculated strain to fracture versus thickness of aluminum, magnesium and alpha-titanium sheets with surface UFG layers. Values of ε_f at tension were calculated at room temperature and strain rates ~ 1 s^{-1}.

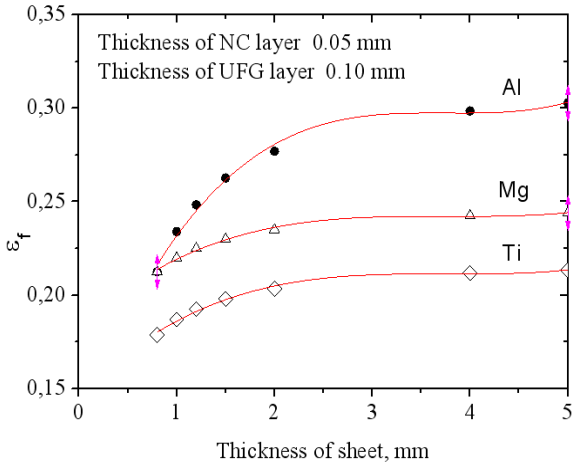


Figure 6. The strain to fracture vs thickness of aluminium, magnesium and titanium sheets.

The influence of surface nanostructured layers on the strain to fracture decreases with increasing thickness of the sheets. Since the thickness of the structured surface layers does not exceed 0.02-0.025 mm, the influence of layers on the resistance to plastic flow and the strain to fracture of metal sheet becomes insignificant when the thickness of the sheet more ~2.5 mm.

4 CONCLUSIONS

The multiscale computer simulation was used for the simulation of fracture of thin sheet of light alloys with distribution of grain sizes in the surface layers.

Fracture of fine-grained alloys under dynamic loading has probabilistic character and depends on parameters of structure heterogeneity. Formation of macro-scale failure zone is a result of several processes of structure evolution including damage nucleation, damage growth, and coalescence of damages. Damage nucleation is associated with strain localization at mesoscale level. Results of computer simulation demonstrate that the high strain localization in UFG alloys under dynamic loadings depends on the ratio between volume concentrations of fine and coarse grains. Fine precipitates in alloys not only affect the hardening but also lead to change the influence of the grains size distribution on volume concentration of shear bands. The dynamic ductility of UFG light alloys is increased when specific volume of coarse grains is greater than 30 %. Results of computer simulation can be used for estimation of grains size distribution influence on the dynamic strength and ductility of UFG alloys processed by severe plastic deformation methods.

ACKNOWLEDGMENT

This work was supported partially by the Grant from the President of Russian Federation, and by a grant from the Foundation of D. I. Mendeleev's National Tomsk State University within the Program of increasing the competitiveness of TSU. The authors are grateful for the support of this research.

REFERENCES

[1] S.O. Gashti, A. Fattah-alhosseini, Y. Mazaheri, M.K. Keshavarz, Effects of grain size and dislocation density on strain hardening behavior of ultrafine grained AA1050 processed by accumulative roll bonding. *Journal of Alloys and Compounds*, **658**, 854e861, 2016.

[2] X. Wu, N. Tao, Y. Hong et al. Microstructure and evolution of mechanically-induced ultrafine grain in surface layer of AL-alloy subjected to USSP. *Acta Materialia*, **50**, 2075–2084, 2002.

[3] Y. Huang, L. Wan, S. Lv, et al. Gradient micro-structured surface layer on aluminum alloy fabricated by in situ rolling friction stir welding. *Materials and Design.* **52**, 821–827, 2013.

[4] Y. Yin, W. Xu, Q. Sun, L. Xiao, J. Sun, Deformation and fracture behavior of commercially pure titanium with gradient nano-to-micron-grained surface layer. *Transaction Nonferrous Metal Society of China*, **25**, 738−747, 2015.

[5] G. Liu, S.C. Wang, X.F. Lou, J. Lu and K. Lu. Low carbon steel with nanostructured surface layer induced by high-energy shot peening. *Scripta materialia*, 44, 1791–1795, 2001.

[6] Yang, D.K. and Hodgson, P.D., Tough ultrafine-grained Ti through multilayering and grading, Scripta materialia, 68, 309–312, 2013.

[7] X. Guo, A.Y.T. Leung, A.Y. Chen, H.H. Ruan, J. Lu Investigation of non-local cracking in layered stainless steel with nanostructured interface. Scripta Materialia, 63, 403-406, 2010.

[8] J. Zhang, X. Ou, Thermal stability of nanocrystalline in surface layer of magnesium alloy AZ91D. Transaction Nonferrous Metal Society of China, 20, 1340−1344, 2010.

[9] R.K. Nalla, I. Altenberger, U. Noster et al. On the influence of mechanical surface treatments/deep rolling and laser shock peening/on the fatigue behavior of Ti64. *Materials Science and Engineering* , **A355**, 216–230, 2003.

[10] Wen M., Liu G., Gu J., et al. Dislocation evolution in titanium during surface severe plastic deformation. *Applied Surface Science*, **255,** 6097–6102, 2009.

[11] H.Q. Sun, Y.N. Shi, M.-X. Zhang, K. Lu. Surface alloying of an Mg alloy subjected to surface mechanical attrition treatment. *Surface and Coatings Technology*, **202**, 3947-3953, 2008.

[12] K. Lu, J. Lu Nanostructured surface layer on metallic materials induced by surface mechanical attrition treatment. *Materials Science and Engineering*, A **375-377**, 38-45, 2004.

[13] Y. Wei, B. Liu, L. Hou, et al. Characterization and properties of nanocrystalline surface layer in Mg alloy induced by surface mechanical attrition treatment. *Journal of Alloys and Compounds*, **452,** 336-342, 2008.

[14] X. Liao and Y. Zhao Surface. Nanocrystallization by surface mechanical attrition treatment. *Materials science forum*, **579**, 91-108, 2008.

[15] N. S. Prabhu, J. Joseyphus, T. S. N. Sankaranarayanan, B. R. Kumar Residual stress analysis in surface mechanical attrition treated (SMAT) iron and steel component materials by magnetic barkhausen emission technique, *IEEE Transactions on Magnetics*, 48, 4713 – 4717, 2012.

[16] K.Y. Luo, J.Z. Lu, Q.W. Wang, et al. Residual stress distribution of Ti-6Al-4V alloy under different ns-LSP processing parameters. *Applied Surface Science*, **285, B**, 607, 2013.

[17] M. Korzynski, T. Zarski, K. Korzynska Surface layer condition and the fatigue strength of an AZ91 alloy after ball peening. *Journal of Materials Processing Technology*, **211,** 1982–1988, 2011.

[18] Vinogradov, V., Hashin Z. Probabilistic energy based model for prediction of transverse cracking in cross-ply laminates. Int. J. of Solids and Structures. **42**, 365–392, 2005.

[19] L. Zhu, S. Shi, K. Lu, and J. Lu, A statistical model for predicting the mechanical properties of nanostructured metals with bimodal grain size distribution, *Acta Materialia*, **60**, 5762–5772, 2012.

[20] V.A. Skripnyak, N.V. Skripnyak, E.G. Skripnyak. Mechanical behavior of light alloys with bimodal grain size distribution. *Applied Mechanics and Materials*, **756**, 205-213, 2015.

[21] V.A. Skripnyak Mechanical behavior of nanostructured and ultrafine-grained materials under shock wave loadings. Experimental data and results of computer simulation, *Shock Compression of Condensed Matter-2012*, AIP Conference Proceedings 1426 , AIP, Melville, NY, 2012, 965-970.

[22] V.A. Skripnyak, E.G. Skripnyak, N.V. Skripnyak, et al. Failure mechanisms of light alloys with a bimodal grain size distribution. *11th. World Congress on Computational Mechanics (WCCM XI)*, Barcelona, Spain, 25-29 June, 2014. Vol. IV. P. 3915-3925.

[23] A. N. Parshikov, S.A. Medin. Smoothed particle hydrodynamics using interparticle contact algorithms. *Journal Computation Physics*, **180**, 358-382, 2002.

[24] K. E. Prasad, B. Li, N. Dixit, et al. The dynamic flow and failure behavior of magnesium and Magnesium Alloys. *JOM*, **66**, (2), 291-304, 2014.

[25] J. Li, F. Li, J. Cai et al. Comparative investigation on the modified Zerilli–Armstrong model and Arrhenius-type model to predict the elevated-temperature flow behaviour of 7050 aluminium alloy. *Computational Materials Science*, **71**, 56–65,2013.

[26] M.A. Meyers, A. Mishra, D.J. Benson, Mechanical properties of nanocrystalline materials. *Progress in Materials Science*, **51**, 427–556, 2006.

[27] T. Mukai, and K. Higashi. Ductility enhancement of ultra fine-grained aluminum under dynamic loadings. *Scripta materialia*, **44**, 1493–1496, 2001.

[28] H.J. Choi, , et. al., Deformation behavior of magnesium in the grain size spectrum from nano- to micrometer. *Material Science Engineering*, **A 527**, 1565–1570, 2010.

[29] N. Ono, R. Nowak, S. Miura. Effect of deformation temperature on Hall–Petch relationship registered for polycrystalline magnesium. *Materials Letters*, **58**, 39– 43, 2003.

[30] L.W. Mayer, M. Hackauf, l. Kruger, I. Shneider Compressive behavior of ultrafine-grained AA6063T6 over a wide range of strains and strain rates. *International Journal of Material Research*, **98**, (3), 1-9, 2007.

[31] S. Zherebtsov, E. Kudryavtsev, S. Kostjuchenko et al. Strength and ductility-related properties of ultrafine grained two-phase titanium alloy produced by warm multiaxial forging. *Materials Science and Engineering*, **A 536**, 190–196, 2012.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

BRITTLE OR QUASI-BRITTLE FRACTURE OF CERAMIC NANOCOMPOSITES UNDER DYNAMIC LOADING

Evgeniya G. Skripnyak[1], Vladimir V. Skripnyak[1], Vladimir A. Skripnyak[1,2], Nataliya V. Skripnyak[1,3], and Irina K. Vaganova[1]

[1] National Research Tomsk State University
36 Lenin Avenue, 634050 Tomsk, Russia
e-mail: skrp@ftf.tsu.ru

[2] Institute of Strength Physics and Materials Sciences SB RAS
2/4, pr. Akademicheskii, Tomsk, 634021, Russia
e-mail: skrp2006@yandex.ru

[3] Linköping University
Linköping, SE-581 83 Sweden
e-mail: natali.skrp@mail.ru

Keywords: Mechanical Properties, Nanostructured Ceramics, Dynamic Loading, Fracture.

Abstract. *Multiscale computer simulation was used for simulation of the damage nucleation and fracture of $ZrB_2 - ZrO_2$, $ZrB_2 – B_4C$, $Al_2O_3 - ZrO_2 - Y_2O_3$, $Al_2O_3 - B_4C$ nanocomposites under dynamic loading. The aim of research was the study transition from brittle to quasi-brittle fracture of nanocomposites under dynamic loading. It was shown that isolated micro- and mesoscale cracks can be nucleate in ceramic nanocomposites near voids under stress pulse amplitude less than the Hugoniot elastic limit. The critical fracture stress on mesoscale level depends not only on relative volumes of voids and particles concentration, but also sizes of corresponding structure elements. Results of simulation have shown the Hugoniot elastic limit and ceramics damage kinetics under dynamic loading depends on a volume concentration of nano-particles and nano-voids clusters*.

1 INTRODUCTION

Ceramic nanocomposites constitute an important class of constructional materials for high temperature applications in the aerospace industry: hypersonic re-entry vehicles, rocket nozzle inserts, and cutting tools, wear resistant parts etc. Thus, for designing materials and predicting the destruction of new ceramic composites for engineering applications need an adequate theoretical model that can be used in computer simulation.

In the present paper, we propose a computational model to predict the strength of quasi-brittle nanocomposites such as $ZrB_2 - ZrO_2$, $ZrB_2 – B_4C$, $Al_2O_3 - ZrO_2 - Y_2O_3$, $Al_2O_3 - B_4C$

in wide range of strain rates. Considered nanocomposites are a class of the ultra-high temperature ceramics (UHTC).

The matrix of UHTC composites make up 15 boride, carbide and nitride compounds which have the melting temperature above 3250 K [1].These compounds are brittle at low homologous temperatures and demonstrate not high enough strength and toughness at quasi-static and dynamic loading.

New UHTC nanocomposites possess enhanced strength and fracture toughness at quasi-static loading [2]. Effective way to increase strength UHTC is the creation on their base nanocomposites containing the ceramic nano-sized particles, graphite flakes, carbon fibres, and metal or ceramic whiskers. Mechanical behaviour of UHTC composites under intensive dynamic loadings is not well understood.

2 MODEL OF MECHANICAL BEHAVIOR OF NANOCOMPOSITES

Mechanical properties of nanocomposites based on refractory compounds depend not only on their structure and composition, but also on the manufacturing technology. High-temperature isostatic pressing UHTC composites occurs at temperatures above 1500 K, while the temperature at selective laser sintering or sintering spark plasma is above 2000 K.

When the ceramic composite is cooled from sintering temperature to room temperature, residual stresses arise around inclusions of strengthening phases. Residual stress and bulk defects in the structure have a negative impact on strength properties of ceramic products.

Residual stresses have to be considered while predicting the mechanical behaviour of ceramic materials under external impacts.

Residual pressure around the particle can be calculated by the formula [3]:

$$p_{res} = [(\alpha_m - \alpha_{inc})(T - T_r) E_m E_{inc} (\frac{d}{r+d})^3] / [E_{inc}(1-2\nu_m) + 2E_m(1-2\nu_{inc})], \qquad (1)$$

where p_{res} is the residual pressure, α_m, α_{inc} are the linear thermal expansion coefficients of matrix and inclusion, respectively, T is the sintering temperature, and T_r is the room temperature, d is the particle diameter, r is the distance from the surface of particle E_m, E_{inc}, are the Young's moduli of matrix and inclusions, and ν_m, ν_{inc} are the Poisson's ratios of matrix and inclusions, respectively.

The calculated residual stresses in ZrB_2 matrix composites and different ceramic inclusions shown in Figure 1. Residual pressures were obtained for the difference between the sintering temperature and the room temperature ~1500 K.

Note, the residual pressures can have a different sign for various inclusions. Residual pressures can causes local tensile or compression of matrix near the particle. If the residual stresses are tensile, the strength of ceramic composites under tension is significantly lower than under compression.

Thus, this effect is advisable to apply for strengthening of nanocomposites using inclusions, which lead the formation positive pressures around nanoparticles.

The value of the residual pressure around the inclusions is less than the Hugoniot elastic limit, but is comparable to the quasi-static tensile strength of ZrB_2 matrix. As shown in Figure 2 the residual pressure lead to changing of the fracture toughness of ZrB_2 matrix composites.

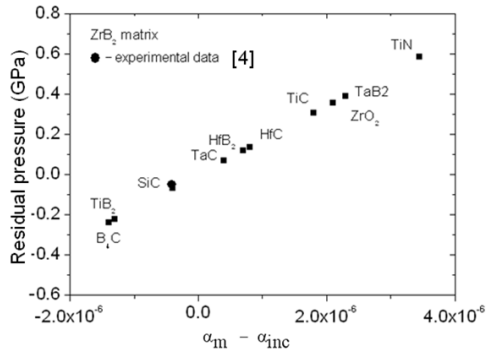


Figure 1. Residual pressure vs the difference between the linear thermal expansion coefficients of matrix and inclusions.

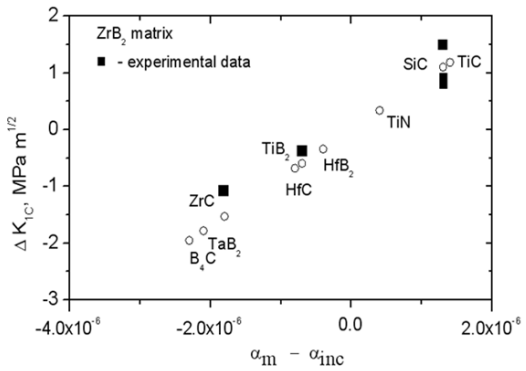


Figure 2. Fracture toughness of ZrB_2 based composites vs the difference between the linear thermal expansion coefficients of matrix and inclusions.

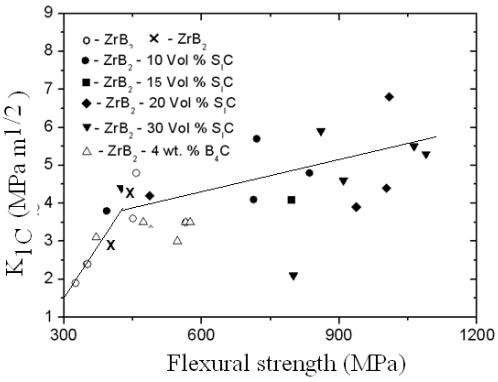


Figure 3. Fracture toughness of ZrB_2 based composites versus flexural strength.

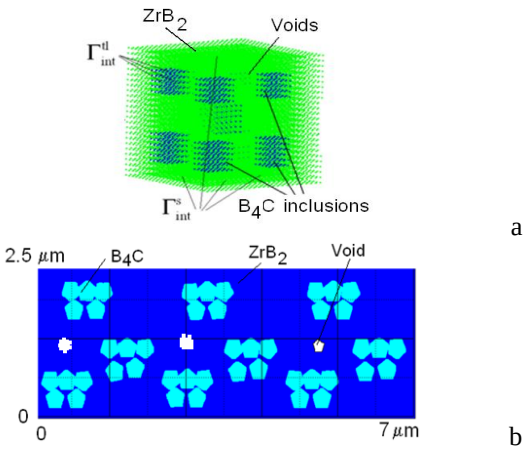


Figure 4. (a) Representative volumes of composites ZrB_2-B_4C with uniform distribution of inclusions. (b) Clusters of inclusions.

The correlation between the fracture toughness and the flexural strength of ZrB_2−B_4C and ZrB_2−SiC composites is shown in Figure 3. Symbols are experimental data [1-4].

When the concentration of inclusions exceeds the percolation threshold (~30 %) the frame structure can be formed in the composites. The scatter of the experimental data shows that the strengthening and toughening of composites is caused not only the residual stress, but also other factors [5].

To study the dynamic strength of UHTC nanocomposites we used the method of multiscale simulation [6-8]. The calculated values of residual stresses were used in the simulation of loading of the elementary volume of ZrB_2 matrix nanocomposite under shock wave loadings.

The mechanical response of the nanocomposites in the amplitude range from 0.8 to 2 of the Hugoniot elastic limit (~15 GPa) was numerically simulated. The Hugoniot elastic limit of ZrB_2 ceramics with a porosity of ~7 % is equal to 7.11 GPa.

The elementary volumes (EV) of nanocomposite materials, used in simulation, are shown in Figure 4. Figure4a shows the volume with regular distribution of nanoparticles in matrix grain. Figure 4b shows the section of volume with clusters of nanoparticles.

The modified smooth particle hydrodynamics (SPH) method was used for numerical simulation [9]. Smoothed particles with sizes of ~20 nm were used in numerical simulation shock loading of ceramic nanocomposites. Grains of matrix and inclusions had a different mechanical impedances ρC_l (ρ is the mass density, and $_{Cl}$ is the longitudinal sound velocity).

Material parameters for the matrix and inclusions of ceramic nanocomposites are shown in Table 1.

Phase	ρ, 10^3 kg/m^3	E, GPa	Poisson Ratio	α, 10^{-6} K^{-1}	C_l, 10^3 m/c	μ, GPa	σ_{HEL}, GPa
Al_2O_3	3.97	395.8	0.254	8.8	10.92	156.5	9-14
t–ZrO_2 (3mol.% Y_2O_3)	6.05	219.0	0.312	12.2	7.07	83.0	8.9-17.0
m–ZrO_2	5,6	236	0.303	8.0-10.6	7.5	86.4	
ZrB_2	6.09	450	0.128	5.9-6.5	8.8	212	7.11
SiC	3.22	454	0.163	4.7	11.89	195	13.2-14.7
B_4C	2.52	432-463	0.151-0.18	4.5	13.42	188	16.0-17.1

Table 1: Material parameters of ceramic phases.

The impedance ρC_l for ZrB_2, B_4C are: 5.08, 3.55 [10^7 kg/m^2 s], respectively. Therefore, on the mesoscale level, shock waves are reflected from boundaries of matrix and inclusions. In the result, particle velocities, stresses, and temperature can be distributed non-uniformly in the elementary volume of nanocomposite during hundreds nanoseconds.

Mechanical behavior of ceramic matrix and inclusions were described using the model of damaged medium [4−9].

The damage parameter D is determined by the relation [4, 5]:

$$D = \int_0^{t_f} \frac{\dot{\varepsilon}_{eq}^n}{\varepsilon_f^n} dt \ , \qquad (2)$$

where $\dot{\varepsilon}_{eq}^n = [(2/3)\dot{\varepsilon}_{ij}^n \dot{\varepsilon}_{ij}^n]^{1/2}$ denote an intensity of inelastic strain rate tensor, $\varepsilon_f{}^n$ is the threshold of inelastic deformations of material particles (the volumes of the phases at the micro level) at fracture , $\varepsilon_f^n = D_1 (P^* + T^*)^{D_2}$, $T^* = \sigma_{sp}/P_{HEL}$, $P^* = p/P_{HEL}$, P_{HEL} is pressure correspond to the Hugoniot elastic limit, $\sigma_{sp} = (3\rho C_l \dot{\varepsilon}_{eq} K_{1C}^2)^{1/3}$, C_l is the longitudinal sound velocity [9].

The thresholds stress of fracture for ceramic matrix and inclusions under compression in wide range of strain rates can be described by Eq. (3), if the volume concentration of inclusions is less than 30 % and porosity is less than 10 % [10,11].

$$\sigma_f / \sigma_0 = 1 + (\dot{\varepsilon}_{eq} / \dot{\varepsilon}_0)^{2/3}, \ \sigma_0 = 4.25 \, K_{1C} \eta^{1/4} / [\alpha]^{1/2}, \ \dot{\varepsilon}_0 = 4.25 K_{1C} \eta^{3/4} / [\alpha \rho E]^{1/2}, \qquad (3)$$

where α is a specific volume of voids, ρ is the mass density, K_{1C} is the fracture toughness, E is the Young's modulus, and η is a square density of unstable microcracks [10, 11].

Note, the fracture toughness of the matrix and inclusions have been corrected for residual stresses.

Averaged in elementary volume of material the threshold stress of fracture is used in Eq (2) for determination of σ_{HEL}.

The following parameters of Eq (3) for ZrB_2 matrix and for B_4C inclusions were used in the calculations:

ZrB_2: E =495 GPa, ρ =6.11 103 kg/m^3, K_{1C} = 3.5 MPa m$^{1/2}$, σ_{HEL} =7.11 GPa, p_{HEL}=3.07 GPa, D_1= 0.1, D_2 =1., Cl = 9.233 km/s, σ_{sp}=0.5 GPa, σ_{sd}=0.5 GPa ,

B_4C: E =432 GPa, ρ =2.52 10^3 kg/m^3, K_{1C} = 5 MPa m$^{1/2}$, σ_{HEL} =16 GPa, p_{HEL}= 7.2 GPa, D_1=0.1 , D_2 =1.0, Cl =12.8 km/s, σ_{sp}=0.32 GPa, σ_{sd}=13.2 GPa.

The Hugoniot elastic limit and the effective sound velocity of composite materials are determined using simulation results of shock wave propagation in the elementary volume of material [6-8].

Using averaged values of sound velocities and the elastic limit of the Hugoniot elastic limit the shear strength of ceramic nanocomposites σ_{sd} can be determined by the relation:

$$\sigma_{sd} = 1.5(1-\frac{<C_b>^2}{<C_l>^2})\sigma_{HEL}, \quad (4)$$

where C_b, C_l are the effective bulk sound velocity and the longitudinal sound velocity, respectively, σ_{HEL} is the Hugoniot elastic limit of composites.

3 RESULTS AND DISCUSSION

Calculated values of the equivalent stress in the ZrB_2 matrix versus specific values of shock pressure are shown in Figure 5. The relaxation of the equivalent stress at high strain rates is mainly caused by a nucleation of micro-cracks around pores and inclusions.

Figure 6 shows the nucleation of cracks around voids under shock compression with amplitude of ~ 10 GPa. Initial distribution of B_4C nanoparticles and voids in section of the volume is shown in Figure 4b.

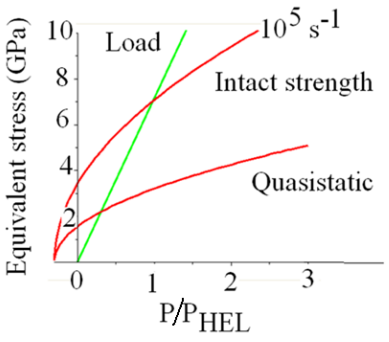


Figure 5. The intact strength and strength for the failed polycrystalline ZrB_2.

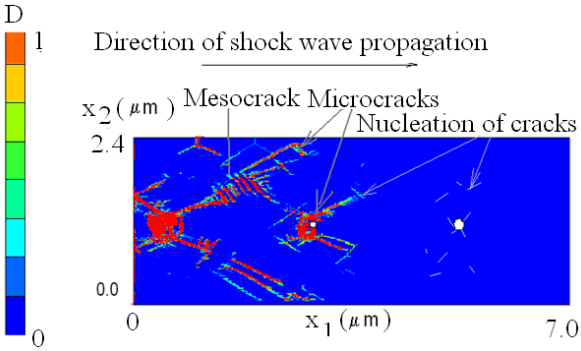


Figure 6. Crack nucleation behind the front of shock wave.

As showed Figure 6, the damage of ZrB_2 – B_4C nanocomposites near voids can be formed under pressure amplitude less than P_{HEL} of matrix. The relaxation of the equivalent stress is associated with cracks nucleation. In consequence of this, the Hugoniot elastic limit of ceramic nanocomposites decreases proportionally to specific volume of voids. Under shock compression stress field around cluster of nano-particles is close to that which occurs near the particle of micron size. Used multiscale model allows considering this phenomenon.

In the considered nanocomposites clusters of inclusions lead to increased resistance to the growth of cracks or even complete stop. As a result the transition from brittle to quasi-brittle fracture of nanocomposites under dynamic loading takes place.

The Hugoniot elastic limits for non-porous nanocomposites were evaluated according to the results of modeling the nucleation of damage in the elementary volume of material under shock compression. Obtained for the ZrB_2 or Al_2O_3 matrix nanocomposites the Hugoniot elastic limits are given in the Table 2.

Volume concentration of inclusions (%)	0	5	10	15	20	25	30
Nanocomposite	Hugoniot elastic limit , σ_{HEL} (GPa)						
ZrB_2–B_4C	7.11	7.43	7.8	8.5	9.1	9.55	10.1
ZrB_2 – t ZrO_2	7.11	7.45	8.9	8.53	9.04	9.4	10.01
ZrB_2 – Al_2O_3	7.11	7.30	7.43	7.64	7.9	7.8	7.7
Al_2O_3–ZrO_2,	9.1	9.35	9.67	9.9	10.01	10.3	10.4
Al_2O_3–B_4C	9.07	9.36	9.77	10.05	10.65	11.5	11.1

Table 2: Theoretical values of the Hugoniot elastic limit of several ceramic nanocomposites.

In considered groups of composites the Hugoniot elastic limits increase with increasing concentration of nano-inclusions from 0 to 30 %. In considered ZrB_2 matrix nanocomposites the greatest relative increase of the Hugoniot elastic limit is observed at same concentrations of nanoparticles B_4C and ZrO_2.

In Al_2O_3 matrix nanocomposites the greatest relative increase in the Hugoniot elastic limits was caused by B_4C nanoparticles.

4 CONCLUSIONS

Multi-scale model proposed for computer simulation of the mechanical behaviour of UHTC nanocomposites.

The model allows investigating the effect of structures of nanocomposites on their mechanical properties under dynamic loading.

Model was used for simulation of $ZrB_2 - ZrO_2$, $ZrB_2 – B_4C$, $Al_2O_3 - ZrO_2 - Y_2O_3$, $Al_2O_3 - B_4C$ nanocomposites under shock compression with amplitudes from 6 to 15 GPa.

In studied UHTC nanocomposites the dynamic shear strength increases proportionally volume concentration of inclusions in the range up to 30 %.

Model predicts that isolated micro- and mesoscale cracks nucleate in UHTC nanocomposites near voids under stress pulse amplitude less than the Hugoniot elastic limit.

The Hugoniot elastic limits of nanocomposites decrease proportionally to porosity in the range from 0 to 10 %.

ACKNOWLEDGMENT

This work was supported partially by the Grant from the President of Russian Federation, and by a grant from the Foundation of D. I. Mendeleev's National Tomsk State University within the Program of increasing the competitiveness of TSU. The authors are grateful for the support of this research.

REFERENCES

[1] S.Q. Guo, Densification of ZrB_2-based composites and their mechanical and physical properties, Journal of the European Ceramic Society, **29**, 995–1011, 2009.

[2] J.J. Melendez-Martınez, A. Domınguez-Rodrıguez, F. Monteverde et al. Characterisation and high temperature mechanical properties of zirconium boride-based materials**,** Journal of the European Ceramic Society, **22**, 2543**–**2549, 2002.

[3] I.A. Stepanov, V.A. Skripnyak, S.P. Andrietz, et.al., *Nuclear physics and engineering*, **2**, 1–10, 2011.

[4] V.V. Skripnyak, V.A. Skripnyak, I.K. Vaganova, et al. Multiscale simulation of porous quasi-brittle ceramics fracture. *Applied Mechanics and Materials*, **756,** 196-204, 2015.

[5] V.V. Skripnyak, E.G. Skripnyak, V.A. Skripnyak, and I.K. Vaganova "Computer simulation of fracture quasi-brittle ceramic nanocomposites under pulse loading", Proc. 11th. World Congress on Computational Mechanics (WCCM XI), edited by Eugenio Oñate, Xavier Oliver and Antonio Huerta, CIMNE Publ., Barselona, Spain, 2014, **IV,** 3904**–**3914, 2014.

[6] C. C. Holland and R. M. McMeeking, The influence of mechanical and microstructural properties on the rate-dependent fracture strength of ceramics in uniaxial compression**.** *International Journal of Impact Engineering*, **81,** 34-49, 2015.

[7] E. G. Skripnyak, V.A. Skripnyak, and V. V. Skripnyak. Fracture of nanoceramics with porous structure at shock wave loadings, *Shock Compression of Condensed Matter*. AIP Conf. Proc, **1426**, N.Y. 965 **–**970, 2012.

[8] E.G. Skripnyak, V.A. Skripnyak, I.K. Vaganova, A.S. Yanyushkin, et. al., *Izvestiya Vuzov. Physica*, (in Ruassian), **55**, (7/2), 119**–**123, 2012.

[9] E.G. Skripnyak, V.V. Skripnyak, I.K. Vaganova, and V.A. Skripnyak. Fracture of Ceramic Materials under Dynamic Loadings. *19th European Conf. on Fracture (ECF19)*, Kazan, Russia, Paper 639, Kazan, 2012.

[10] J. Kimberley, K.T. Ramesh, and N.P. Daphalapurkar, A Scaling law for the dynamic strength of brittle solids. *Acta Mater.*, **61**, 3509–3521, 2013.

[11] K.T. Ramesh, J. D. Hogan, J. Kimberley, and A Stickle, A review of mechanisms and models for dynamic failure, strength, and fragmentation. *Planetary and Space Science*, **107**, 10–23, 2015.

[12] B. Paliwal and K.T. Ramesh, Effect of crack growth dynamics on the rate sensitive behavior of hot-pressed boron carbide. *Scripta Materialia*, **57**, 481–484, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

REMOVING THE STABILIZATION PARAMETER IN FITTED AND UNFITTED SYMMETRIC NITSCHE FORMULATIONS

Christoph Lehrenfeld[1]

[1] Institute for Computational and Applied Mathematics, University of Münster
Einsteinstr. 62, D - 48149 Münster, Germany
e-mail: christoph.lehrenfeld@gmail.com

Keywords: Nitsche's method, unfitted finite elements, discrete lifting, stability parameter

Abstract. *In many situations with finite element discretizations it is desirable or necessary to impose boundary or interface conditions not as essential conditions – i.e. through the finite element space – but through the variational formulation. One popular way to do this is Nitsche's method. In Nitsche's method a stabilization parameter λ has to be chosen "sufficiently large" to provide a stable formulation. Sometimes discretizations based on a Nitsche formulation are criticized because of the need to manually choose this parameter. While in the discontinuous Galerkin community variants of the Nitsche method – known as "interior penalty" method in the DG context – are known which do not require such a manually chosen stabilization parameter, this has not been considered for Nitsche formulations in other contexts. We introduce and analyse such a parameter-free variant for two applications of Nitsche's method. First, the classical Nitsche formulation for the imposition of boundary conditions with fitted meshes and secondly, an* unfitted *finite element discretizations for the imposition of interface conditions is considered. The introduced variants of corresponding Nitsche formulations do not change the sparsity pattern and can easily be implemented into existing finite element codes. The benefit of the new formulations is the removal of the Nitsche stabilization parameter λ while keeping the stability properties of the original formulations for a "sufficiently large" stabilization parameter λ.*

1 Introduction

Nitsche's method was originally introduced in [1] and has been applied in many different contexts in order to weakly enforce interface or boundary conditions. The idea is to impose boundary or interface conditions not into the finite element space as essential condition but to change the variational formulation to achieve the condition at least in a weak sense. For the boundary value problem $-\Delta u = 0$ in Ω with $u|_{\partial\Omega} = 0$ the classical Nitsche variational formulation is: Find $u \in V_h$ such that

$$\int_\Omega \nabla u \nabla v \, dx - \int_{\partial\Omega} \partial_n u \, v \, ds - \int_{\partial\Omega} \partial_n v \, u \, ds + \frac{\lambda}{h} \int_{\partial\Omega} uv \, ds = 0 \quad \forall \, v \in V_h,$$

where V_h is a finite element space with $V_h \subset H^1(\Omega)$ and $\partial_n v_h \in L^2(\partial\Omega)$ for all $v_h \in V_h$. Here h is the characteristic mesh size and λ is a parameter that has to be chosen sufficiently large. The criticism on Nitsche's method is sometimes related to this parameter λ and the "sufficiently large" condition. Although theoretical error analyses can give a precise lower bound on the choice of λ in practice the parameter is typically chosen by estimating the order of magnitude or a trial and error approach. If the parameter is chosen too small, the formulation gets unstable. If however the stabilization parameter is chosen too large the condition number of the system matrix increases and error bounds degenerate. A suitable choice for λ is specifically difficult if problems with a large contrast in the diffusion parameter are considered.

In this paper we focus on *unfitted* finite element discretizations with a weak enforcement of interface or boundary conditions with a *symmetric* Nitsche-type formulation. In unfitted finite element methods a computational mesh is used which is not aligned to the geometries with respect to which a solution to a PDE should be approximated. We consider an elliptic interface problem with an unfitted interface. To transition into this problem we also consider the model problem which has been considered in the original paper [1], a Poisson problem on domain with a fitted boundary.

We give a brief overview on the literature of fitted and especially unfitted discretizations using Nitsche's method. Nitsche's method has a close relation to (stabilized) Lagrange multiplier methods. This connection allows to derive suitable Nitsche formulations from stabilized Lagrange multiplier formulations and has been nicely discussed in [2, 3, 4].

While we want to focus on *symmetric* Nitsche-type formulations we mention the papers [5, 6, 7] which investigate *non-symmetric* Nitsche formulations and get rid of the "sufficiently large" condition on λ, [5], or the whole penalty term, [6, 7]. We aim at achieving the same goal with a *symmetric* formulation.

Many discontinuous Galerkin (DG) schemes – especially the (symmetric) interior penalty method [8] – have a very close relation to Nitsche's method. The continuity of the solution is not build into the finite element space but has to be implemented by the variational formulation. For DG schemes for elliptic problems many variants are known, cf. [9] for an overview of different methods. Among them there exist versions which get rid of the parameter λ and are able to provide consistency, continuity, symmetry and stability. The approach discussed in this paper is in the same virtue as the DG method presented in [10, 11] and analyzed in [12].

For unfitted interface problems a stable Nitsche formulation has been introduced in the original paper [13]. One important aspect of this method is the choice of a geometry-dependend averaging operator which is crucial for stability. We detail on this method below. Many discretization are based on this formulation, for instance [14, 15, 16, 17, 18].

The major contribution of this paper is the introduction of new finite element formulations for the weak imposition of boundary and interface conditions for *fitted* and *unfitted* meshes.

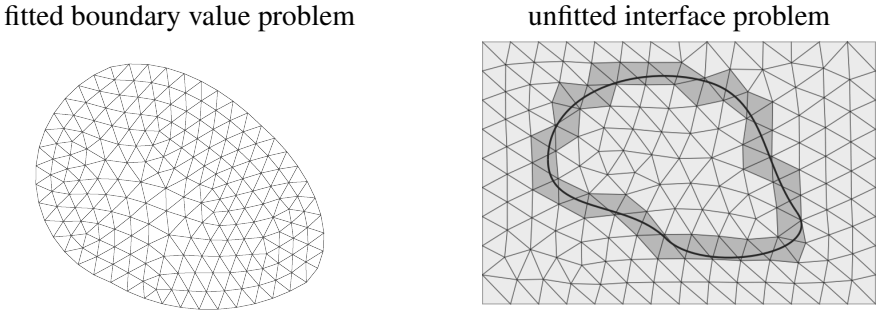


Figure 1: Two situations where Nitsche's method is applied: A fitted Boundary value problems (left) and an unfitted interface problem (right).

These new formulations inherit the stability from known Nitsche-type formulations but remove the stabilization parameter λ.

The paper is organised as follows. In Section 2 we introduce notation for Nitsche-type formulations and their analysis. Based on a common framework we present Nitsche-type formulations from the literature and discuss their theoretical properties in Section 3. Modified formulations which get rid of the penalty parameter λ are then introduced and discussed in Section 4. The modified formulations involve a lifting operator. That an implementation of this lifting operator is practically feasible is shown in Section 5.

2 Preliminaries

2.1 Problem classes

We consider two classes of problems on which Nitsche's method is applied, cf. Fig. 1. A Boundary value problem with a boundary-*fitted* mesh and an elliptic interface problem with an interface-*unfitted* mesh. In Section 3 we introduce suitable Nitsche-type finite element formulations which are known to be stable for both problems. To provide stability different versions of trace inverse inequalities play an important role. The stability of the classical Nitsche method for the model problem can be shown based on standard trace inverse inequalities for polynomials. For the interface problem with an interface-unfitted mesh a special weighting in the Nitsche method is required to obtain a similar result.

Remark 1. *We do not consider the case of an interface problem with a fitted mesh as the same techniques as for the Poisson problem with a boundary-fitted mesh can directly be carried over. Further the case of an unfitted boundary value problem is not consider. Existing stable Nitsche-type formulations for this problem introduce a second stabilization term with a second stabilization parameter, cf. [19]. With the techniques discussed in this paper it is not clear how to remove also this parameter.*

2.2 Unified notation for Nitsche formulations

In both cases we assume to have a domain $\tilde{\Omega}$ on which a shape regular simplex triangulation $\mathcal{T}_h$ is available. Throughout the paper we assume quasi-uniformity with a characteristic mesh size h. This restriction simplifies the presentation but is not essential and can be generalized by exchanging h with a localized version, h_T, a characteristic length of an element T. All scalar product's are L^2 scalar products, i.e. $(u,v)_S = \int_S u\, v\, ds$ for a corresponding domain S.

We can write a suitable version of Nitsche's method for all considered problems in the form:

Find $u \in V_h$ such that

$$A(u,v) := a(u,v) + N^c(u,v) + N^c(v,u) + N^s(u,v) = f(v) \quad \text{for all } v \in V_h, \tag{N}$$

with bilinear and linear forms which are yet to be defined. Here, the bilinear form $a(\cdot,\cdot)$ is responsible to provide consistency with respect to the volume terms, i.e. the inner part of the domain(s). The bilinear form $N^c(\cdot,\cdot)$ is responsible for the consistency w.r.t. boundary terms stemming from partial integration of the elliptic operator. The second term which involves $N^c(\cdot,\cdot)$ is added for symmetry (or adjoint consistency) reasons. Finally $N^s(\cdot,\cdot)$ is a term added to provide stability. In the standard formulation, we have

$$N^s(\cdot,\cdot) = N^s_\lambda(\cdot,\cdot) = \lambda \cdot N^s_1(\cdot,\cdot) \tag{1}$$

where $N^s_\lambda(\cdot,\cdot)$ is the λ-scaled penalty term.

2.3 Analysis of (symmetric) Nitsche formulations

The analysis of corresponding Nitsche's methods uses standard arguments. With coercivity, continuity and consistency of the bilinear form w.r.t. suitable spaces and norms a Céa-like Lemma can be obtained. We briefly introduce the three criteria – in terms of the equations (A)-(D) below – together with the notation for the involved norms and spaces. The criteria are fulfilled for both formulations mentioned below.

We use the notation $a \lesssim b$ ($a \gtrsim b$) if there exists a constant c independent of discretization or material parameters such that $a \leq cb$ ($a \geq cb$). If $a \lesssim b$ and $a \gtrsim b$ we write $a \simeq b$.

Coercivity The crucial component with respect to the stabilization parameter λ is coercivity. For sufficiently large λ all considered Nitsche formulations provide that the following holds:

$$A(u_h,u_h) \gtrsim |||u_h|||_A^2 \text{ for all } u_h \in V_h \text{ with } |||v|||_A := \sqrt{a(v,v) + N^s(v,v)}. \tag{A}$$

In the analysis a stronger norm is needed to bound the flux terms in the bilinear form $N^c(\cdot,\cdot)$. We denote this norm by $|||\cdot|||$ and require that there holds

$$|||v_h||| \simeq |||v_h|||_A \text{ for all } v_h \in V_h, \tag{B}$$

so that (A) yields $A(u_h,u_h) \gtrsim |||u_h|||^2$ for all $u_h \in V_h$.

Continuity On the space $V_h + V_{\text{reg}}$ with V_{reg} so that $A(u,v)$ (especially $N^c(u,v)$) is well-defined for $u,v \in V_{\text{reg}}$ we then have continuity in the sense that

$$A(u,v) \lesssim |||u||| \; |||v||| \quad u,v \in V_h + V_{\text{reg}}. \tag{C}$$

Consistency With $u \in V_{\text{reg}}$ the solution of the original problem and $u_h \in V_h$ the solution to the discrete one, we demand Galerkin orthogonality, i.e.

$$A(u-u_h,v_h) = 0 \text{ for all } v_h \in V_h. \tag{D}$$

With these consistency, continuity and coercivity results we have a version of the well-known Céa Lemma:

Lemma 1. *Let $u \in V_{reg}$ be the solution to the original problem and $u_h \in V_h$ the solution to the discrete one. If the equations* (A)-(D) *are fulfilled, then there holds*

$$|||u - u_h||| \lesssim \inf_{v_h \in V_h} |||u - v_h|||. \tag{2}$$

Below, in Section 4 we propose an alternative to the stabilization form (1) of the form

$$N^s(\cdot,\cdot) = N^s_{nn}(\cdot,\cdot) + N^s_1(\cdot,\cdot)$$

where $N^s_{nn}(\cdot,\cdot)$ is responsible to provide non-negativity in the sense that

$$A(u,u) - N^s_1(u,u) \gtrsim a(u,u)$$

which implies (A) with $\lambda = 1$. Then, $N^s_1(\cdot,\cdot)$ is only required to provide that $A(\cdot,\cdot)$ induces a norm. The scaling of this term is however independent of any stability requirement. In the following section we specify the problem classes and the bilinear forms, spaces and norms corresponding to suitable discretization which are known to be stable (for sufficiently large λ).

3 Stable Nitsche formulations

In the following subsections we will introduce suitable discretizations for the considered problems and show the conditions (A)-(D).

3.1 Nitsche formulation for a boundary value problem with a body-fitted mesh

In the fitted case we have $\tilde{\Omega} = \Omega$ and the model problem is

$$-\Delta u = f \text{ in } \Omega,\ u = g \text{ on } \partial\Omega \tag{3}$$

In this case the Nitsche's method, as introduced in [1], takes the form of (N) with

$$a(u,v) = (\nabla u, \nabla v)_\Omega,\ N^c(u,v) = (-\partial_n u, v)_{\partial\Omega} \text{ and } N^s(u,v) = N^s_\lambda(u,v) = \frac{\lambda}{h}(v,v)_{\partial\Omega}.$$

We briefly verify (A)-(D) with respect to suitable spaces and norms. As the finite element space V_h we consider a standard finite element space of continuous piecewise polynomials up to degree k. The stronger norm on $V_h + V_{\text{reg}}$ with $V_{\text{reg}} = H^2(\Omega)$ is

$$|||u|||^2 = |||u|||^2_A + h\|\nabla u\|^2_{\partial\Omega},$$

and continuity on $V_h + V_{\text{reg}}$ with respect to $|||\cdot|||$ follows directly with Cauchy-Schwarz, so that we have (C). Consistency follows directly from partial integration, so that also (D) holds. An essential estimate to prove (A) is the following trace inverse inequality:

$$h\|\nabla u\|^2_{\partial\Omega\cap T} \le h\|\nabla u\|^2_{\partial T} \le c_{\text{tr}}\|\nabla u\|^2_T, \quad u \in \mathcal{P}^k(T),\ T \in \mathcal{T}_h$$

with a constant c_{tr} only depending on the shape regularity of the element T and the polynomial degree k, cf. [20]. Assuming that c_{tr} denotes the upper bound of the shape regularity constant on all elements in the mesh, then we can easily show (B) and the estimate

$$2N^c(u_h,u_h) \le 2h^{\frac{1}{2}}\|\nabla u_h\|_{\partial\Omega} \cdot h^{-\frac{1}{2}}\|u_h\|_{\partial\Omega} \le \frac{8c_{\text{tr}}}{\lambda}a(u_h,u_h) + \frac{1}{2}N^s_\lambda(u_h,u_h) \tag{4}$$

from which we deduce (A) for λ sufficiently large (for instance $\lambda > 16c_{\text{tr}}$).

3.2 Nitsche formulation for an interface problem with an unfitted mesh

For the case of an unfitted interface we assume $\tilde{\Omega} = \Omega$ but have an additional internal interface $\Gamma \subset \Omega$ which separates Ω into two disjoint sub-domains Ω_1 and Ω_2 which are unfitted, i.e. not aligned with the mesh. Inside the domains only diffusion takes place, the diffusion coefficient α may however be discontinuous across the interface Γ. On the boundary we consider homogeneuous Dirichlet conditions to be implemented as essential conditions. This simplifies the presentation. For a weak imposition of the boundary conditions, we refer to the treatment of the case in section 3.1. At the interface two conditions are posed: continuity of the solution and conservation of the flux.

$$-\alpha\Delta u = f \text{ in } \Omega_i, \quad u = 0 \text{ on } \partial\Omega, \quad [\![u]\!] = 0 \text{ on } \Gamma, \quad [\![-\alpha\partial_n u]\!] = 0 \text{ on } \Gamma. \tag{5}$$

In unfitted FEM a famous formulation of the problem goes back to the original paper [13]. The finite element space depends on Γ and is defined as $V_h = V_h^1 + V_h^2$ where $V_h^i = W_h|_{\Omega_i}$ is the fictitious domain finite element space related to domain Ω_i with the underlying standard finite element space W_h which again is the space of continuous piecewise polynomial functions. The variational formulation proposed in [13] can be casted into the form of (N) with

$$a(u,v) = (\alpha\nabla u, \nabla v)_{\Omega_1\cup\Omega_2},\ N^c(u,v) = (\{\!\{-\alpha\partial_n u\}\!\}, [\![v]\!])_\Gamma \text{ and } N^s(u,v) = \frac{\lambda}{h}([\![u]\!], [\![v]\!])_\Gamma.$$

where $[\![\cdot]\!]$ is the usual jump operator across the interface Γ. The averaging operator $\{\!\{v\}\!\} = \kappa_1 v|_{\Omega_1} + \kappa_2 v|_{\Omega_2}$ takes an important role to provide stability. With $V_{\text{reg}} := H^2(\Omega_1) \cup H^2(\Omega_2)$ and the norm $|||u|||^2 = |||u|||_A^2 + h\|\{\!\{\nabla u\}\!\}\|_\Gamma^2$ one easily shows (C) and (D). It remains to show (A) and (B). In the following we assume that a stable weighting is used, so that the following inverse estimate holds:

$$\alpha_i^2\kappa_i^2 h\|\nabla u\|_{\Gamma\cap T}^2 \le c_{\text{tr}}\alpha_i\|\nabla u\|_{T_i}^2, \quad u \in \mathcal{P}^k(T_i),\ T_i := T \cap \Omega_i,\ i = 1, 2, .$$

for a constant c_{tr} only depending on the shape regularity of T (not T_i!), the polynomial degree k and the diffusion parameter α_i. We note that a carefully chosen stabilization parameter λ in $N^s(u,v) = N_\lambda^s(u,v)$ scales with c_{tr} and hence with the diffusion parameter α.

For piecewise linears ($k = 1$) and the weighting $\kappa_i = |T_i|/|T|$ the inverse inequality has been proven in [13]. For higher order discretizations we refer to the discussion of weightings that lead to stable discretizations in [21, Remark 6]. Analogously to the estimates in (4) this allows to prove (A) and (B). Together with the optimal approximation properties of V_h, see also [13], we can obtain optimal order error bounds provided that λ is chosen sufficiently large.

4 Modified formulations to remove the parameter λ

In the following we introduce a new stabilization bilinear form based on a lifting operator and reestablish (A)-(D) with this new parameter-free stabilization term and $\lambda = 1$ in the penalty term. For the formulations in the Sections 3.1-3.2 the bilinear forms $a(\cdot,\cdot)$, $N^c(\cdot,\cdot)$ and $N^s(\cdot,\cdot)$ can easily be decomposed into element contributions which are indicated by an index T, e.g. $a_T(\cdot,\cdot)$, so that $a(\cdot,\cdot) = \sum_{T\in\mathcal{T}_h} a_T(\cdot,\cdot)$. We introduce and analyse the modified formulation for the fitted Poisson case and afterwards translate it to the unfitted interface problem.

4.1 The modified formulation for the fitted Poisson problem

The discrete local lifting operator $\mathcal{L}(\cdot)$

On elements which are located at the boundary we introduce an element-wise lifting into the space of polynomials which are orthogonal to constants, $\mathcal{P}_0^k(T) := \mathcal{P}^k(T) \setminus \mathcal{P}^0(T)$. The lifting

is defined element-wise as

$$\mathcal{L}_T : H^1(T) \to \mathcal{P}_0^k(T), \qquad a_T(\mathcal{L}_T(u), v) = N_T^c(v, u) \quad \forall v \in \mathcal{P}_0^k(T) \tag{6}$$

We note that $\mathcal{L}_T$ is an element-local operator and that (6) also holds for $v \in \mathcal{P}^k(T)$ as $a_T(\cdot, w) = N_T^c(w, \cdot) = 0$ for $w \in \mathcal{P}^0(T)$. Further, by construction there holds $\mathcal{L}_T(u) \perp \mathcal{P}^0(T)$ and $\mathcal{L}_T(u) = 0$ if $u|_{\partial\Omega \cap T} = 0$. We redefine $a(u, v) := \sum_{T \in \mathcal{T}_h} a_T(u, v)$ with $a_T(u, v) = (\nabla u, \nabla v)_T$, i.e. in a broken sense so that $a(\cdot, \cdot)$ is also well defined for only element-wise $H^1(T)$ functions. On elements which are not located at the boundary we set $\mathcal{L}_T(\cdot) = 0$ and define the global lifting

$$\mathcal{L} : H^1(\mathcal{T}_h) \to \bigoplus_{T \in \mathcal{T}_h} \mathcal{P}_0^k(T) \quad \text{with} \quad \mathcal{L}(u)|_T := \mathcal{L}_T(u).$$

Note that in general $\mathcal{L}(u)$ is discontinuous across element interfaces. With the definition of the lifting we have $a(\mathcal{L}(u), v) = N^c(v, u)$ for all $v \in V_h^{\text{disc}} := \bigoplus_{T \in \mathcal{T}_h} \mathcal{P}^k(T)$ and hence especially for $v \in V_h \subset V_h^{\text{disc}}$.

Analysis of the modified Nitsche formulation for fitted boundary value problems

The lifting allows to characterize the nonsymmetric term $N^c(\cdot, \cdot)$ in terms of the symmetric bilinear form $a(\cdot, \cdot)$ and the lifting operator applied on one of the arguments. This characterization will help to control $N^c(\cdot, \cdot)$ in terms of $a(\cdot, \cdot)$ and a new symmetric bilinear form involving the lifting. We exploit this in the following central result.

Lemma 2. *The bilinear form*

$$A^{nn}(u, v) := a(u, v) + N^c(u, v) + N^c(v, u) + \underbrace{2a(\mathcal{L}(u), \mathcal{L}(v))}_{=:N_{nn}^s(u,v)}, \quad u, v \in V_{reg} = H^2(\Omega) \tag{7}$$

is non-negative in V_h and there holds $A^{nn}(u_h, u_h) \geq \frac{1}{2} a(u_h, u_h)$ for all $u_h \in V_h$.

Proof. Consider $u_h \in V_h$. Then with Young's inqualitywe have

$$2N^c(u_h, u_h) = 2a(\mathcal{L}(u_h), u_h) \leq 2a(\mathcal{L}(u_h), \mathcal{L}(u_h)) + \frac{1}{2} a(u_h, u_h) = N_{nn}^s(u_h, u_h) + \frac{1}{2} a(u_h, u_h)$$

from which we can deduce the claim. □

Note that $N_{nn}^s(u, v)$ is well defined on $H^1(\Omega)$. The bilinear form $A^{nn}(u, v)$ only offers control on the semi-norm $\sqrt{a(u, u)}$. In order to have control of a reasonable norm, i.e. in order to have control on the boundary values, we add the bilinear form $N_1^s(\cdot, \cdot)$. Note however that we can choose $\lambda = 1$ (or any other positive number) as the only purpose of this part is to add control on the boundary integral. Due to the lifting part in $a(\mathcal{L}(\cdot), \mathcal{L}(\cdot))$ the bilinear form $N_\lambda^s(\cdot, \cdot)$ is no longer needed to control the non-symmetric terms $N^c(\cdot, \cdot)$. With

$$A^*(u, v) := A(u, v) \text{ as in (N) with } N^s(\cdot, \cdot) = 2a(\mathcal{L}(\cdot), \mathcal{L}(\cdot)) + N_1^s(\cdot, \cdot)$$

we can directly conclude the following corollary from which (A) follows.

Corollary 3. *The bilinear form $A^*(\cdot, \cdot)$ is well-defined on $V_{reg} + V_h$ and coercive on V_h w.r.t. the norm $|||v|||_A$ with coercivity constant $\frac{1}{2}$.*

Consistency in the form of (D) immediately follows from the fact that $\mathcal{L}(u) = 0$ for $u|_{\partial\Omega} = 0$. One might get the impression that the trace inverse inequality constant c_{tr} does not play a role in the formulation with $A^*(\cdot,\cdot)$. That this is not true becomes obvious if we investigate the continuity of the bilinear form $a(\mathcal{L}(\cdot),\mathcal{L}(\cdot))$. We have

$$\begin{aligned} a_T(\mathcal{L}(u),\mathcal{L}(u)) = N_T^c(\mathcal{L}(u),u) \leq \|\nabla\mathcal{L}(u)\|_{\Gamma\cap T}\|u\|_{\Gamma\cap T} \leq c_{\text{tr}}^{\frac{1}{2}} h^{-\frac{1}{2}} \|\nabla\mathcal{L}(u)\|_T \|u\|_{\Gamma\cap T} \\ \leq \frac{1}{2} a_T(\mathcal{L}(u),\mathcal{L}(u)) + \frac{1}{2}\frac{c_{\text{tr}}}{h}\|u\|^2_{\Gamma\cap T} \quad \Longrightarrow \quad a(\mathcal{L}(u),\mathcal{L}(u)) \leq \frac{c_{\text{tr}}}{h}\|u\|^2_\Gamma \;\; \forall\, u \in H^1(\Omega) \end{aligned} \tag{8}$$

and hence we have (B) and (C) with constants depending on c_{tr}. We see that the constant from the trace inverse estimate is still very important and enters the error analysis. However, we do not need to estimate the constant c_{tr} explicitly and hence get rid of the penalty parameter λ and the "sufficiently large" condition.

4.2 The modified formulation for the unfitted interface problem

The idea from the previous section can be carried over to unfitted interface problems. For this case the idea has already been introduced in [22, Section 2.2.3.2]. We define a lifting operator similar to (6), but use the space of piecewise polynomials which are orthogonal to constants on each sub-element, $\mathcal{P}_0^k(T_{1,2}) := \{v \in L^2(T), v|_{T_1} \in \mathcal{P}_0^k(T_i),\ i = 1,2\}$. On cut elements we set

$$\mathcal{L}_T : H^1(T_1 \cup T_2) \to \mathcal{P}_0^k(T_{1,2}),\ \text{s.t.} \quad a_T(\mathcal{L}_T(u),v) = N_T^c(v,u) \quad \forall v \in \mathcal{P}_0^k(T_{1,2}), \tag{9}$$

on uncut elements we set $\mathcal{L}_T(\cdot) = 0$. Overall we again have $a(\mathcal{L}(u),v) = N^c(u,v)\ \ \forall v \in V_h$. With this property Lemma 2, Corollary 3 and (8) can directly be carried over and we have (A), (B) and (C). For (D) we note that $\mathcal{L}(u) = 0$ if $[\![u]\!] = 0$.

5 Implementational aspects

In this section we want to explain that the implementation of the bilinear form $2a(\mathcal{L}(\cdot),\mathcal{L}(\cdot))$ is fairly simple. We explain the procedure for the case of the fitted Poisson problem. It however easily translates to the unfitted interface problems (or similar problems). To implement the element-local lifting $w_h = \mathcal{L}_T(u_h)$ of a local finite element function u_h and an element T that is located at the boundary, we solve for w_h with $u_h, w_h \in \mathcal{P}^k(T)$, such that

$$a_T(w_h,v_h) + k_T(w_h,v_h) = N_T^c(v_h,u_h), \quad \forall v_h \in \mathcal{P}^k(T) \tag{10}$$

with the bilinear form

$$k_T(w_h,v_h) := h^{-(d+2)}\,(w_h,1)_T\,(v_h,1)_T$$

which is taylored to eliminate the kernel $\{u_h|_T = const\}$. By testing with a constant function one easily sees that if w_h solves (10), we have $k_T(w_h,v_h) = 0$ for all $v_h \in \mathcal{P}_0^k(T)$. Instead of adding $k_T(\cdot,\cdot)$ any other approach which makes (10) uniquely solvable and ensures $a_T(w_h,v_h) = N_T^c(v_h,u_h)$ for all $v_h \in \mathcal{P}^k(T)$ is possible. Note that the local element matrices $\mathbf{A}$ and $\mathbf{N}_c$ corresponding to the bilinear form $a_T(\cdot,\cdot)$ and $N_T^c(\cdot,\cdot)$ have to be computed anyway. The only new component, the rank 1 element matrix $\mathbf{K}$ corresponding to $k_T(\cdot,\cdot)$ is easily obtained by scaling of a corresponding element matrix with respect to a reference element. We thus get the coefficients $\mathbf{w}$ of the local lifting ($\mathbf{w}_i = \mathcal{L}(\varphi_i)$) as $\mathbf{w} = \mathbf{L}\cdot\mathbf{u}$ with $\mathbf{L} = (\mathbf{A}+\mathbf{K})^{-1}\mathbf{N}_c^T$. The overall element contribution to the bilinear form $A^*(\cdot,\cdot)$ in matrix notation is

$$\mathbf{A} + \mathbf{N}_c + \mathbf{N}_c^T + 2\cdot\mathbf{L}^T\mathbf{A}\mathbf{L} + \mathbf{N}_{s,1}$$

where $\mathbf{N}_{s,1}$ is the element matrix corresponding to $N_1^s(\cdot,\cdot)$. The element matrix corresponding to the original formulation, cf. Section 3.1, takes the form $\mathbf{A} + \mathbf{N}_c + \mathbf{N}_c^T + \lambda \mathbf{N}_{s,1}$. We note that the only additional effort of the new formulation is in the computation of the local matrices $\mathbf{K}$, $\mathbf{L}$ and $2 \cdot \mathbf{L}^T \mathbf{A} \mathbf{L}$.

6 Numerical example

We consider a simple numerical example to compare the original and the modified method. The impact of a penalty parameter that has been chosen too large is small for the fitted case as diagonal preconditioning gives robustness of the linear system with respect to λ. As this is not true for the unfitted interface problem, cf. [18], we consider the discretization of an unfitted interface problem and take an example from [21, Section 4.1]. On the domain $\Omega = [2.01, 2.01]^2$ we consider the interface $\Gamma = \{\|x\|_4 = 1\}$ which is approximated with a piecewise planar reconstruction Γ_h. We choose the diffusion coefficients $(\alpha_1, \alpha_2) = (1, 2)$ so that the solution has a kink across Γ. For the approximation of the solution we consider the finite element formulation as in Section 3.2 with $k = 1$, i.e. piecewise linear functions and the modified formulation as in Section 4.2. The right hand side and the Dirichlet data are chosen such that the solution is given as

$$u(x) = \begin{cases} 1 + \frac{\pi}{2} - \sqrt{2} \cdot \cos(\frac{\pi}{4}\|x\|_4^4), & x \in \Omega_1, \\ \frac{\pi}{2}\|x\|_4 \ , & x \in \Omega_2. \end{cases}$$

We consider a uniform triangular grid obtained by dividing a rectangular 16×16 mesh into triangles which results in 512 unknowns. Note that this implies a very shape regular mesh. We computed the spectral condition number κ of the diagonally preconditioned system matrix for different values of λ. The results are shown in Table 1. We observe that the standard

λ	1	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192
κ	-	-	-	-	86.3	81.6	79.2	83.2	88.0	91.3	116.0	221.4	427.4	830.3

Table 1: Dependency of the conditioning of the Nitsche formulation on the penalty parameter λ.

discretization with $N^s(\cdot,\cdot) = N_\lambda^s(\cdot,\cdot)$ is only stable for values $\lambda > 8$. Within a range of one order of magnitude the condition number seems to be essentially independent of λ. However, for $\lambda \geq 256$ the condition number increases linearly with λ. For the modified (parameter-free) formulation with $N^s(\cdot,\cdot) = N^{nn}(\cdot,\cdot) + N_1^s(\cdot,\cdot)$ we obtain a condition number of 86.9, i.e. a condition number which is close to an optimal choice of λ.

Independent of λ (for $\lambda \geq 16$) we further observed optimal order of convergence for the $L^2(\Omega)$ and the $H^1(\Omega_1 \cup \Omega_2)$ norm of the error. This holds also true for the modified formulation. Further, we observed that the error seems to be essentially independent of the choice of λ. This does not hold for the $L^2(\Gamma)$ norm of the jump which decreases linearly with increasing λ.

7 Conclusion

We presented an approach to eliminate the penalty parameter from symmetric Nitsche formulations. Modified Nitsche formulations have been proposed which are based on well-known formulations from the literature. The new formulations inherit the good properties while the stability of the formulations no longer depends on a stabilization parameter that has to be chosen sufficiently large. We explained the procedure for two important applications of Nitsche's method and demonstrated its potential on a simple numerical example.

REFERENCES

[1] J. Nitsche, "Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 36, no. 1, pp. 9–15, 1971.

[2] R. Verfürth, "Finite element approximation of incompressible Navier-Stokes equations with slip boundary conditions II," *Numer. Math.*, vol. 59, no. 1, pp. 615–636, 1991.

[3] R. Stenberg, "On some techniques for approximating boundary conditions in the finite element method," *J. Comp. Appl. Math.*, vol. 63, pp. 139–148, 1995. relation between stabilized lagrange multiplier techniques and Nitsche.

[4] M. Juntunen, "On the connection between the stabilized Lagrange multiplier and Nitsche's methods," *Numerische Mathematik*, vol. 131, no. 3, pp. 453–471, 2015.

[5] J. Freund and R. Stenberg, "On weakly imposed boundary conditions for second order problems," in *Proceedings of the Ninth International Conference on Finite Elements in Fluids, M. Cecchi et al., eds.,*, pp. 327–336, Universita di Padova, 1995.

[6] E. Burman, "A penalty-free nonsymmetric Nitsche-type method for the weak imposition of boundary conditions," *SIAM J. Numer. Anal.*, vol. 50, no. 4, pp. 1959–1981, 2012.

[7] T. Boiveau and E. Burman, "A penalty-free Nitsche method for the weak imposition of boundary conditions in compressible and incompressible elasticity," *IMA Journal of Numerical Analysis*, 2015.

[8] D. N. Arnold, "An interior penalty finite element method with discontinuous elements," *SIAM Journal on Numerical Analysis*, vol. 19, no. 4, pp. 742–760, 1982.

[9] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, "Unified analysis of discontinuous Galerkin methods for elliptic problems," *SIAM J. Numer. Anal.*, vol. 39, no. 5, pp. 1749–1779, 2002.

[10] F. Bassi and S. Rebay, "A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations," *J. Comput. Phys.*, vol. 131, pp. 267–279, March 1997.

[11] F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti, and M. Savini, "A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows," in *Proceedings of 2nd European Conference on Turbomachinery, Fluid Dynamics and Thermodynamics*, pp. 99–108, Technologisch Instituut, Antwerpen, Belgium, 1997.

[12] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo, "Discontinuous finite elements for diffusion problems," *Atti Convegno in onore di F. Brioschi (Milano 1997), Istituto Lombardo, Accademia di Scienze e Lettere*, pp. 197–217, 1999.

[13] A. Hansbo and P. Hansbo, "An unfitted finite element method, based on nitsches method, for elliptic interface problems," *Computer methods in applied mechanics and engineering*, vol. 191, no. 47, pp. 5537–5552, 2002.

[14] P. Hansbo, "Nitsche's method for interface problems in computational mechanics," *GAMM-Mitt.*, vol. 47, no. 2, pp. 183–206, 2005.

[15] C. Lehrenfeld and A. Reusken, "Nitsche-XFEM with Streamline Diffusion stabilization for a two-phase mass transport problem," *SIAM J. Sci. Comput.*, vol. 34, pp. 2740–2759, 2012.

[16] C. Annavarapu, M. Hautefeuille, and J. E. Dolbow, "A robust Nitsche's formulation for interface problems," *Comput. Meth. Appl. Mech. Eng.*, vol. 225–228, no. 0, pp. 44–54, 2012.

[17] P. Hansbo, M. G. Larson, and S. Zahedi, "A cut finite element method for a Stokes interface problem," *Appl. Num. Math.*, vol. 85, pp. 90–114, 2014.

[18] C. Lehrenfeld and A. Reusken, "Optimal preconditioners for Nitsche-XFEM discretizations of interface problems," *arXiv preprint arXiv:1408.2940*, 2014.

[19] E. Burman and P. Hansbo, "Fictitious domain finite element methods using cut elements: II. a stabilized Nitsche method," *Applied Numerical Mathematics*, vol. 62, no. 4, pp. 328–341, 2012.

[20] T. Warburton and J. Hesthaven, "On the constants in hp-finite element trace inverse inequalities," *Comput. Methods Appl. Mech. Engrg.*, vol. 192, pp. 2765–2773, 2003.

[21] C. Lehrenfeld and A. Reusken, "Analysis of a high order unfitted finite element method for elliptic interface problems," *arXiv preprint arXiv:1602.02970*, 2016.

[22] C. Lehrenfeld, *On a Space-Time Extended Finite Element Method for the Solution of a Class of Two-Phase Mass Transport Problems*. PhD thesis, RWTH Aachen, February 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A SEMI-UNSTRUCTURED TURBOMACHINERY MESHING LIBRARY WITH FOCUS ON MODELING OF SPECIFIC GEOMETRICAL FEATURES

Marco Stelldinger[1], Thomas Giersch[2], Felix Figaschewsky[1], Arnold Kühhorn[1]

[1] Chair of Structural Mechanics and Vehicle Vibration Technology
Brandenburg University of Technology Cottbus - Senftenberg
D-03046 Cottbus, Germany
e-mail: marco.stelldinger@b-tu.de

[2] Rolls-Royce Deutschland Ltd & Co KG
D-15827 Blankenfelde-Mahlow, Germany
e-mail: thomas.giersch@rolls-royce.com

Keywords: Semi-unstructured mesh generation, Mesh quality, Turbomachinery, CFD

Abstract. *Computational Fluid Dynamics is widely used for the analysis and the design of turbomachinery blade rows. A well established method is the application of semi-unstructured meshes, that uses a combination of structured meshes in the radial direction and unstructured meshes in the axial as well as the tangential direction. This takes advantage of the approximately two dimensional flow field through the blade rows, whereby a fine radial discretization, excepting the near wall region, is not necessary. Otherwise, it is possible to discretize particular regions, e.g. the leading and trailing edge regions, in the axial and tangential direction without generating unnecessary nodes in the far field. The meshing approach is based on the projection of a two dimensional unstructured mesh defined at a reference surface. Once, the two dimensional mesh is generated the projection is achieved by transfinite interpolation from the reference surface to further radial surfaces using a structured mesh. Due to the modeling of geometrical features, especially fillets, advanced methods for the generation of structured meshes and mesh smoothing algorithms are required.*

The paper presents two different approaches for the generation of an appropriate structured mesh. The first is based on the solution of elliptic partial differential equations. The second approach is based on the split of the domain into fourteen appropriately arranged blocks. Furthermore, two smoothing methods for two dimensional unstructured meshes, a constrained Laplace smoothing and an optimization based approach, are presented. Regarding a more realistic representation of the geometry, methods for the modeling of cavities, variable clearance sizes and fillets are presented. Finally, a comparison of the smoothing techniques applied to a rotor passage is presented and the influence of chosen geometrical features on the flow solution is evaluated.

1 INTRODUCTION

Computational Fluid Dynamics (CFD) is a widely used method for the analysis and the design of aircraft engines. Nowadays, the design process of an aircraft engine based nearly to 90% on CFD simulations [1]. An integral part of CFD simulations is the generation of an appropriate finite volume mesh. This paper is concerned with the mesh generation for the analysis of flows in turbomachinery. Due to the fact, that the computational resources are continuously growing, it is possible to use meshes with a fine grid spacing in CFD simulations. Otherwise, large models are required for specific CFD simulations. These are for instance unsteady forced response analyses, that require the modeling of several blade rows as a full passage, and steady state analyses of an entire high pressure compressor. Hence, it is necessary to generate meshes, that can capture the flow features with good accuracy employing a moderate number of nodes. Another challenge is to generate a mesh, that represents the real geometry as accurately as possible. Thereby, the modeling of specific geometrical features is required. In addition, for multi blade row computations an accurate interpolation of the flow field at the interfaces is required. It is common to generate a finite volume mesh only for a single blade, so that the imposition of periodicity in circumferential direction is mandatory.

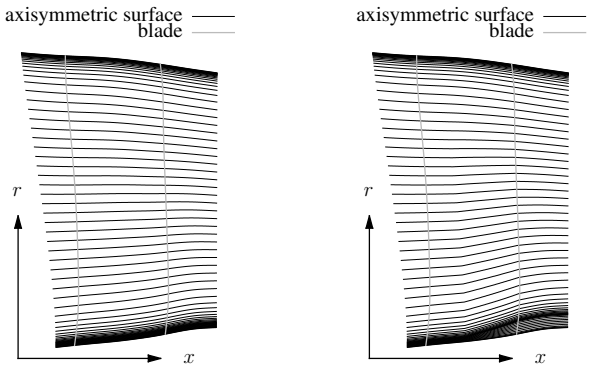
Historically, mainly structured hexahedral meshes are used for the discretization of blade passages, since the generation of such meshes is comparatively easy. Structured meshes are obtained either by solving a system of elliptic partial differential equations [2, 3, 4, 5, 6] or by using algebraic approaches [6, 7, 8]. A major disadvantage of such meshes is the generation of unnecessary nodes in the far field, because of required local fine discretization, e.g. at the leading edge and trailing edge of the blade. Another approach is the application of unstructured triangular meshes, primarily used for two dimensional CFD simulations. Such meshes capture flow features of interest, like wake and shock effects, via mesh refinement, without generating unnecessary nodes in the far field. In the three dimensional case the fine resolution at leading edge and trailing edge leads to a large number of nodes in the radial direction. Considering the fact that the flow field in the blade passage is approximately two dimensional a fine resolution in the radial direction is not necessary. Taking into account these facts an alternative and well established method is the application of semi-unstructured meshes, e.g. proposed by Sbardella et al. [9] as well as Kim and Cizmas [10]. This method uses a combination of structured meshes in the radial direction and unstructured meshes in the axial as well as the tangential direction. Furthermore, it is possible to start with a structured O-grid around the blade surface to resolve the boundary layer. Such meshes compose of prismatic and tetrahedral cells and are widely used for viscous CFD simulations, because of a good control of the cell size in the viscous region [11]. They can capture the flow features with good accuracy applying a moderate number of nodes. Another advantage is that only two dimensional meshing algorithms are required. An overview of this strategy is presented in section 2. One part of the generation of semi-unstructured meshes is the projection of a two dimensional unstructured mesh, defined at a two dimensional reference surface, to further surfaces using a two dimensional structured mesh, also denoted as background mesh [10, 12]. In section 3 two approaches for the generation of an appropriate background mesh are presented. The first approach is based on the solution of a system of elliptic partial differential equations. The second approach is based on the split of the domain into fourteen blocks.

In the last years blade geometries became more three dimensional. Thereby, the projected unstructured mesh can be heavily skewed. The modeling of fillets amplifies this effect because the domain of the unstructured mesh constricts near the fillet. This leads to the requirement of

suitable methods to improve the quality of the two dimensional unstructured triangular mesh. In this paper a combination of edge swapping and smoothing is used. In section 4 two smoothing algorithms, a constrained Laplace smoothing and an optimization-based smoothing, are presented. Section 5 presents the modeling of some specific geometrical features, those are cavities, fillets and variable clearance sizes.

2 SEMI-UNSTRUCTURED GRID GENERATION

The geometry of a turbomachinery blade is usually defined as a number of radial sections, defined in cylindrical coordinates by (r, ϑ, x). The radial distribution of these sections must not be confused with the radial distribution of the three dimensional finite volume mesh. Additionally, the position of the inlet, outlet and inner as well as outer casing is usually represented by streamlines, defined in the (x, r) plane. At the beginning, the radial distribution of the mesh has to be defined. E.g. in case of viscous flow simulations a fine radial discretization near the boundaries is necessary. In this paper the definition of the axisymmetric surfaces, that define the radial distribution of the three dimensional mesh, is variable along the machine axis. These fact allows the modeling of clearances of a varying size. Figure 1 shows the radial distribution of a variable stator vane with a constant (a) and a variable clearance size (b).



(a) Constant clearance size (b) Variable clearance size

Figure 1: Axisymmetric surfaces, those define the radial distribution of the three dimensional mesh

To apply the implemented mapping procedure the surfaces are defined by two dimensional coordinates (u, v). Applying the two dimensional discrete geometry information the blade surface and the outer boundaries can be defined by linear or cubic splines, see figure 2. The outer boundaries are the inlet, outlet and periodic boundaries. An O-grid is generated around the blade at each spanwise surface. Afterwards an unstructured triangular mesh, that fills the area between the outer boundaries and the boundary of the O-grid, is generated at a chosen reference surface. In this paper the unstructured triangular mesh is generated by applying an advancing front algorithm, that is a widely used method for the generation of unstructured meshes for CFD simulations [9, 13, 14, 15, 16, 17]. The O-grid and the unstructured mesh at the reference surface of a rotor passage are shown in figure 3,

An essential part of the meshing strategy is the projection of the generated unstructured mesh to all spanwise surfaces. In this paper the Transfinite Interpolation (TFI), explained in

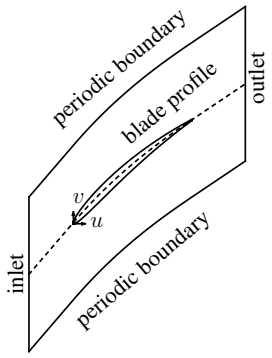
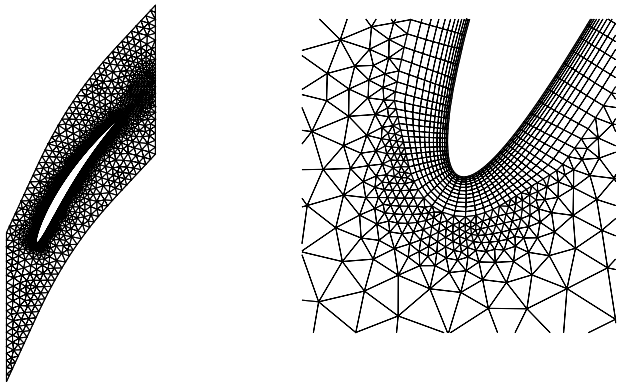


Figure 2: Blade surface and outer boundaries defined in two dimensional coordinates (u, v)



(a) Entire domain (b) Detail of the leading edge

Figure 3: Initial two dimensional mesh of a rotor at the reference surface

section 3, in combination with a background mesh is used. The starting point of the mapping procedure is the generation of appropriate background meshes with identical topology at all spanwise surfaces. Each quadrilateral of the background mesh (figures 6, 7) is represented as an uniform rectangular domain with curvilinear coordinates (ξ, η) applying TFI. In the next step each node i of the unstructured mesh must be located on a quadrilateral j of the corresponding background mesh. Afterwards the coordinates of the node are represented in local curvilinear coordinates (ξ, η) of the corresponding quadrilateral. These coordinates are determined by a Newton-Raphson method, where this process is referred to as inverse mapping. Now the two dimensional coordinates (u, v) of each point i of the unstructured mesh can be obtained at all spanwise surfaces using TFI which is called direct mapping, see figure 4. As a last step the three dimensional grid is obtained by connecting the corresponding points of adjacent spanwise surfaces.

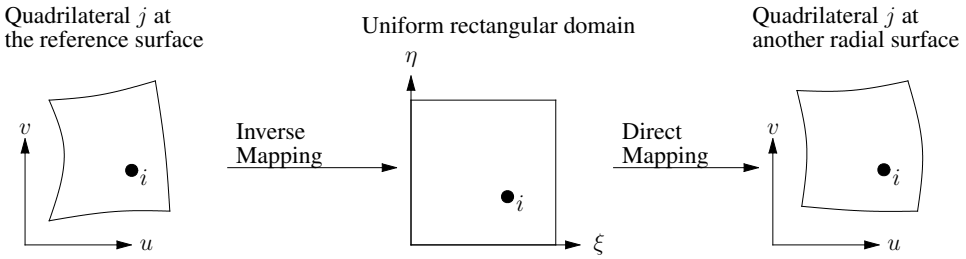


Figure 4: Mapping procedure via Transfinite Interpolation

3 GENERATION OF A BACKGROUND MESH

3.1 Transfinite Interpolation

TFI is a widely used algebraic approach to generate structured meshes and was first described by Gordon [18]. TFI has the advantage to generate grids, these ensure conformity to the boundaries. Thompson et al. [6] give a detailed overview of different interpolation methods used by TFI. In this paper the TFI is not used for mesh generation but entirely for the mapping of quadrilaterals from the physical domain with cartesian coordinates (x, y) into the logical domain with

curvilinear coordinates (ξ, η), as shown in figure 5. The expression for a TFI applying linear interpolation functions, these are used in this paper, is:

$$\begin{aligned}\underline{x}(\xi,\eta) =& \big(1-\eta\big)\underline{x}_s(\xi) + \eta\underline{x}_n(\xi) + \big(1-\xi\big)\underline{x}_w(\eta) + \xi\underline{x}_e(\eta) \\ &- \xi\Big[\big(1-\eta\big)\underline{x}_s(1) + \eta\underline{x}_n(1)\Big] - \big(1-\xi\big)\Big[\big(1-\eta\big)\underline{x}_s(0) + \eta\underline{x}_n(0)\Big] \quad .\end{aligned} \tag{1}$$

Here $\underline{x}_s(\xi)$, $\underline{x}_n(\xi)$, $\underline{x}_w(\eta)$ and $\underline{x}_e(\eta)$ are the boundary curves of the quadrilateral in cartesian coordinates, defined by linear or cubic splines.

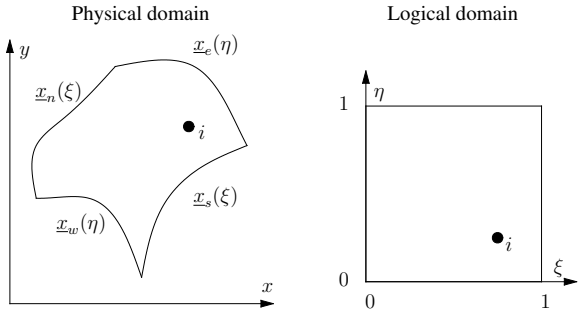


Figure 5: Transfinite Interpolation

3.2 System of elliptic partial differential equations

Solving a system of elliptic partial differential equations is a widely used method to generate structured grids. To obtain an accurate solution, the physical domain is transformed into a uniform rectangular domain by employing a boundary-conforming coordinate transformation. A detailed description is given by Thompson, Thames and Mastin [2, 3, 19]. Consider a system of elliptic partial differential equations in the physical domain (x, y) formulated as a system of Poisson equations:

$$\begin{aligned}\xi_{xx} + \xi_{yy} &= P(\xi,\eta) \\ \eta_{xx} + \eta_{yy} &= Q(\xi,\eta) \quad .\end{aligned} \tag{2}$$

Thereby the curvilinear coordinates (ξ, η) are the solution of the system in the physical domain. The non-homogeneous terms P and Q are the control functions. It is desired to solve the system in the uniform rectangular domain. To formulate the equations (2) into the transformed domain the dependent and independent variables must be interchanged [3]. The equations (2) become

$$\begin{aligned}\alpha x_{\xi\xi} - 2\beta x_{\xi\eta} + \gamma x_{\eta\eta} &= -J^2(\alpha P x_\xi + \gamma Q x_\eta) \\ \alpha y_{\xi\xi} - 2\beta y_{\xi\eta} + \gamma y_{\eta\eta} &= -J^2(\alpha P y_\xi + \gamma Q y_\eta)\end{aligned} \tag{3}$$

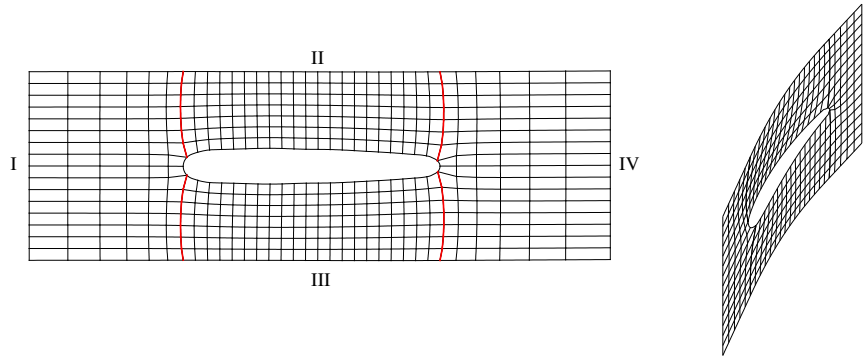
where

$$\begin{aligned}\alpha &= x_\eta^2 + y_\eta^2, \quad \beta = x_\xi x_\eta + y_\xi y_\eta \\ \gamma &= x_\xi^2 + y_\xi^2, \quad J = x_\xi y_\eta - x_\eta y_\xi \quad .\end{aligned} \tag{4}$$

The control functions P and Q affect the intersecting angles at the boundaries and the grid spacing. The choice of the control functions has a large influence of the solution on the system.

Some forms of the control functions are proposed by Thompson et al. [3], Thomas and Middlecoff [20] as well as Hsu and Lee [21].

The mapping procedure of section 2 required the generation of a structured mesh in a two dimensional domain, that is located between the outer boundaries of the blade passage and the boundary of the O-grid. Firstly, the surface is divided into four parts, marked by the red lines in figure 6(a). Within these four parts a structured mesh is generated by solving the system of elliptic partial differential equations, that contains suitable control functions. To obtain the solution of the system a Gauss-Seidel method, especially the method of Successive Over Relaxation, is used. A resulting mesh for a rotor domain is shown in figure 6(b).

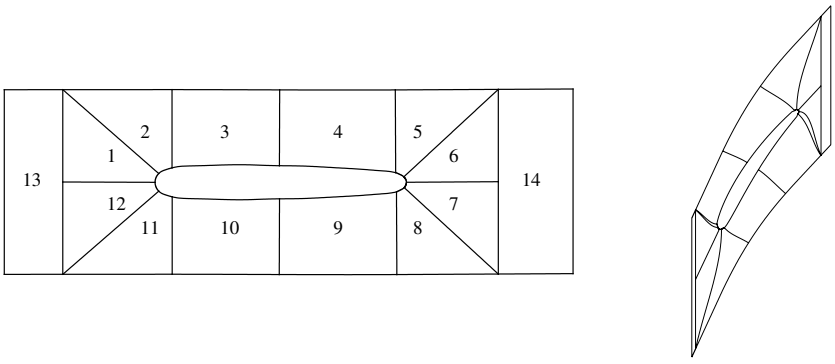


(a) Schematic depiction (b) Rotor domain

Figure 6: Background meshes, generated via solution of elliptic partial differential equations

3.3 Fourteen appropriate arranged blocks

Another way to generate a background mesh is an assembly of fourteen blocks inside the domain. The boundaries of the blocks are defined by linear or cubic splines. Figure 7(a) shows a schematic depiction of the background mesh and figure 7(b) an example for a rotor domain. This kind of background mesh has benefits concerning the computational time for its generation since no system of partial differential equations has to be solved. Furthermore, the inverse mapping procedure requires less time since fewer blocks have to be checked, during the search of the block that contains the regarded node of the unstructured mesh.



(a) Schematic depiction (b) Rotor domain

Figure 7: Background meshes consisting of fourteen blocks

4 Mesh quality

The mesh quality affects the efficiency and accuracy of CFD simulations. There are many investigations regarding this topic in the literature. Exemplary Katz and Sankaran [22] studied the mesh quality effects on the accuracy of CFD simulations on unstructured meshes by perturbing them randomly. The quality can be improved through several techniques, including point insertion/deletion, edge swapping and mesh smoothing. Canann et al. [23] give a detailed literature review about mesh improvement techniques, especially mesh smoothing. Batdorf et al. [24] use edge swapping as well as mesh smoothing to improve unstructured triangular meshes and performed CFD simulations, that quantify the effect on the solution time and convergence rate. Especially, mesh smoothing algorithms have been shown to be effective in improving the mesh quality. The most common and simplest smoothing technique is the Laplace smoothing. Another common technique is optimization-based smoothing, while it is more computationally expensive than Laplace Smoothing. But it gives better results, especially near concave regions [23]. Freitag [25] as well as Canann et al. [23] studied approaches to combine Laplace and optimization-based smoothing.

Quality measures, also denoted as distortion metrics, are used to determine the quality of a mesh. Stimpson et al give a collection of metrics for evaluating triangles, quadrilaterals, tetrahedra and hexahedra [26]. Amenta et al. [27] represent quality measures especially for triangles. Following Canann et al. [23] the quality measure q for a triangular element can be formulated as follows

$$q = \epsilon \frac{4\sqrt{3}\,A}{l_1^2 + l_2^2 + l_3^2} \quad \text{with} \quad \begin{cases} \epsilon = 1, \text{regular element} \\ \epsilon = -1, \text{inverse element} \end{cases} \tag{5}$$

where A is the area of the triangle and l_1, l_2 and l_3 are the edge lengths. To improve the quality of the unstructured two dimensional mesh a combination of edge swapping and several smoothing techniques is applied. The smoothing techniques are a constrained Laplace smoothing and an optimization-based smoothing.

4.1 Constrained Laplace smoothing

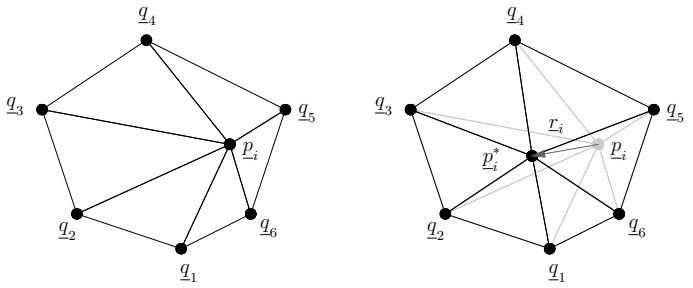


Figure 8: Submesh consisting of six triangular elements

The standard Laplace smoothing is an iterative process, where each node is placed at the average of the adjacent nodes. The left hand side of figure 8 shows a submesh consisting of six triangular elements, where $\underline{p}_i$ is the position vector of the regarded node and $\underline{q}_j$ are the position vectors of the m adjacent nodes. The new position vector $\underline{p}_i^*$ of the regarded node results to

$$\underline{p}_i^* = \frac{1}{m} \sum_{j=1}^{m} \underline{q}_j \quad . \tag{6}$$

It is recommendable to introduce a relaxation factor ω to increase the stability of the smoothing process. The factor should lie within the range $0.0 < \omega < 1.0$. The adapted formula to determine $\underline{p}_i^*$ reads

$$\underline{p}_i^* = \underline{p}_i + \omega \cdot \underline{r}_i \tag{7}$$

where

$$\underline{r}_i = \frac{1}{m} \sum_{j=1}^{m} (\underline{q}_j - \underline{p}_i) \quad . \tag{8}$$

This method usually works quite well for meshes in convex regions, but can results in distorted or inverted elements near concave boundaries [23]. In the flow passage the leading edge and trailing edge of the blade are such concave boundaries. Hence, the constrained Laplace smoothing, that is a variant of the Laplace smoothing, is used. This algorithm relocates the regarded node only if the submesh is improved in terms of a chosen quality measure. The mesh quality of the submesh is defined as the minimum quality q_{min} of the m triangular elements calculated by equation (5):

$$q_{min} = \min_{j \in \{1,\dots,m\}} (q_j) \tag{9}$$

Now the algorithm relocates the regarded node only if $q_{min}^* \geq q_{min}$, where q_{min}^* is the minimum quality obtained by equation 9 if the node would be relocated. To increase the flexibility of the algorithm it is useful to introduce inner iterations n_i within that the relaxation factor ω is halved if $q_{min}^* < q_{min}$. The resulting algorithm is presented for an initial ω of 1.0 in algorithm 1. In some cases, especially near concave boundaries, the constrained Laplace smoothing is unable to improve heavily skewed elements [23]. Hence, it is necessary to use alternative techniques, e.g. optimization-based smoothing to repair these elements.

Algorithm 1 Constrained Laplace smoothing

for $i = 1, n$ **do**
  $\omega = 1.0$
  $q_{min} = \min_{j \in \{1,\dots,m\}} (q_j)$
  $\underline{r}_i = \frac{1}{m} \sum_{j=1}^{m} (\underline{q}_j - \underline{p}_i)$
  for $k = 1, n_i$ **do**
    $\underline{p}_i^* = \underline{p}_i + \omega \cdot \underline{r}_i$
    $q_{min}^* = \min_{j \in \{1,\dots,m\}} (q_j^*)$
    if $q_{min}^* < q_{min}$ **then**
      $\omega = \omega / 2$
    else
      $exit$
    end if
  end for
end for

4.2 Optimization-based smoothing

Optimization-based smoothing directly improves the mesh quality, whereby heavily skewed and inverted elements can be improved. Canann et al. [23] give a detailed literature review about existing optimization-based smoothing techniques and distortion metrics. There are global and local optimization techniques, whereby recursive local optimization is the more feasible option [23]. In this paper a local optimization is presented. The goal of the optimization is to determine the new position vector $\underline{p}_i^*$ of the regarded node inside a submesh that maximizes the composite function $f_i(\underline{p}_i)$

$$min(-f_i(\underline{p}_i)) \quad , f_i(\underline{p}_i) = \sigma f_{1,i}(\underline{p}_i) + (1-\sigma) f_{2,i}(\underline{p}_i) \tag{10}$$

where

$$f_{1,i}(\underline{p}_i) = \frac{1}{m}\sum_{j=1}^{m} q_j \quad , f_{2,i}(\underline{p}_i) = min\{q_1, \ldots, q_m\} \quad . \tag{11}$$

Here σ is a relaxation factor and should lie within the range $0.0 < \sigma < 1.0$. The function $f_{1,i}$ conforms to the mean value of the qualities q_j and $f_{2,i}$ is the minimum quality q_j of all concerned triangles. For the investigated turbomachinery meshes $\sigma = 0.0$ has exposed as the most feasible choice. To obtain $min(-f_i(\underline{p}_i))$ a line search strategy in gradient direction is used.

4.3 Application

The outlined smoothing algorithms are applied to improve the mesh of a rotor blade of a high pressure compressor consisting of rather 473 thousand nodes. Figure 9 shows the percentage of triangles, that have a quality measure inside a specific interval, in the non-smoothed mesh and after application of the several smoothing algorithms. In the non-smoothed case it is obvious that some triangles have a really small quality. The smallest quality measure q_{min} is even negative, which signifies the presence of inverse elements. The application of the smoothing algorithms considerably improves the mesh quality. The standard Laplace smoothing as well as the constrained Laplace smoothing lead to a significant reduction of triangles with really small quality. The optimization-based smoothing leads to a minimum quality measure of $q_{min} = 0.59$, while the number of equilateral triangles decreases. Figure 10 shows the distribution of the inner angle ϕ of the triangles within the mesh. The standard Laplace smoothing as well as the constrained Laplace smoothing lead to a reduction of small and large angles. The optimization-based smoothing leads similar to the results of Freitag [25] to more angles in the regions of $45°$ and $90°$, while the number of angels near $60°$ decreases. Table 1 gives an overview of the minimum quality measures q and the minimum as well as the maximum angles ϕ. Additionally, the computational time required for the smoothing process including edge swapping is shown.

	$q_{min}[-]$	$\phi_{min}[°]$	$\phi_{max}[°]$	time $[s]$
no smoothing	-0.02	0.04	179.8	-
Laplace smoothing	0.23	8.6	152.6	4.7
Constr. Laplace smoothing	0.34	12.9	138.3	12.9
Opt.-based smoothing	0.59	23.4	114.5	420.0

Table 1: Results of mesh quality improvement

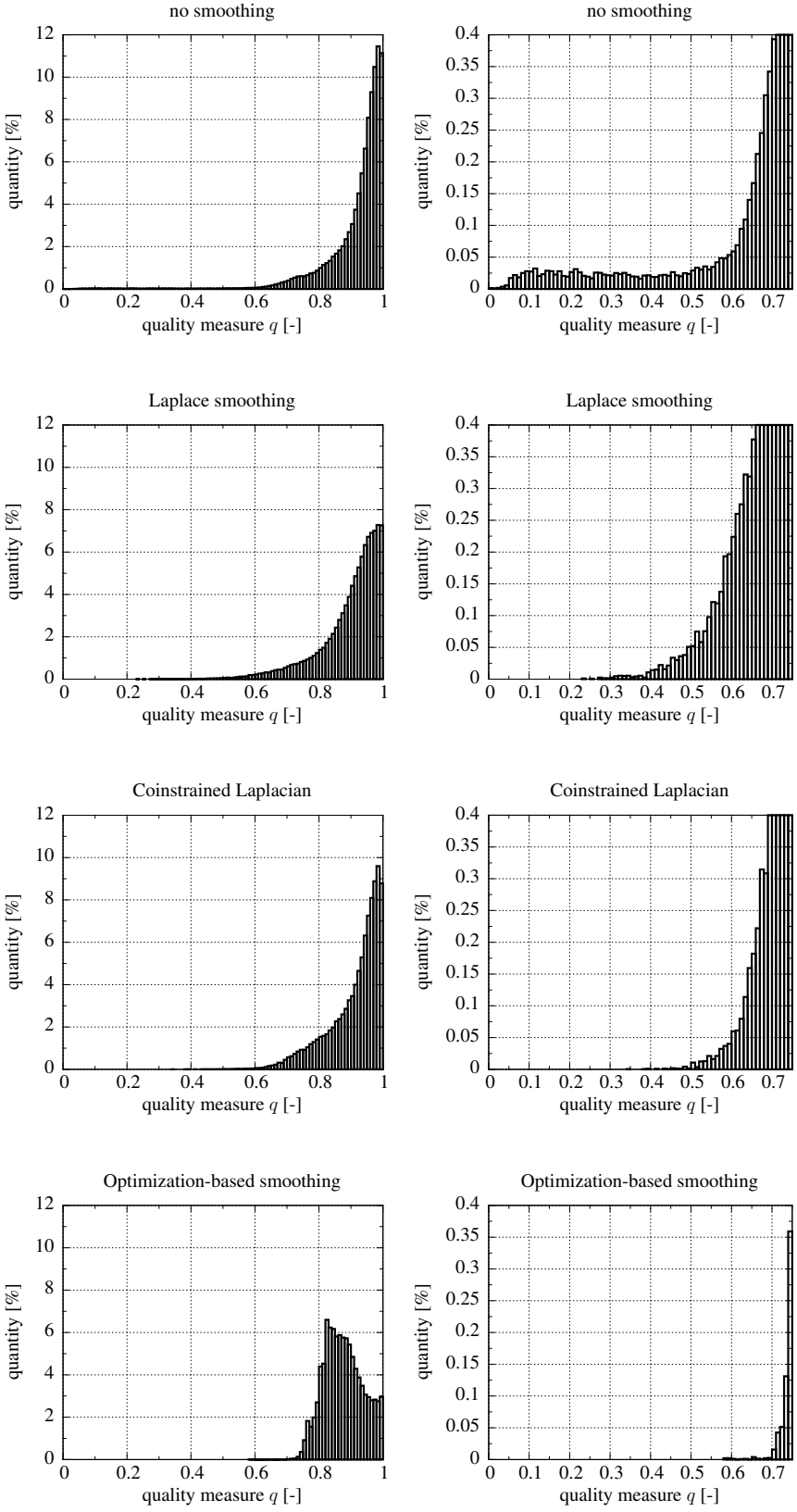


Figure 9: Distribution of quality measures q in a non-smoothed rotor passage and after application of several smoothing techniques

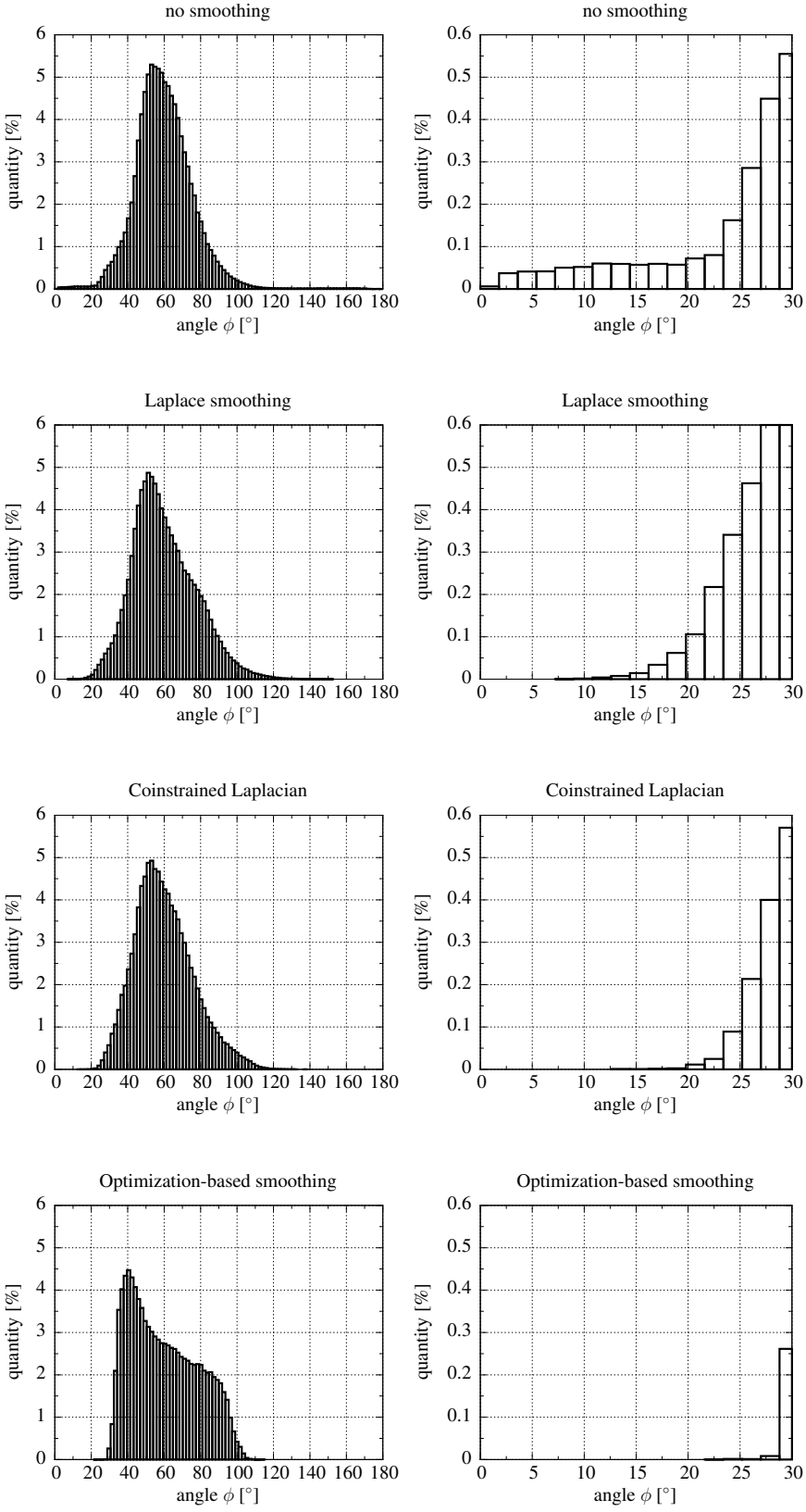
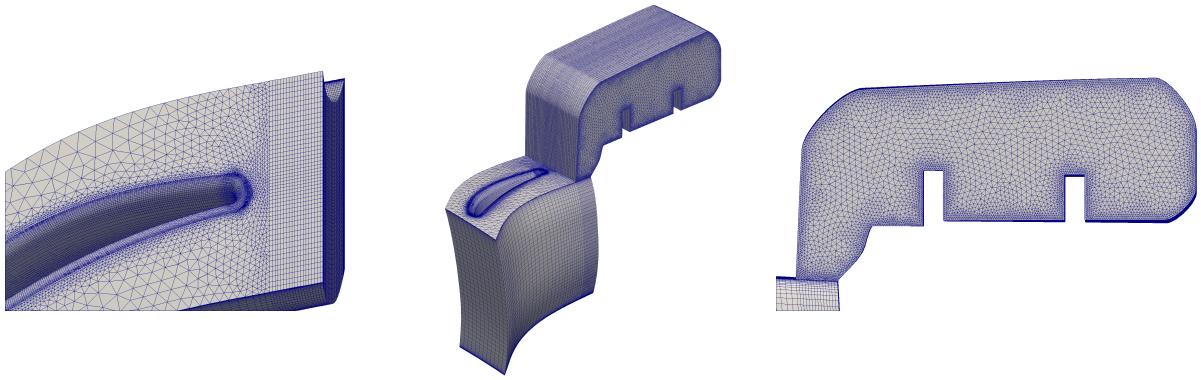


Figure 10: Distribution of angle ϕ in a non-smoothed rotor passage and after application of several smoothing techniques

5 GEOMETRICAL FEATURES

5.1 Modeling of cavities

Traditionally, the flow in turbomachinery blade rows has been simulated using a simplified computational domain consisting of just the blade surface and the main annulus. A more realistic representation of the geometry requires the modeling of cavities and bleed volumes. Burgos et al. [28] presented a method to generate semi-unstructured meshes, especially tailored for the meshing of turbomachinery blade passages and their associated cavities. In this work an analogous method is presented. The geometry of a cavity exhibits a large variation inside the (x, r) plane, while the variation in circumferential direction is usually negligible. Hence, the cavities are meshed in the (x, r) plane using an unstructured triangular mesh, that is extruded in circumferential direction. To connect the three dimensional mesh of the blade passage and the three dimensional mesh of the cavity it is necessary to ensure conformal grids at the interfaces. Since the interface mesh of the cavity has a structured topology, an adaption of the mesh of the blade passage is required. Therefore, a structured mesh at the inlet or outlet is attached. Figure 11(a) shows the structured mesh at the outlet of a stator domain. The figures 11(b,c) show the merged three dimensional mesh of a stator passage including a modeled bleed volume.



(a) Interface mesh (b) Overall mesh (c) Bleed volume

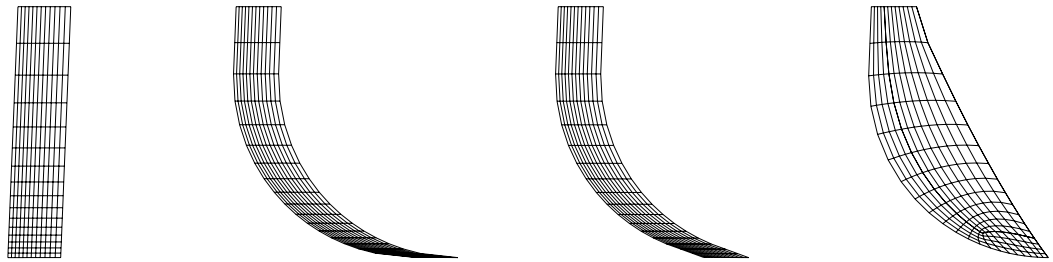
Figure 11: Modeling of a bleed volume

5.2 Modeling of fillets

The influence of the fillet between the blade and casing on flows in turbomachinery is widely investigated [29, 30, 31, 32, 33]. Zess and Thole [29] performed experimental measurements and CFD simulations, that verified the effectiveness of the leading edge fillet on a gas turbine vane by eliminating the horseshoe vortex. Pieringer and Sanz [30] investigated numerically the influence of the fillet between the blade and casing of a transonic turbine vane on the aerodynamic performance. The results show that considering the fillet in the simulation leads to a significant reduction of mass flow and a change of the specific angular momentum, depending on the flow situation. Furthermore, the efficiency increases when the fillet leads to an attached flow at the fillet and vice versa, decreases when the fillet additionally blocks the flow. Kügeler et al. [31] compared the results of CFD simulations of a 15-stage compressor with measurement data. The study shows differences in the numerical simulation for a geometry model with and without fillets. The fillets decrease flow separations and lead to a different flow behaviour, es-

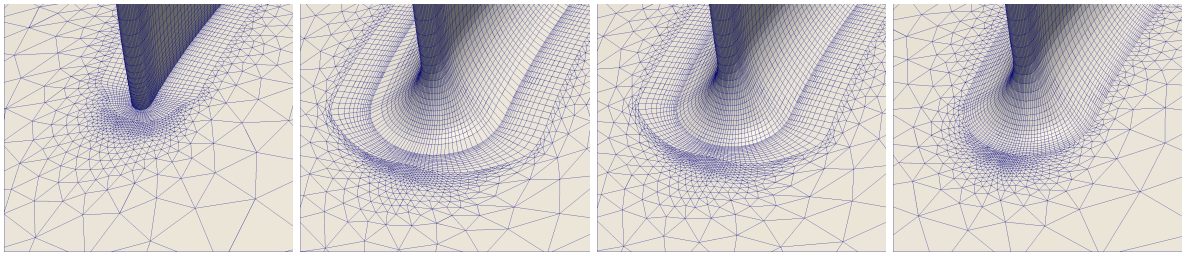
pecially at the endwalls. This results in a different stage loading, while the overall performance characteristics are similar.

In this work three methods to model blade fillets are presented. Within the first and the second one the topology of the semi-unstructured mesh remains. The blade surface is adapted, while the O-grid is generated as usual. Figure 12(a) shows a plane of the O-grid in normal direction of the blade surface, when no fillet is modeled. Figure 12(b) shows the case, if the fillet is reproduced completely (method 1). This obviously leads to heavily skewed cells near the casing. The heavily skewed cells can be prevented, when the fillet isn't reproduced completely (figure 12(c), method 2). Another method (method 3) to model blade fillets is the application of an alternative mesh topology within the O-grid near the casing. Here the two dimensional structured mesh, that is shown in figure 12(d), is generated by solving a system of elliptic partial differential equations. Figure 13(a) shows the surface meshes of the blades as well as the inner casing near the leading edge of a rotor blade without a fillet. Figure 13(b-d) show the surface meshes, whereas fillets are modeled applying the outlined methods.



(a) No fillet (b) Method 1 (c) Method 2 (d) Method 3

Figure 12: Implemented methods for fillet modeling



(a) No fillet (b) Method 1 (c) Method 2 (d) Method 3

Figure 13: Mesh at the leading edge of a rotor applying different methods for fillet modeling

Steady state CFD simulations were performed for a rotor passage, using the outlined methods to model a fillet. Figure 14 shows the residuals of the simulations with and without modeled fillet. The model including a fillet, that is modeled by the first method shows a relatively poor convergence rate because of heavily skewed cells near the trailing edge. That fact accords with the investigations of Batdorf et al. [24], that the mesh quality effects the convergence rate. Additionally the attained residual has a relatively poor level. In the other cases the simulations converge to an equal level, while the convergence rates are different.

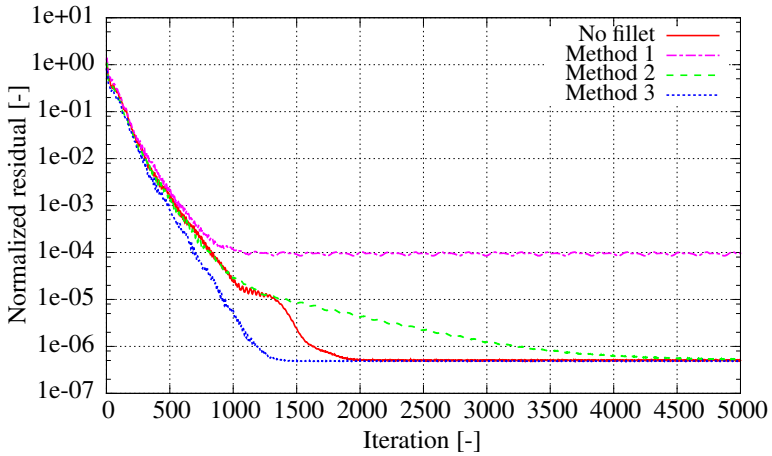


Figure 14: Residuals of steady state simulations applying different methods for fillet generation

5.3 Modeling of variable clearance sizes

Due to a varying stagger angle of variable stator vanes over the speed range of a high pressure compressor, the clearances between the blade and the casing vary too. At the design speed the clearance is approximately constant along the chord length of the blade. If the vanes are closed at part speed the clearances, especially between blade and inner casing, increase from penny to trailing edge. The chosen definition of the axisymmetric surfaces (section 2) allows the modeling of increasing clearances. Figure 15 shows the mesh of a variable stator vane including a constant clearance size between blade and inner casing (red mesh) as well as another one including a varying clearance size (blue mesh).

Steady state CFD simulations were performed for both cases. Due to the pressure difference from the pressure side to the suction side of the blade, the fluid propagates from the pressure side to the suction side through the clearance. Figure 16 shows, that in the case with the constant clearance size, the vortex generated at the suction side is significantly larger than in the other case. This leads to a different flow situation associated with a different pressure field at the blade surface, caused by the several clearance sizes. Consequently, the solution of unsteady CDF simulations, e.g. flutter analyses, can differ noticeable, depending on the modeled geometry.

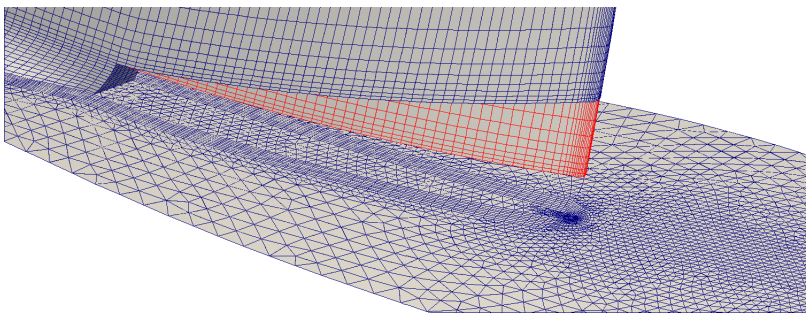
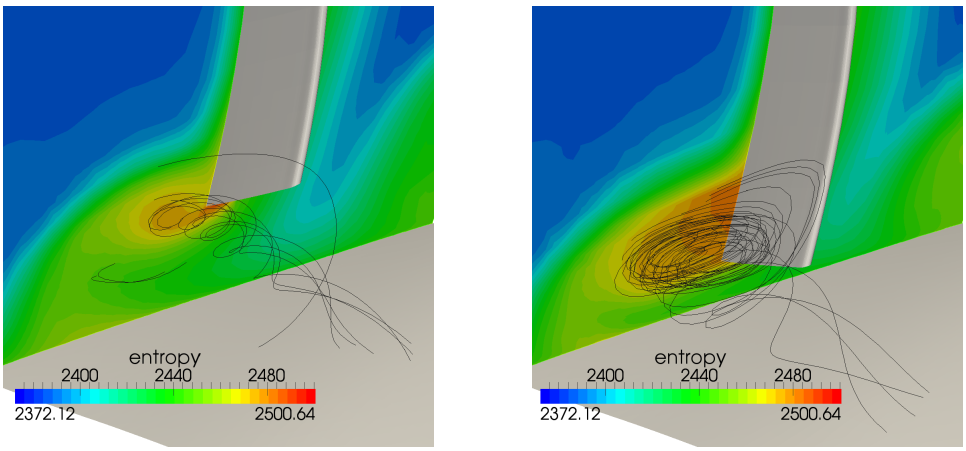


Figure 15: Different modeled clearance sizes of a variable stator vane (blue grid: varying clearance size, red grid: constant clearance size)



(a) Variable clearance size (b) Constant clearance size

Figure 16: Entropy and streamlines near the trailing edge

6 CONCLUSIONS

In this work a library for turbomachinery meshing was implemented. It enables the generation of semi-unstructured meshes for turbomachinery blade passages, including cavities, fillets and varying clearance sizes. The focus lies on the generation of a mesh, that represents the real geometry as accurately as possible, while the mesh quality is preserved.

Two approaches for the generation of background meshes were presented. The first approach divides the blade passage into four parts. Inside of these parts a structured grid is generated by solving a system of elliptic partial differential equations. The second approach is based on the split of the domain into fourteen blocks. It has benefits concerning computational time towards the first one, because of a faster generation procedure as well as a faster performance of the inverse mapping.

Another key aspect in mesh generation is the improvement of the mesh quality applying suitable methods. Since especially mesh smoothing algorithms have been shown to be effective in improving the mesh quality two smoothing algorithms, a constrained Laplace smoothing and an optimization-based smoothing, were presented. Both algorithms have benefits concerning the achieved mesh quality compared to the standard Laplace smoothing, while the computational time is longer. For the investigated turbomachinery meshes the constrained Laplace smoothing has exposed as the most feasible choice, because of a suitable combination of mesh quality and computational time.

Several methods for the modeling of fillets between blade and the casing were presented. The methods provide meshes with different qualities, that results into different convergence rates and residuals. Furthermore, the axisymmetric surfaces are dependent on the axial position, that enables the modeling of clearances with a variable size. CFD simulations for a variable stator vane with a constant clearance size between blade and inner casing as well as with a variable clearance size were performed. The results show a different flow behaviour near the clearance. This emphasizes the requirement of an accurate representation of the real geometry for CFD simulations of turbomachinery flows.

NOMENCLATURE

Symbols

A	Area of triangular element
f	Composite function
i, j, k	Indices
J	Jacobian
l_1, l_2, l_3	Edge lengths of triangular element
m	Number of adjacent nodes/elements
n	Number of free nodes within the unstructured mesh
n_i	Number of inner iterations (constrained Laplace smoothing)
P, Q	Control functions
q	Quality measure
(u, v)	Local coordinates
(x, y)	Cartesian coordinates (2D)
(r, ϑ, x)	Cylindrical coordinates
α, β, γ	Coefficients (system of elliptic partial equations)
ϵ	sign $(1, -1)$
(ξ, η)	Curvilinear coordinates
σ	Relaxation factor (optimization-based smoothing)
ϕ	Inner angle of triangular element
ω	Relaxation factor (constrained Laplace smoothing)
$\underline{p}$	Position vector of the regarded node
$\underline{q}$	Position vector of the adjacent node
$\underline{r}$	Displacement vector of the regarded node
$\underline{x}$	Vector of cartesian coordinates (x, y)
$\underline{x}_e, \underline{x}_n, \underline{x}_s, \underline{x}_w$	Boundary curves in cartesian coordinates (x, y)

Subscripts/Superscripts

*	Value after smoothing
$_{max}$	Maximum value
$_{min}$	Minimum value

REFERENCES

[1] F. Montomoli, M. Carnevale, A. D'Ammaro, M. Massini, *Uncertainty Quantification in Computational Fluid Dynamics and Aircraft Engines*. Springer, Cham Heidelberg New York Dordrecht London, 2015.

[2] J.F. Thompson, F.C. Thames, C.W. Mastin, *Boundary-Fitted Curvilinear Coordinate Systems for Solution of Partial Differential Equations on Fields Containing Any Number of Arbitrary Two-Dimensional Bodies*. NASA-CR-2729, 1977

[3] J.F. Thompson, F.C. Thames, C.W. Mastin, *TOMCAT—A Code for Numerical Generation of Boundary-Fitted Curvilinear Coordinate Systems on Fields Containing Any Number of Arbitrary Two-Dimensional Bodies*. Journal of Computational Physics, **24**(3), 274-302, Juli 1977

[4] J.L. Steger, R.L. Sorenson, *Automatic Mesh-Point Clustering Near a Boundary in Grid Generation With Elliptic Partial Differential Equations*. Journal of Computational Physics, **33**, 405-410, 1979

[5] J.F. Thompson, Z.U.A. Warsi, C.W. Mastin, *Numerical Grid Generation: Foundations and Applications*. North-Holland, 1985

[6] J.F. Thompson, B.K. Soni, N.P. Weatherill, *Handbook of Grid Generation*. CRC Press, 1999

[7] R.E. Smith, *Algebraic Grid Generation*. Applied Mathematics and Computation, **10-11**, 137-170, 1982

[8] T.I.-P. Shih, R.T. Bailey, H.L. Nguyen, R.J. Roelke, *Algebraic Grid Generation for Complex Geometries*. International Journal for Numerical Methods in Fluids, **13**, 1-31, June 1991

[9] L. Sbardella, A.I. Sayma, M. Imregun, *Semi-Structured Meshes for Axial Turbomachine Blades*. Int. J. Numer. Meth. Fluids, **32**, 569-584, 2000

[10] K. Kim, P.G.A. Cizmas, *Three-Dimensional Hybrid Mesh Generation for Turbomachinery Airfoils*. Journal of propulsion and power, **18**(3), May - June 2002

[11] A. Khawaja, Y. Kallinderis, *Hybrid Grid Generation for Turbomachinery and Aerospace Applications*. Int. J. Numer. Meth. Engng, **49**, 145–166, 2000

[12] M.L. Staten, S.A. Canann and S.J. Owen, *Locating Interior Nodes During Sweeping*. Engineering with Computers, **15**, 212–218, 1999

[13] R. Lohner, P. Parikh, *Generation of Three-Dimensional Unstructured Grids by the Advancing-Front Method*. International Journal of Numerical Methods in Fluids, **8**, 1135-1149, 1988

[14] J. Peraire, J. Peiró, K. Morgan, *Adaptive Remeshing for Three-Dimensional Compressible Flow Computations*. Journal of Computational Physics, **103**(2), 269-285, 1992

[15] S. Pirzadeh, *Unstructured Viscous Grid Generation by Advancing-Front Method*. NASA Constractor Report 191449, 1993

[16] D.J. Mavriplis, *Unstructured Mesh Generation and Adaptivity*. Technical Report ICASE 95-26, 1995

[17] J. Contreras, R. Corral, J. Fernandez-Castañeda, G. Pastor, C. Vasco, *Semi-Unstructured Grid Methods for Turbomachinery Applications*. GT2002-30572, ASME Turbo Expo 2002, Amsterdam, The Netherlands, June 3–6, 2002

[18] W.N. Gordon, C.A. Hall, *Construction of Curvilinear Coordinate Systems and Application to Mesh Generation*. International J. Num. Methods in Eng, **7**, 461-477, 1973

[19] J.F. Thompson, F.C. Thames, C.W. Mastin, *Automatic Numerical Generation of Body-Fitted Curvilinear Coordinate System for Field Containing Any Number of Arbitrary Two-Dimensional Bodies*. Journal of Computational Physics, **15**:3, 299-319, Juli 1974

[20] P.D. Thomas, J.F. Middlecoff, *Direct Control of the Grid Point Distribution in Meshes Generated by Elliptic Equations*. AIAA Journal, **18**, 652-656, 1980

[21] K. Hsu, S.L. Lee, *A Numerical Technique for Two-Dimensional Grid Generation with Grid Control at All of the Boundaries*. Journal of Computational Physics, **96**, 451-469, 1991

[22] A. Katz, V. Sankaran, *Mesh Quality Effects on the Accuracy of CFD Solutions on Unstructured Meshes*. Journal of Computational Physics, **230**, 7670–7686, 2011

[23] S.A. Canann, J.R. Tristano, M.L. Staten, *An Approach to Combined Laplacian and Optimization-Based Smoothing for Triangular, Quadrilateral, and Quad-Dominant Meshes*. 7th International Meshing Roundtable, 479-494, 1998

[24] M. Batdorf, L.A. Freitag, C. Ollivier-Gooch, *Computational Study of the Effect of Unstructured Mesh Quality on Solution Efficiency*. 13th Annual Computational Fluid Dynamics Meeting, Snowmass Village, 1997

[25] L.A. Freitag, *On Combining Laplacian and Optimization-Based Mesh Smoothing Techniques*. Trends in Unstructured Mesh Generation, ASME Applied Mechanics Division, **220**, 37-43, 1997

[26] C. Stimpson, C. Ernst, P. Knupp, P. Pebay, D. Thompson, *The Verdict Geometric Quality Library*. Sandia National Laboratories, 2007

[27] N. Amenta, M. Bern, D. Eppstein, *Optimal Point Placement for Mesh Smoothing*. Journal of Algorithms, **30**(2), 302-322, February 1999

[28] M.A. Burgos, R.Corral, J. Fernández-Castañeda, C. López, *Rapid Meshing of Turbomachinery Rows Using Semi-Unstructured Conformal Grids*. Engineering with Computers, **26**(4), 351-362, 2010

[29] G.A. Zess, K.A. Thole, *Computational Design and Experimental Evaluation of using a Leading Edge Fillet on a Gas Turbine Vane*. ASME Journal of Turbomachinery, **124**, 167-175, 2002

[30] P. Pieringer, W. Sanz, *Influence of the Fillet Between Blade and Casing on the Aerodynamic Performance of a Transonic Turbine Vane*. GT2004-53119, ASME Turbo Expo 2004, Wien, Austria, June 14-17, 2004

[31] E. Kügeler, D. Nürnberger, A.Weber, K. Engel, *Influence of Blade Fillets on the Performance of a 15 Stage Gas Turbine Compressor*. GT2008-50748, ASME Turbo Expo 2008, Berlin, Germany, June 9-13, 2008

[32] L. Castillon, B. Billonnet, S. Péron, C. Benoit, *Numerical Simulations of Technological Effects Encountered on Turbomachinery Configurations with the Chimera Technique*. 27th Congress of International Council of the Aeronautical Sciences, Nice, France, September 19-24, 2010

[33] R. Meyer, S. Schulz, K. Liesner, H. Passrucker, H. Wunderer, *A Parameter Study on the Influence of Fillets on the Compressor Cascade Performance*. Journal of Theoretical and Applied Mechanics, **50**, 131-145, 2012

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

CARTESIAN MESHES WITH DYNAMIC LOCAL WAVELET-BASED REFINEMENT FOR FLOW AROUND BODY PROBLEMS

Kirill D. Merkulov[1], **Igor S. Menshov**[1,2], **and Andrew V. Plenkin**[2]

[1] Dukhov All-Russia Research Institute of Automatics
22, Suschevskaya str., Moscow, Russia
e-mail: parovoz1991@yandex.ru, imen57@mail.ru

[2] Keldysh Institute of Applied Mathematics
4, Miusskaya sq., Moscow
e-mail: menshov@kiam.ru, golden_dragon@mail.ru

Keywords: Adaptive mesh refinement, finite volume methods, Godunov fluxes, free boundary method, wavelets, WENO-technology.

Abstract. *This document describes our project of making Cartesian multilevel AMR-method for flow around body problems in Eulerian statement. The main idea of Cartesian meshes with local refinement is based on the simplicity of Cartesian algorithms and the necessity of time and memory economy during the calculations. Tree-like data format for mesh representation [1] makes all the algorithms of mesh performance (generation, reconfiguration and traversal) rather simple and fast. We use Free boundary method [2] for geometry treatment on Cartesian meshes, WENO-technology for interlevel solution projecting and B-spline wavelet for data analysis.*

.

1 INTRODUCTION

We present modern computational method for fluid dynamics based on both adaptive Cartesian meshes [1] and Free Boundary Method [2] for geometry treatment on Cartesian meshes. Suggested format of data representation based on quadric trees simplify procedures of mesh generation, reconfiguration and traversal. Specific way of cell numeration accelerates neighbor searching. Mesh configuration on each time step depends on the current position of the solids inside the computational domain and the local solution behavior. Free Boundary Method allows us to handle static or moving solids of complex geometry on the meshes with the simple rectangular cells. We use Finite Volume Method with Godunov or Rusanov fluxes as the main calculation scheme. Wavelet decomposition is used for quick parameter-free analysis of numerical solutions. Several different algorithms of multilevel mesh adaption that used binary criteria of function smoothness are compared.

2 MATHEMATICAL MODEL

In this paper, we propose a variant of mesh adaptation to solve the problem of an ideal gas movement, suitable for massive parallelization. The mathematical formulation of the problem is reduced to solving the system of equations

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + div(\rho \mathbf{U}) = 0, \\ \dfrac{\partial (\rho \mathbf{U})}{\partial t} + div(\rho \mathbf{U}\mathbf{U}) + grad(p) = 0, \\ \dfrac{\partial (\rho E)}{\partial t} + div(\rho \mathbf{U} H) = 0, \end{cases} \tag{1}$$

in a region $V(\Gamma) \subset \mathbf{R}^2$ bounded by a given closed curve $\Gamma = \Gamma(x, y)$. Here we introduce the standard notation: $\mathbf{U}$ - velocity vector, ρ - density, p – pressure, $H = E + p/\rho$ - the total specific enthalpy, $E = e + 0.5\mathbf{U}^2$ - the total specific energy, e - specific internal energy.

Equation of the system correspond to the laws of conservation of mass, momentum and energy. To close these systems serve as the equation of state. The gas is assumed ideal, i.e $p = (\gamma - 1)\rho e$.

For the Euler equations put the standard initial boundary value problem in the area.

Also the computational domain can include a solid body bounded be known curve $\Gamma_S \subset \Gamma$.

3 NUMERICAL METHOD

3.1 Basic calculation scheme

This system (1) written in vector form is solved numerically in the time step on a fixed Cartesian grid. This is done using the finite volume method using the approximate solution of the Riemann Problem by Rusanov [3] or its exact solution and counting flows by Godunov [4].

3.2 Free boundary method

This method was designed for numerical solving the Euler equations of compressible fluid dynamics in domains that contain solid, impermeable, and in general case moving entries (objects) with stationary Cartesian grids. The method proposed is based on the approach of immersed boundary, in which calculations are carried out over all cells of a Cartesian grid that

covers the whole computational domain including solid entries. The influence of the solid surface on the gas flow is taken into account by introducing effective fluxes of mass, momentum, and energy on the right-hand side of the governing equations. We use Newton iterations for integrating implicit equation with non-zero right-hand side. The Jacobean was derived analytically. The stability of numerical method is guaranteed by satisfying the CFL condition.

4 CARTESIAN MESH REFINEMENT

4.1 Data format

To adapt the originally structured Cartesian grids, where the partition is meant for any preselected the law, it is convenient to use different tree structures. In addition, they are well-known fast recursive algorithms for crawling and rebuild.

We use quadric trees for the 2D adaptive grids and OC3 format for 3D-fields [5]. Each cell can only be broken into four equal division in half in each direction. Take the original rectangular grid (available in uneven) size $M \times N$.

Each cell is described by its level (for the basic mesh size is equal to zero), the virtual position at this level - a couple of parameters: row on level and column on level. In addition, it has a divided flag, indicating whether this cell is the final settlement (the corresponding leaf of the tree) or it has 4 children cells of the next level. For divided cells we save pointers to all its children, for leaf cell – the pointer to corresponding gas dynamic parameters.

4.2 Wavelet-based analyzer

Suppose on a rectangular grid of discrete specified function $\{g_{ij}\}, 0 \le i < 2N, 0 \le j < 2N$ (e.g., density or pressure of the gas-dynamic flow). Also considered to be specified threshold (the value of which is caused by a step of the computational grid and the order of approximation calculation method). Then we can calculate the wavelet decomposition of a two-dimensional array of data and define a set of flags [6]:

$$flag_{ij} = \left(lh_{\lfloor i/2 \rfloor \lfloor j/2 \rfloor} > \delta \right) \vee \left(hl_{\lfloor i/2 \rfloor \lfloor j/2 \rfloor} > \delta \right) \vee \left(hh_{\lfloor i/2 \rfloor \lfloor j/2 \rfloor} > \delta \right), 0 \le i < 2N, 0 \le j < 2N \qquad (2)$$

Flags define the area where you want to save a fine mesh, and in other areas it is possible to use a coarse grid.

4.3 Multilevel adaptation

Now consider the analyzer as a black box, issuing 1 if the cell is necessary to be divided and 0 if it is not necessary. Black box is applying to some window of cell environment (cross from 5 axis-aligned quasicells (real cells or parent cells without data) equal to chosen by size with the center of chosen cell or a square with a side of 5 cells and center in a given).We projected all the data to current pattern, if one pattern element is part of a larger cell, we give parent's value, if it consists of several subcells – we compute the value as conservative averaged.

Now we need to construct an algorithm that generates a grid configuration that is adequate to real physical field, written on current mesh.

We propose the following procedure:

1. First, hold the mesh coarsening procedure: those cells that can be made larger, according to the wavelet analysis are combined into one without violating the existing tree structure. In the cycle at levels ranging from the penultimate to the zero (base) are looking for a virtual cell, which may be roughened (have four physical descendants) and the criteria coarsening (black box for the parent cell produces zero).

2. After that we make refinement procedure. We pass through the levels from zero to the penultimate, looking at them physical cell, for which the dividing criteria is satisfied (black box provides 1) and divide them into 4 parts.

3. Draw the mesh smoothing process: remove all the collisions, where two neigboring cells differs more than twice by size. We do it by dividing too large cells into subcells until we rich the correct mesh.

5 NUMERICAL RESULTS

5.1 Gas dynamic tests

The advantages of suggested approach were demonstrated by some significant numerical calculations. We have computed single point explosion Sedov test [7] and Liska test 6 [8] on adaptive meshes with different value of maximal adaptation level. Results and statistics of this calculations are presented on figures1 and 2.

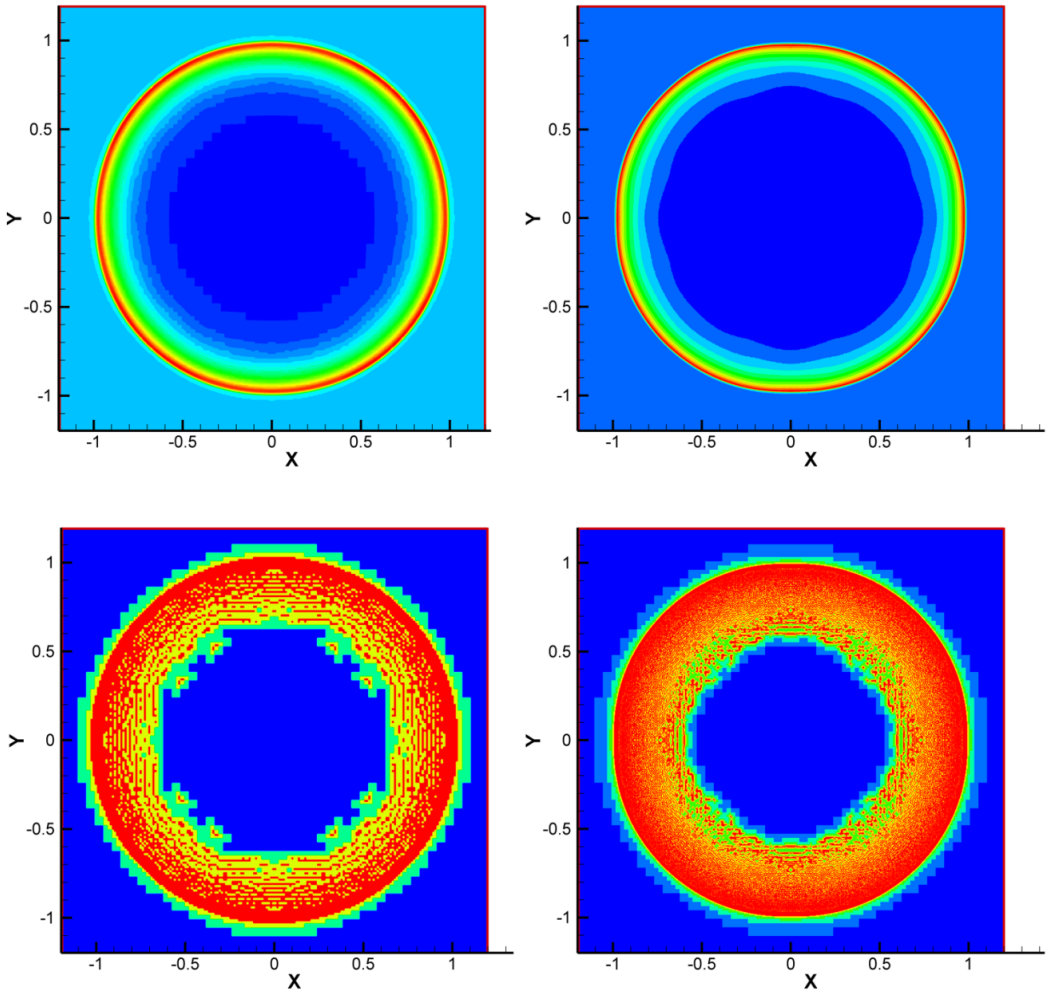


Figure 1: The results of Sedov problem calculations (top row: field density by color from blue to red) in the multi-level adaptive grid and the corresponding configuration of computational grids by the end of the calculation(colors from blue to red shows the level of physical cell). The left column – 4 adaptation levels, right – 6 adaptation levels, the base mesh 50x50.

With an increasing number of adaptation levels significantly increases the accuracy of the calculations (density at the shock wave is approaching to analytical value), decreases the percentage of the smallest. This fact talks about better localization features of algorithm with bigger number of levels. All the calculations were performed with the Rusanov flow to avoid carbuncle effect [9].

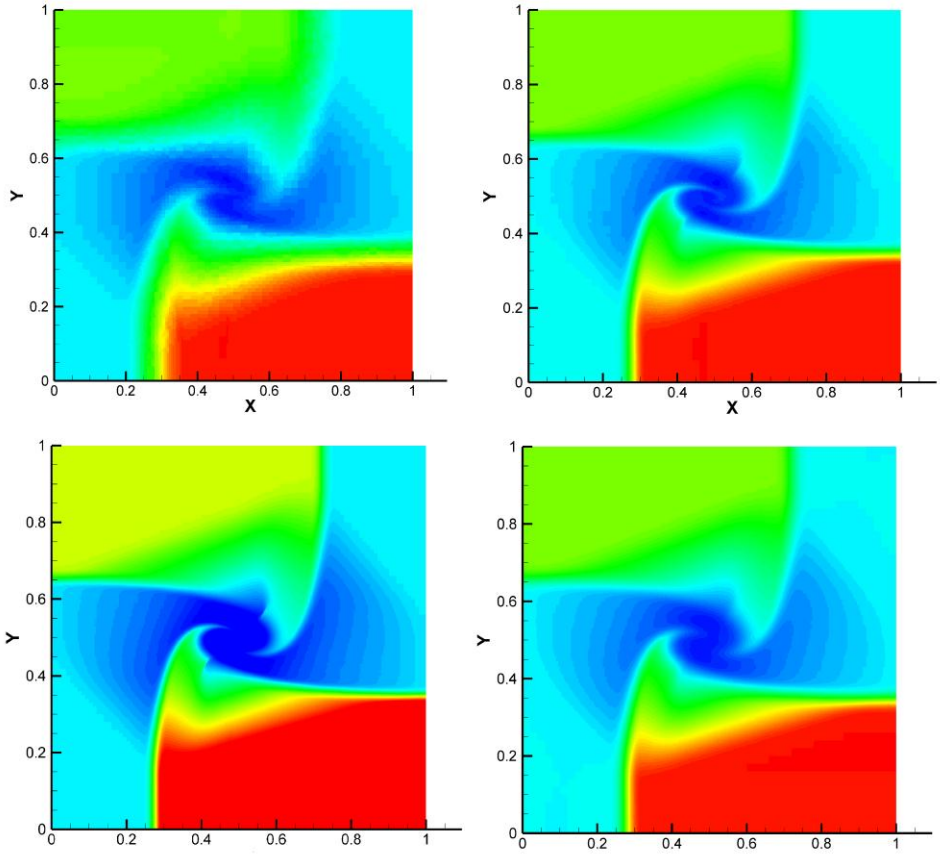


Figure 2: The calculated density fields for Liska test 6 by Godunov method with 3 levels of adaptation (top left), 5 levels of adaptation (top right), 7 levels of adaptation (bottom left) and Rusanov method with 7 levels of adaptation with the base grid of 50x50 cells.

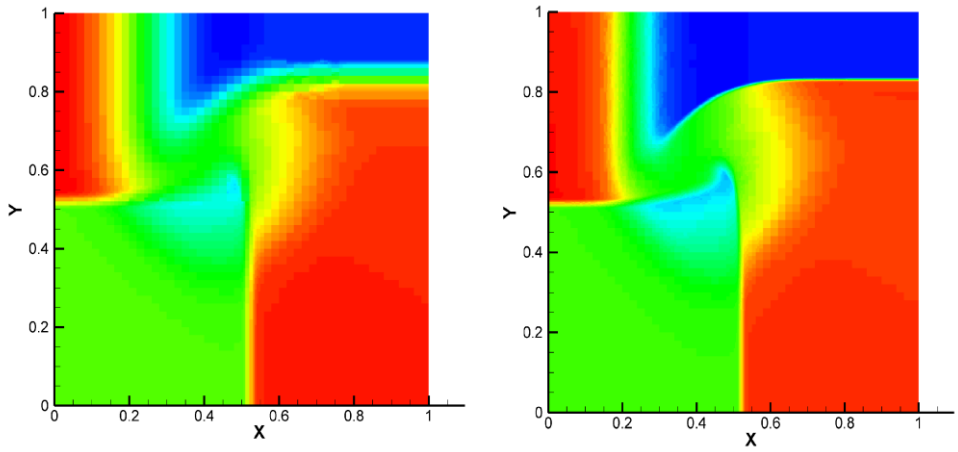


Figure 3: The calculated density fields for Liska test 16 by Godunov method with 3 levels of adaptation (left) and 6 levels of adaptation (right) with the base grid of 50x50 cells.

Increasing the number of adaptation levels when using Godunov method allows to identify the spatial effects of Rayleigh-Taylor instability for the test №6 along the coordinate axes and the rarefication zone for the test №16 (see. figure 3). Calculations using Rusanov doesn't shov this features even using a lot of adaptation levels.

5.2 Flow around body problems

To show the opportunities of the our approach with geometrical solids we present the result of calculating the flow around aircraft fuselage (see fig. 4).

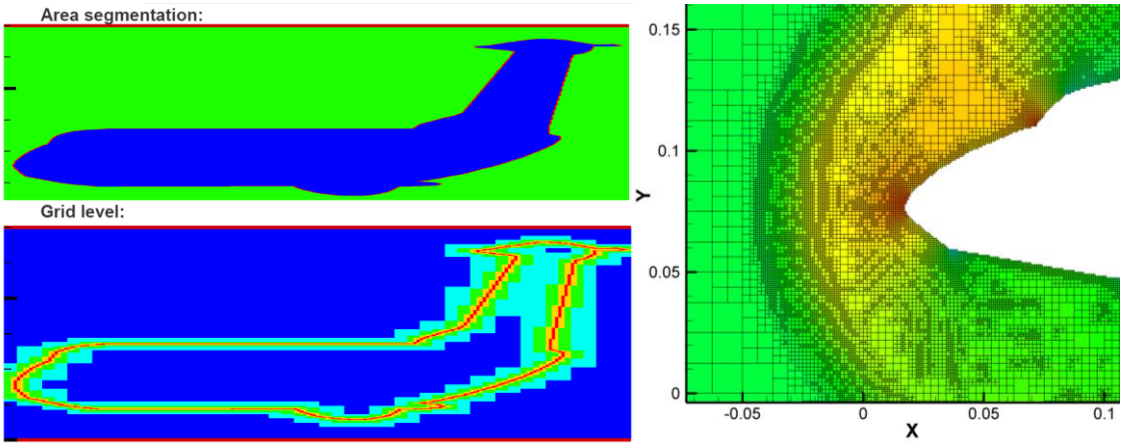


Figure 4: Area segmentation and pressure field for flow around aircraft model

6 CONCLUSIONS

We have designed the program code that allow us to calculate different CFD problems in the computational domains with several solids of complex shape using Cartesian meshes that dynamically rebuilds according to solids positions and solution behavior.

This work was supported by the grant No 14-11-00872 from Russian Scientific Fund.

REFERENCES

[1] K. Merkulov. Wavelet-Based Local Mesh Adaptation with Application to Gas Dynamics. Parallel Computing Technologies. Springer International Publishing, 2015, Pages 426-435.

[2] I. Menshov, M. Kornev. Free boundary method for numerical solving gas dynamics equations in domains with varying geometry. Matem. Mod., 2014, Volume 26, Number 5, Pages 99–112.

[3] Godunov S. K. et al. Numerical solution of multidimensional problems of gas dynamics //Moscow Izdatel Nauka. – 1976. – T. 1.

[4] Rusanov V. V. The calculation of the interaction of non-stationary shock waves and obstacles //USSR Computational Mathematics and Mathematical Physics. – 1962. – T. 1. – №. 2. – C. 304-320.

[5] Sukhinov A. A. Construction of Cartesian meshes with dynamic adaptation to the solution //Matematicheskoe Modelirovanie. – 2010. – T. 22. – №. 1. – C. 86-98.

[6] Afendikov A. L., Merkulov K. D., Plenkin A. V. Local mesh adaptation in gas dynamic problems with the use of wavelet analysis. – 2014.

[7] Sedov L. I. Similarity and dimensional methods in mechanics. – CRC press, 1993.

[8] Liska R., Wendroff B. Comparison of several difference schemes on 1D and 2D test problems for the Euler equations //SIAM Journal on Scientific Computing. – 2003. – T. 25. – №. 3. – C. 995-1017.

[9] Afendikov A. L., Merkulov K. D., Plenkin A. V. Comparative analysis of approaches for modeling of gasdynamic flows on two-level adaptive grids. – 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MESH GENERATION METHOD FOR HISTORICAL MONUMENTAL BUILDINGS: AN INNOVATIVE APPROACH

Giovanni Castellazzi[1], Antonio Maria D'Altri[1], Stefano de Miranda[1], Francesco Ubertini[1], Gabriele Bitelli[1], Alessandro Lambertini[1], Ilenia Selvaggi[1], Antonio Tralli[2]

[1] Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM), University of Bologna
V.le Risorgimento 2, Bologna, Italy
e-mail: giovanni.castellazzi@unibo.it, antoniomaria.daltri2@unibo.it, stefano.demiranda@unibo.it, francesco.ubertini@unibo.it, gabriele.bitelli@unibo.it, alessandro.lambertini@unibo.it, ilenia.selvaggi2@unibo.it

[2] Engineering Department, University of Ferrara
Via Saragat 1, Ferrara, Italy
e-mail: tra@unife.it

Keywords: Semi-automatic mesh generation, Simplified FE discretization, Historical monumental buildings.

Abstract. *The numerical modeling of historical monumental buildings is a challenging task for contemporary civil engineers. One of the main reasons for this is that the use of traditional simplified structural schemes is inadequate due to the complex geometry of such historical structures. Therefore, it is necessary to resort to a fully 3D modeling technique that is performed using Computer Aided Design (CAD). In general, CAD-based modeling is an expensive and complex process and it is often performed manually by the user, which inevitably leads to the introduction of geometric simplifications or interpretations. In this study, an innovative mesh generation approach for the structural analysis of historical monumental buildings is presented. The method consists in a peculiar breakdown of the geometry starting from laser scanner or photogrammetric surveys. Moreover, this new approach involves a structural discretization that guarantees the generation of 3D finite element meshes as well as their mechanical characterization. The most relevant feature of the proposed method is the possibility to directly exploit 3D and detailed point clouds surveyed on historical buildings for structural purposes. As a result, a large reduction in the required time in comparison to CAD-based modeling procedures is achieved. A geometrical and structural validation of the method is carried out on a masonry tower application. The findings show good reliability and effectiveness of the mesh generation approach.*

1 INTRODUCTION

Nowadays, when dealing with structural analyses of historical buildings the Finite Element Method (FEM) is very often used in order to manage the huge complexity of the problem. However, the numerical modeling of historical monumental buildings is still facing significant difficulties linked to the description of nonlinear non-homogeneous material such as masonry, of complex geometry and boundary conditions [1, 2, 3]. Indeed, the use of traditional simplified structural schemes is frequently inadequate. Consequently, the recourse to a fully 3D modeling using the Computer Aided Design (CAD) is commonly unavoidable. Generally, CAD based modeling is a laborious operation which inevitably leads to the adoption of geometric simplifications (*Defeaturing*) or interpretations to speed up the process.

In order to reduce the time required by the creation of the geometrical model of these structures, a crucial support can be supplied by automatic advanced survey techniques such as Terrestrial Laser Scanner (TLS) or photogrammetry, which can rapidly generate 3D detailed point clouds. In particular, in the field of architectural heritage several TLS and photogrammetric applications have been performed: from simple documentation to monitoring the condition of historical buildings, and also in order to support restoration works or structural checks [4, 5]. Moreover, aiming at designing the urgent rehabilitation interventions of damaged structures after seismic events the automated survey techniques (e.g. TLS or photogrammetry) are the only tools which can lead to a rapid and detailed 3D geometric acquisition of buildings.

Nevertheless, aiming at the rapid generation of FE meshes the high geometrical accuracy of the TLS and photogrammetric surveys become a drawback. Indeed, several studies try to transform 3D point clouds in FE models, but in most cases the output is too simple or considerably simplified. For instance, in [6] a 3D point cloud is used to generate models of the cross sections of historical walls for structural analysis application, while in [7] an example of FE analysis of a historic theater is performed using laser scanning data limited to the inner surfaces of the building. Other interesting contributions are proposed in [8, 9], where an attempt to precisely capture the geometry of the building through the automatic reconstruction of its boundary is presented. Moreover, in [10] a method to automatically transform point cloud data into solid models for computational modeling is developed. The resulting model captures the three-dimensionality of the survey, but does not capture the whole structure, since it is designed for the façade only.

When dealing with complex historical buildings, one of the most frequent problems is represented by the impossibility to generate "closed surfaces" from the point cloud of the surveyed object. Therefore, it is not possible to directly transform the outer surfaces into a solid geometry and consequently into a FE mesh as done for various objects (see for instance [11, 12]).

Aiming at solving this lack of numerical tools, a new procedure for the rapid transformation of a 3D point cloud into a FE model has been presented by the authors in [13]. The procedure, called *CLOUD2FEM*, starts from an automated survey of historical monumental buildings and semi-automatically generates a FE model. In this study, in order to deepen the potentialities of the method some further findings are presented thought an application to the case study of the main tower of the San Felice sul Panaro (Italy) fortress.

2 AN INNOVATIVE MESH GENERATION APPROACH

The new mesh generation approach aims at offering a simple and effective tool for practitioners to be able to build FE models of complex monumental buildings with a minimal time investment. Indeed, the method principal innovation is the simple transformation of complex point clouds (originating from TLS or photogrammetric surveys) into FE models.

In order to achieve this goal, we propose to break down the geometry of the structure through the application of a systematic procedure that recursively performs the subdivision of the 3D domain into bi-dimensional sub-domains. This operation can be carried out by slicing the structure perpendicularly to an opportune direction (typically the vertical direction). The slicing step is chosen according to the complexity of the building along the slicing direction. A boundary polygon that encloses the points of each slice can be computed using a concave hull algorithm. Consequently, it is possible to obtain a filled region for each slice of the building that describes the entire structure.

In order to speed up the meshing operation and to guarantee the automatic mesh generation, each slice is idealized as a digital image composed of picture elements (pixels) with a certain resolution. Since the digitalization is performed on each slice with a fixed space region, they are stackable. The slices stacking sequence generates the volume elements (voxels). The full reconstruction of the original 3D geometry is obtained by stacking all of the slices: this produces a 3D matrix composed by voxels. The resulting dataset is simple and easy to use with the finite element technique: each voxel is automatically transformed into an eight-node hexahedral finite element. For practical usage this operation could be performed selectively: for instance only on filled parts (certain voxel values). Therefore, the structure is completely discretized as an unique continuum generated by the assembly of eight-node hexahedral elements. The resulting FE model is characterized by elements of the same dimensions. This aspect introduces an automatic defeaturing of the model that the user can set according to the structural complexity.

Once the mesh is built, its material assignment can be generally conducted by selecting single or groups of elements. In addition to this, the mechanical characterization of the FE model can be conducted in a further simpler way. Indeed, once the digitalization of each slice has been concluded, the material characterization of the FE mesh can be simply conducted on each slice by the user through a material ID assignment before the stacking operation (i.e. through the pixel code in a bi-dimensional environment).

Upon these features the resulting discretized geometry is already containing all the information to use with the FE model including the mechanical properties associated with the material features. The new mesh generation approach guarantees by means of a simple procedure the construction of a discretized geometry and then an automatic creation of a fine FE solid model. This rational organization is certainly the key novelty introduced by the method. Furthermore, since we are dealing directly with the definition of the FE nodal coordinates and with the connectivity matrix the proposed method is generally customized to work with any commercial FE software.

3 APPLICATION TO THE MASTIO TOWER

In order to deepen the potentialities of the proposed mesh generation method, the application to the main tower (Mastio) of the San Felice sul Panaro Fortress is presented and discussed, see Figure1. Such a monumental building is located near the city of Modena, in the town of San Felice sul Panaro (Italy). In 2012, it has been hit by the Emilia earthquake with two magnitude peaks on May 20^{th} ($M_W = 5.86$) and on May 29^{th} ($M_W = 5.66$), and it is object of several studies that aim to preserve its integrity [14, 15, 16].

3.1 Mesh generation and optimization

A morphological survey for the San Felice sul Panaro Fortress was planned by request of the Municipality in order to generate a functional representation of the state of the building after

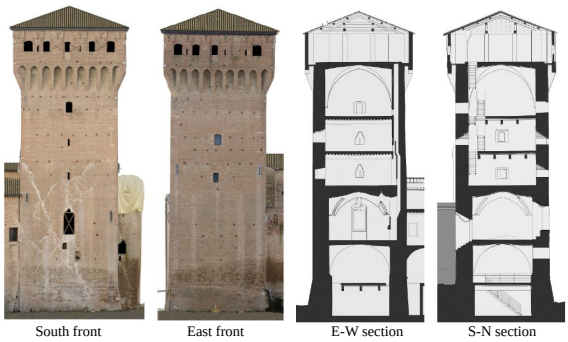


Figure 1: The main tower of the San Felice sul Panaro fortress.

the earthquake. In particular, a fine laser scanner point cloud of the whole building (including inner rooms) has been surveyed. In order to apply our mesh generation method, this point cloud has been exploited.

An accurate description of the point cloud manipulation phases as well as the FE mesh generation of the Mastio tower is reported in [13]. The first points cloud slicing has been performed using a vertical slicing step of $0.200\ m$ with the resolution of the digitalization of each slice equal to $0.115\ m$. This assumption produced a fine FE model which counts 661,105 elements and 745,668 nodes (i.e. 2,237,004 dofs). Such a model, which accurately reproduces the geometrical features of the structure as shown in [13], is evidently usable for linear static and dynamic analyses. However, it ends to be very time consuming when used with nonlinear analyses. Here, we present some comparisons using a optimized (coarse) model characterized by a vertical slicing step equal to $0.250\ m$ with the resolution of each slice also equal to $0.250\ m$, which is, more or less, the greater dimension of a brick. Therefore, the consequent mesh is only composed by cubic eight-node hexahedral finite elements (each one $0.250 \times 0.250 \times 0.250\ m$). The resulting stacking sequence is composed of 121 horizontal slices, where each one is represented by a grid of 52×50 pixels. In Figure 2, the stacking sequence of three illustrative slices is sketched. The final mesh is then characterized by a

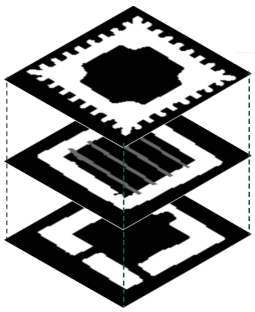
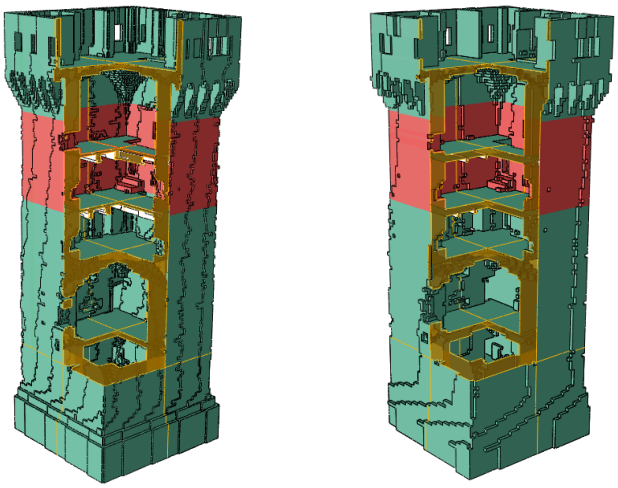


Figure 2: Example of stacking sequence.

considerable reduction of the number of degrees of freedom (443,076). More precisely, it counts 147,692 nodes and 118,554 elements. By comparing the two meshes in Figure 3 (the finer mesh in Figure 3(a) and the coarse mesh in Figure 3(b)) as well as in Table 1 it appears that the two models are equivalent from the structural point of view. Figure 4 illustrates the detail of the vault of the 6th level, which is in particular a groined vault (Figure 4(a)). As can be noted, when the surface is irregular (curved) (e.g. for vaults) the resulting FE mesh is a jagged representation of the original geometry (Figure 4(b)). Despite this fact, it is always possible to



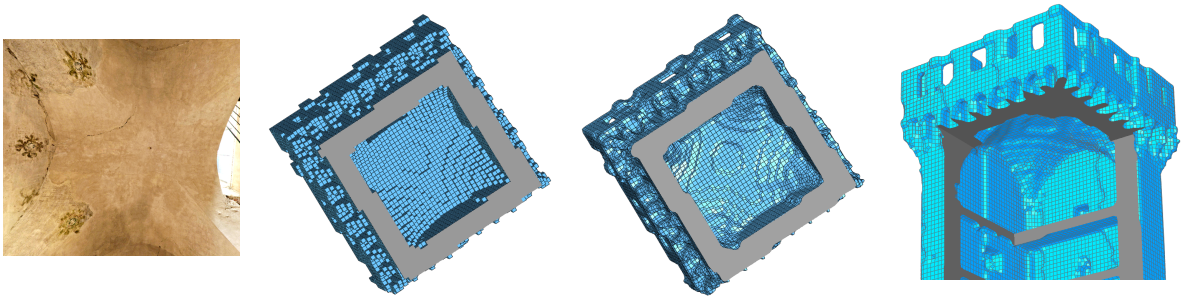
(a) Finer mesh (b) Coarse mesh

Figure 3: Meshes. Each color is correspond to a material index: green portions indicate masonry, red portions point at reinforced masonry and white regions refer to timber elements.

Model	Mass $[ton]$	Max dimensions $\{L \times B \times H\}[m]$	h_g $[m]$
Finer	3,032.11	$9.90 \times 9.80 \times 30.60$	14.07
Optimized	3,081.99	$10.00 \times 10.00 \times 30.50$	13.89

Table 1: Mass, overall dimensions and center of mass height.

improve the mesh accuracy using a smoothing method to reduce the faceting, see Figure 4(c) and Figure 4(d). These methods decrease high curvatures variations (jag) and have to be chosen in order to not produce shrinkage, see for instance [17]. Despite the geometrical improvement, the mesh enhancements are limited by the performance of the parametric finite elements [18]. Nevertheless, if we aim to assess the global behavior of a historical structure, the geometrical accuracy of the raw mesh can be considered satisfactory even if vaults are present. Furthermore, the recovered fields can be improved by standard recovery procedure [19].



(a) Groined vault (b) Raw mesh (c) Smoothed mesh (d) Smoothed mesh (detail)

Figure 4: Detail of the groined vault of the 6th level.

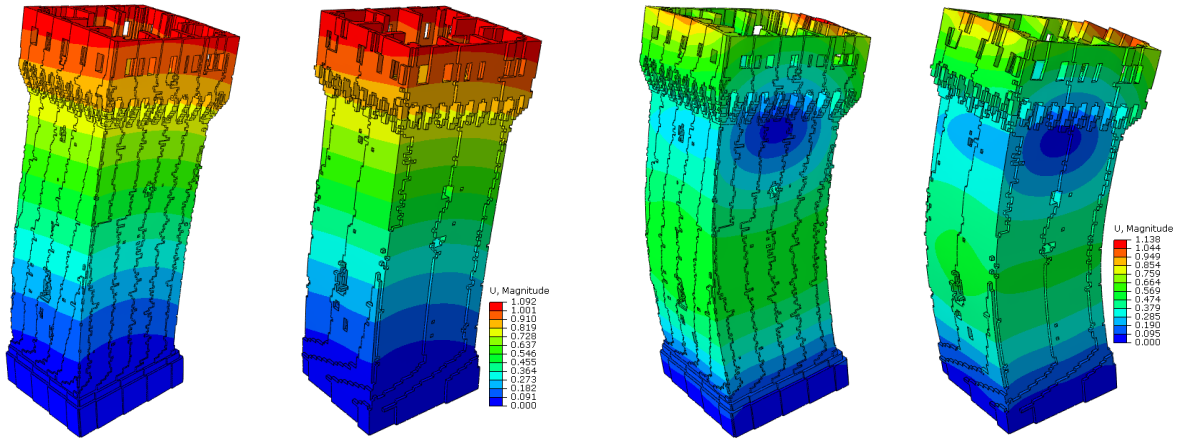
3.2 Structural Analyses

The semi-automatically generated FE models are tested within structural analyses. In order to assess the accuracy of the coarse model, a linear natural frequency analysis (eigenvalue analysis) is carried out and a comparison with the finer model is performed. For simplicity, clamped boundary conditions have been considered for nodes located at the ground level. Table 2 collects the obtained results in terms of computed frequencies and computed errors. It appears that the computed error is always less than 5.5% for the first six modes. Figure 5 illustrates the

Mode no.	Finer Fr. (Hz)	Coarse Fr. (Hz)	Error (%)	Mode descr.
1	1.9131	1.8635	2.592%	1^{st} bending mode (E-W)
2	1.9276	1.8705	2.962%	1^{st} bending mode (N-S)
3	4.5437	4.4004	3.154%	torsional mode
4	7.0804	6.6999	5.374%	2^{nd} bending mode (E-W)
5	7.1654	6.7792	5.390%	2^{nd} bending mode (N-S)
6	8.1623	7.8249	4.133%	axial mode

Table 2: Natural frequencies analysis: frequencies of the main mode shapes of the Mastio tower of the San Felice sul Panaro Fortress. Comparison between *CLOUD2FEM* based finer and coarse model.

Mode shape no.2 and no.5 of the finer model (Figure 5(a) and Figure 5(c), respectively) and of the coarse model (Figure 5(b) and Figure 5(d), respectively): colors are associated to the magnitude of the computed amplitude (normalized). Mode shapes are in very good agreement.



(a) Mode no. 2. Finer mesh freq. =1.9276 Hz

(b) Mode no. 2. Coarse mesh freq. =1.8705 Hz

(c) Mode no. 5. Finer mesh freq. =7.1654 Hz

(d) Mode no. 5. Coarse mesh freq. =6.7792 Hz

Figure 5: Mode no. 2 and mode no.5. Displacements magntiude.

Consequently, the coarse model is structurally equivalent to the finer one. In this way, it can be utilized to carry out advanced numerical analyses, such as nonlinear analyses, with a reasonable computational cost.

4 CONCLUSIONS

The mesh generation approach introduces a simplified description of the geometry and leads to a FE model able to precisely catch the geometry features and the corresponding mass properties. The mechanical properties are defined by a punctual characterization, which leads to a

very accurate description of the structure since each voxel can be automatically associated with a particular property definition. The most relevant feature of the proposed method is the possibility to directly exploit detailed 3D point clouds surveyed on historical buildings for structural purposes. As a result, a large reduction in required time in comparison to CAD-based modeling procedures is achieved.

REFERENCES

[1] Milani, G.; Casolo, S.; Naliato, A.; Tralli, A. Seismic Assessment of a Medieval Masonry Tower in Northern Italy by Limit, Nonlinear Static, and Full Dynamic Analyses. *International Journal of Architectural Heritage* 2012, *6(5)*, 489-524.

[2] Minghini, F.; Milani, G.; Tralli, A. Seismic risk assessment of a 50m high masonry chimney using advanced analysis techniques. *Engineering Structures* 2014, *69*, 255-270.

[3] Roca, P.; Cervera, M.; Gariup, G.; Pelá, L. Structural analysis of masonry historical construction. Classical and advanced approaches. *Archives of Computational Methods in Engineering* 2010, *17(3)*, 299-325.

[4] Zvietcovich, F.; Castaneda, B.; Perucchio, R. 3D solid model updating of complex ancient monumental structures based on local geometrical meshes. *Digital Applications in Archaeology and Cultural Heritage* 2015, *2*, 12-27.

[5] Riveiro, B.; Caamaño, J. C.; Arias, P.; Sanz, E. Photogrammetric 3D modelling and mechanical analysis of masonry arches: An approach based on a discontinuous model of voussoirs. *Automation in Construction*, 2011, *20(4)*, 380-388.

[6] Guarnieri, A.; Pirotti, F.; Pontin, M.; Vettore, A. Combined three-dimensional surveying techniques for structural analysis applications. In Proceedings of the Third IAG, 12th FIG symposium, 2006.

[7] Guarnieri, A.; Milan, N.; Vettore, A. Monitoring Of Complex Structure For Structural Control Using Terrestrial Laser Scanning (Tls) And Photogrammetry. *International Journal of Architectural Heritage* 2013, *7*, 54-67.

[8] Truong-Hong, L.; Laefer, D. Octree-based, automatic building facade generation from LiDAR data. *Computer-Aided Design* 2014, *53*, 46-61.

[9] Truong-Hong, L.; Laefer, D. Validating Computational Models from Laser Scanning Data for Historic Facades. *Journal of Testing and Evaluation* 2013, *41*, 481-496.

[10] Hinks, T.; Carr, H.; Truong-Hong, L.; et al. Point Cloud Data Conversion into Solid Models via Point-Based Voxelization. *Journal of Surveying Engineering* 2013, *139*, 72-83.

[11] Sienz, J.; Szarvasy, I.; Hinton, E.; Andrade, M. L. Computational modelling of 3D objects by using fitting techniques and subsequent mesh generation. *Computers & Structures* 2000, *78(1-3)*, 397-413.

[12] Borri, A.; Grazini, A. Diagnostic analysis of the lesions and stability of Michelangelos David. *Journal of Cultural Heritage* 2006, *7*, 273-285.

[13] Castellazzi, G.; D'Altri, A.M.; Bitelli, G.; Selvaggi, I.; Lambertini, A. From Laser Scanning to Finite Element Analysis of Complex Buildings by Using a Semi-Automatic Procedure. *Sensors* 2015, *15*, 18360-18380.

[14] Cattari, S.; Degli Abbati, S.; Ferretti, D.; Lagomarsino, S.; Ottonelli, D.; Tralli, A. Damage assessment of fortresses after the 2012 Emilia earthquake (Italy). *Bulletin of Earthquake Engineering* 2014, *12*, 2333-2365.

[15] Castellazzi, G.; D'Altri, A.M.; de Miranda, S.; Magagnini, S.; Tralli, A. On the seismic behavior of the main tower of San Felice sul Panaro (Italy) fortress. In Proceeding of ICCMSE 2016, March 17-20, 2016, Athens, Greece.

[16] Bitelli, G.; Castellazzi, G.; D'Altri, A.M.; de Miranda, S.; Lambertini, A.; Selvaggi, I. Automated voxel model from point clouds for structural analysis of cultural heritage. In Proceeding of XXIII ISPRS Congress 2016, July 12-19, 2016, Prague, Czech Republic.

[17] Taubin, G. Curve and surface smoothing without shrinkage. In *ICCV '95*, Proceedings of the Fifth International Conference on Computer Vision, Cambridge, MA, USA, June 20-23th, 1995; IEEE Computer Society Washington, DC, USA 1995.

[18] Castellazzi, G. On the performances of parametric finite elements when geometry distortions occur. *Finite Elements in Analysis and Design* 2011, *47(12)*, 1306-1314.

[19] Castellazzi, G.; de Miranda, S.; Ubertini, F. Patch based stress recovery for plate structures. *Computational Mechanics* 2011, *47(4)* , 379-394.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

HIGH-ORDER CURVED MESH GENERATION BY USING A FINE LINEAR TARGET MESH

Verena S. Schmid[1,2], Hadrien Beriot[1], Onur Atak[1], and Gwenael Gabard[2]

[1]Siemens Industry Software, Simulation and Test Solutions
Interleuvenlaan 68,B-3001,Leuven, Belgium
e-mail: {hadrien.beriot, onur.atak}@siemens.com

[2] Institute of Sound and Vibration Research, University of Southampton
SO17 1BJ, UK
e-mail: v.s.schmid@soton.ac.uk, gabard@southampton.ac.uk

Keywords: Mesh curving, High-order, Lagrange interpolation, Hierarchic shape functions, Projection based interpolation.

Abstract.

This paper examines different mesh boundary curving algorithms in the particular case where no exact geometric description is available. The starting point for the curving is typically a coarse linear mesh, whereas the target geometry is represented by a refined linear mesh. Both can be generated using a classical linear mesh generator.

Two different boundary curving algorithms are examined. The first algorithm is based on Lagrange nodal high-order polynomial interpolation where the interpolation nodes are iteratively moved towards the target curve. Relocation steps are included in each iteration to approximately preserve the original node spacing. The second algorithm is based on hierarchic modal shape functions. In a reference frame, projection based interpolation is applied that minimizes the distance between the interpolating function and the local target function in the H^1-seminorm.

The performance and accuracy of the two methods are evaluated and compared. Thereby, the area between the target and the approximating curve is used as error measure.

In general, both methods exhibit similar levels of performance. A lower bound on the accuracy is observed that depends on the level of refinement of the target mesh. Differences lie in the applicability of the two methods. For the method based on modal interpolation, several initial requirements have to be fulfilled. The method based on nodal interpolation on the other hand is simpler but less robust. Overall, the modal interpolation approach is preferable, where applicable.

1 INTRODUCTION

High-order unstructured methods are gaining popularity for solving large scale engineering problems. These methods generally lead to increased accuracy and lower computational costs when compared to the commonly used low-order methods. While low-order methods typically rely on planar-faced elements, high-order methods call for the introduction of high-order curvilinear finite element meshes, in which the faces and edges of the elements are curved to align with the underlying geometry. The need for an accurate geometric description when employing high-order approximations was shown in the literature both analytically [1] and numerically [2].

Mesh curving algorithms are usually based on a-posteriori curving. A linear mesh is first generated, taking advantage of the robustness of the existing mesh generation algorithms. In a second step these elements are curved to better describe the geometry of the domain, e.g. [3, 4].

If the exact geometry is known, e.g. in the form of a parametrization, this curving step is relatively straightforward. For a nodal based polynomial representation of the geometry, the high-order nodes are defined in the reference space and then mapped onto the exact geometry. The node positions in the reference space influence the resulting curve and can be used to optimize the approximation or to prevent element tangling already during the curving process [3].

Surface reconstruction methods are applied where no or only little additional information is given about the exact geometry. Such problems are for instance encountered in medical imaging, where the definition of the computational domain is given exclusively by scattered data, obtained from data scanning. Examples for surface reconstruction methods are the *Weighted Averaging of Local Fittings (WALF)* and the *Continuous Moving Frame (CMF)* algorithms [5] or the piecewise quadratic mesh extrapolation based on local approximations of the curve Hessian [6].

In this paper an intermediate case is considered. An exact representation is not available, but instead a fine linear boundary mesh is considered, as could be used for a conventional low order simulation. This fine linear mesh forms the target of the curving process. A similar approach can be found in [7] where the fine representation is obtained from scattered scanning data. It is improved by a surface reconstruction based on spherigon patches [8]. The curved mesh is formed of Bézier curves and patches. Here, it is assumed that the target mesh is fine enough such that no further surface reconstruction is required. Two curving algorithms are presented. For both, the curved mesh is represented by piecewise polynomial parametrizations. For the first algorithm, nodal basis functions are used and for the second, hierarchic, modal basis functions.

Note that this paper focuses only on the curving of the boundary elements. Post-processing steps such as element untangling may still be required once the full 2D-mesh is considered.

2 METHODS

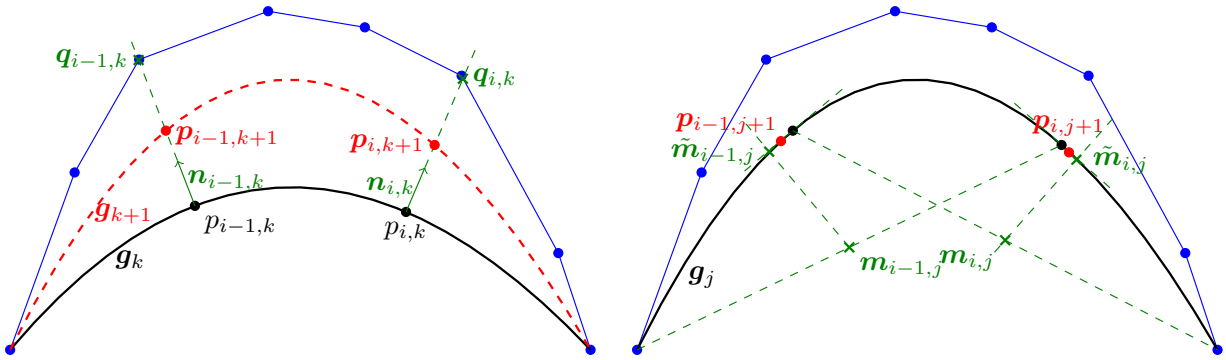
2.1 Notations

In the following, the fine target mesh is named *model mesh* or *Mmesh*. It is denoted $\mathcal{M} := (\mathcal{V}_\mathcal{M}, \mathcal{E}_\mathcal{M})$ with $\mathcal{V}_\mathcal{M} := \left\{\boldsymbol{v}_{M,1}, \boldsymbol{v}_{M,2}, \ldots, \boldsymbol{v}_{M,N^v_\mathcal{M}}\right\}$, $\boldsymbol{v}_{M,i} \in \mathbb{R}^2$ the $N^v_\mathcal{M}$ mesh vertices and $\mathcal{E}_\mathcal{M} := \left\{e_{M,1}, \ldots, e_{M,N^e_\mathcal{M}}\right\}$ the edges that connect them. The number of Mmesh edges is denoted $N^e_\mathcal{M}$.

The coarse mesh that is curved during the algorithm is named *computational mesh* or *Cmesh*. It is denoted $\mathcal{C}$ and the notations $\mathcal{V}_\mathcal{C}$, $\mathcal{E}_\mathcal{C}$, and $N^e_\mathcal{C}$ are defined analogously to $\mathcal{V}_\mathcal{M}$, $\mathcal{E}_\mathcal{M}$, and $N^e_\mathcal{M}$.

The Cmesh geometric order obtained from the mesh curving procedure is denoted q.

2.2 Nodal approach



(a) Curving step (b) Correction step

Figure 1: Schematics of the two different iteration steps in the nodal method. In both figures, the target curve $\mathcal{M}$ is shown in blue, the current $\mathcal{C}$ element at the beginning of the iteration step in black, and the result of the iteration step in red. Helplines and intersection points are drawn in green.

For the nodal method, $\mathcal{C}$ is represented by an elementwise polynomial mapping from the reference element $[-1, 1]$ to the physical space. Lagrange polynomials with equidistant nodes in the reference space form the basis of the mapping. In the physical space, corresponding nodes have to be found in order to elementwise curve the Cmesh.

As a first step, $q-1$ high-order interpolation nodes are placed equidistantly on each edge of the linear Cmesh, forming a polynomial $\boldsymbol{g}_k : [-1, 1] \rightarrow R^2$ of order q . These nodes are iteratively moved towards the Mmesh during the curving process. The iteration consists of two operations which are implemented in a nested iteration:

1. In an outer loop of n_{iter} iteration steps, the Cmesh nodes are progressively moved towards the target geometry, following the local Cmesh normal direction.

   For the calculation of the new node positions, each node $\boldsymbol{p}_{i,k}$ is treated individually. The normal $\boldsymbol{n}_{i,k}$ of $\boldsymbol{g}_k$ at $\boldsymbol{p}_{i,k}$ is first evaluated. To find the distance to $\mathcal{M}$, the closest intersection point $\boldsymbol{q}_{i,k}$ between $\mathcal{M}$ and the line $\boldsymbol{s}_{i,k}(\alpha) := \boldsymbol{p}_{i,k} + \alpha \boldsymbol{n}_{i,k}$ is determined. The updated node position is then calculated as

   $$\boldsymbol{p}_{i,k+1} \quad = \boldsymbol{p}_{i,k} + \sigma_k \left(\boldsymbol{q}_{i,k} - \boldsymbol{p}_{i,k} \right), \tag{1}$$

   with $\sigma_k = \frac{k}{n_{\text{iter}}}$. An example of a mesh curving iteration step is illustrated in Figure 1(a)).

2. After each mesh curving iteration, several correction steps are applied. The spacing between points is readjusted, aiming for an approximately equidistant spacing in the physical space (see Figure 1(b)).

   The tangent of $\boldsymbol{g}_j$ in a node $\boldsymbol{p}_{i,j}$ is denoted $\boldsymbol{t}_{i,j}$. For each point $\boldsymbol{p}_{i,j}$, a midpoint is computed between its two neighboring points and denoted $\boldsymbol{m}_{i,j} := \frac{\boldsymbol{p}_{i-1,j}+\boldsymbol{p}_{i+1,j}}{2}$. The midpoint is then projected onto the tangent line $\boldsymbol{t}_{i,j}$. The resulting point $\tilde{\boldsymbol{m}}_{i,j}$ represents the target point towards which the node $\boldsymbol{p}_{i,j}$ is moved. All nodes are updated at the same time. In order to reduce the likelihood to overshoot, the node is updated by

   $$\boldsymbol{p}_{i,j+1} = \boldsymbol{p}_{i,j} + \tau \left(\boldsymbol{p}_{i,j} - \tilde{\boldsymbol{m}}_{i,j} \right), \tag{2}$$

   with $\tau \in (0, 1)$. In this study, $\tau = 0.4$ is chosen.

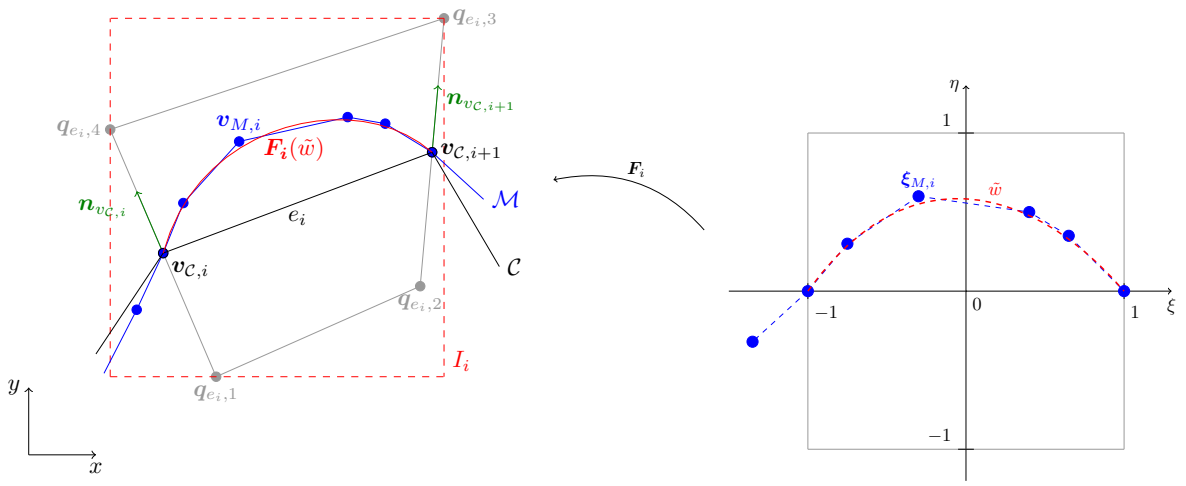


Figure 2: Schematic of the modal method. On the left, the physical space is shown with $\mathcal{M}$ in blue, $\mathcal{C}$ with edge e_i in black, the quadrangle spanned by e_i in grey, and the bounding interval I_i in red. On the right, the reference space is shown with the reference quadrangle in grey, the local target function w_i in blue and the polynomial interpolation $\tilde{w}_i$ in red.

After the last iteration step, one additional mesh deformation step is performed with $\sigma_{n_{\text{iter}}+1} = 1$ to move the nodes exactly onto the Mmesh.

The number of iteration steps for the two iteration loops are increased adaptively until a convergence of the method is observed. For both iteration loops the number of iteration steps are increased simultaneously, but those of the outer loop is increased in bigger steps. As criterion for convergence, the distance between nodes obtained with the current number of iteration steps and those obtained with the previous number of iteration steps is evaluated. Once the maximal distance is smaller than a predefined nodal tolerance of 10^{-14}, it is decided that the curving method converged. The algorithm breaks down if a predefined upper limit of iteration numbers is hit without reaching convergence. In this study, the limit is set to 200 iteration steps for the outer loop.

2.3 Modal approach

The modal method is based on a polynomial fitting in a reference space, using a hierarchic, modal polynomial basis. A schematic of the method is provided in Figure 2.

The algorithm begins as follows. In the physical space, the vertices of each Cmesh element $e_i \in \mathcal{E}_\mathcal{C}$ are extruded by an extrusion length λ_i in positive and negative edge normal direction. Connecting these extruded nodes, a quadrangle element $\mathcal{Q}_{e_i} = (\boldsymbol{q}_{e_i,1}, \boldsymbol{q}_{e_i,2}, \boldsymbol{q}_{e_i,3}, \boldsymbol{q}_{e_i,4})$ is constructed around e_i. Note that the union $\bigcup_{e_i \in \mathcal{E}_\mathcal{C}} \mathcal{Q}_{e_i}$ forms a neighborhood of $\mathcal{C}$. Furthermore, it is required that $\mathcal{M} \in \bigcup_{e_i \in \mathcal{E}_\mathcal{C}} \mathcal{Q}_{e_i}$ and $\mathcal{Q}_{e_i} \cap \mathcal{Q}_{e_j} = \emptyset$ for $i \neq j$. For each quadrangle $\mathcal{Q}_{e_i}$ the mapping $\boldsymbol{F}_i : \mathbb{R}^2 \to \mathbb{R}^2$, $\mathcal{Q}_{ref} \mapsto \mathcal{Q}_{e_i}$ is defined by

$$\begin{aligned} \boldsymbol{F}_i(\xi,\eta) = \; & \tfrac{1}{4}\left((\xi-1)(\eta-1)\boldsymbol{q}_{e_i,1} - (\xi+1)(\eta-1)\boldsymbol{q}_{e_i,2}\right. \\ & \left. +(\xi+1)(\eta+1)\boldsymbol{q}_{e_i,3} - (\xi-1)(\eta+1)\boldsymbol{q}_{e_i,4}\right), \end{aligned} \tag{3}$$

where $\mathcal{Q}_{ref}$ designates the reference quadrangle element. For all $\boldsymbol{v}_{M,i} \in \mathcal{V}_\mathcal{M}$ that lie within the bounding interval

$$I_i := \left[\min_j \left(\boldsymbol{q}_{e_i,j}\right)_x, \max_j \left(\boldsymbol{q}_{e_i,j}\right)_x\right] \times \left[\min_j \left(\boldsymbol{q}_{e_i,j}\right)_y, \max_j \left(\boldsymbol{q}_{e_i,j}\right)_y\right], \tag{4}$$

the corresponding points in the reference space

$$\Xi_{I_i} := \{\boldsymbol{\xi} = (\xi, \eta) \,|\boldsymbol{F}_i(\boldsymbol{\xi}) \in \mathcal{V}_{\mathcal{M}} \cap I_i\} \tag{5}$$

are determined. Since $\boldsymbol{F}_i$ is generally not linear, no closed-form expression could be found for the inverse mapping. Instead, the objective function $g(\xi, \eta) := |\boldsymbol{x} - \boldsymbol{F}_i(\xi, \eta)|$ is minimized to find the inverse mapping of the fixed point $\boldsymbol{x}$. The points in Ξ_{I_i} are then piecewise linearly interpolated by $w_i : \mathbb{R} \mapsto \mathbb{R}, w_i(\xi_{I_i,j}) = \eta_{I_i,j}$ for $(\xi_{I_i,j}, \eta_{I_i,j}) \in \Xi_{I_i}$. Note that this linear interpolation forms an approximation. In general, the image $\boldsymbol{F}_i(w_i)$ of w_i does not lie on $\mathcal{M}$ because of the non-linearity of $\boldsymbol{F}_i$.

The function w_i is approximated by a polynomial interpolation

$$\tilde{w}_i = \sum_{j=0}^{q} \alpha_{i,j} l_j. \tag{6}$$

The basis of the interpolation consists of Lobatto shape functions [9]

$$\begin{aligned} l_0(\xi) &= \tfrac{1-\xi}{2}, \\ l_1(\xi) &= \tfrac{\xi+1}{2}, \quad \text{and} \\ l_k(\xi) &= \tfrac{1}{\|L_{k-1}\|_2} \int_{-1}^{\xi} L_{k-1}(x)\,\mathrm{d}x, \ \text{ for } 2 \le k,\ \xi \in [-1, 1]\,, \end{aligned} \tag{7}$$

with $L_i(\xi)$ the i^{th} Legendre polynomial. From the orthogonality of the Legendre polynomials in L^2, the Lobatto shape functions inherit the property

$$\int_{-1}^{1} l_i'(\xi) l_j'(\xi)\,\mathrm{d}\xi = \delta_{ij}, \ \text{with } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}. \tag{8}$$

This property can be used to obtain the coefficients $\alpha_{i,j}$, $j = 2, \ldots, q$ in (6) that minimize the distance between the $\tilde{w}_i$ and w_i in the H^1-seminorm [9]

$$\left|w_i - \tilde{w}_i^v - \tilde{w}_i^b\right|_{H^1} \to \min\,. \tag{9}$$

The function $\tilde{w}_i$ is thereby splitted into the linear part $\tilde{w}_i^v = \sum_{j=0}^{1} \alpha_{i,j} l_j$ and the high-order part $\tilde{w}_i^b = \sum_{j=2}^{q} \alpha_{i,j} l_j$. Note that $\alpha_{i,1}$ and $\alpha_{i,2}$ are uniquely defined by $w_i(-1)$ and $w_i(1)$. Therefore, the optimization only considers $\alpha_{i,2}, \ldots, \alpha_{i,q}$.

The minimization (9) is equivalent to

$$\int_{-1}^{1} \left(w_i - \tilde{w}_i^v - \tilde{w}_i^b\right)' l_j'\,\mathrm{d}\xi = 0, \ \text{for } j = 2, \ldots, q, \tag{10}$$

which can be reformulated as

$$\alpha_{i,j} = \int_{-1}^{1} w_i'(\xi) l_j'(\xi)\,\mathrm{d}\xi. \tag{11}$$

This approach is referred to as projection based interpolation [9]. Lobatto shape functions are also hierarchic, meaning that $\mathcal{B}^{n-1} \subset \mathcal{B}^n$, with $\mathcal{B}^n$ the Lobatto basis of order n. This is beneficial for adaptive schemes, because it implies that a change in the order q does not require a recalculation of the shape functions l_j or coefficients α_j.

3 NUMERICAL TESTS

3.1 Geometry and Meshes

The test case geometry is based on an ellipse with a high aspect ratio. An additional cosine term is included to introduce bumps along the curve:

$$\begin{pmatrix} x(\theta) \\ y(\theta) \end{pmatrix} = \begin{pmatrix} (r_x + h_b \cos(\theta n_b)) \cos(\theta) \\ (r_y + h_b \cos(\theta n_b)) \sin(\theta) \end{pmatrix} \quad \text{for } \theta \in [0, 2\pi] \,, \tag{12}$$

where n_b is the number of bumps and h_b controls their height. The parameters were set to $r_x = 0.5$m, $r_y = 5$m, $h_b = 0.1$m, and $n_b = 10$. A plot of the geometry can be seen in Figure 3(a).

In a first step, n_v equidistant nodes $\tilde{\theta}_i \in [0, \pi]$ are defined in the reference space. In a second step, perturbed nodes are introduced, $\theta_i = \tilde{\theta}_i + \gamma_i$, with γ_i uniformly distributed pseudo random values in the interval $[-\frac{1}{3n_v}, \frac{1}{3n_v}]$ and $\gamma_0 = \gamma_{n_v} = 0$.

As a last step before the actual curving, the Cmesh vertices are snapped onto the closest Mmesh vertices. This ensures that the assumption $\mathcal{V}_\mathcal{C} \subset \mathcal{V}_\mathcal{M}$ holds.

3.2 Evaluation Method

To assess the accuracy of the two algorithms, the area between the final high-order curve and the target curve $\mathcal{M}$ needs to be evaluated. The area between the two curves is divided in small quadrangles $Q_i = (\boldsymbol{q}_{i,1}, \boldsymbol{q}_{i,2}, \boldsymbol{q}_{i,3}, \boldsymbol{q}_{i,4})$. For each quadrangle, the area is calculated with the formula

$$A_{Q_i} = \frac{1}{2} \left| (\boldsymbol{q}_{i,2} - \boldsymbol{q}_{i,4}) \times (\boldsymbol{q}_{i,1} - \boldsymbol{q}_{i,3}) \right| \tag{13}$$

for convex quadrangles. The size of the quadrangles is adapted in each case to ensure an accurate evaluation of the area. Restrictions on the node positions are applied to reduce the number of non-convex quadrangles. The area based error measure is defined as

$$A_Q := \frac{\sum_i A_{Q_i}}{l_\mathcal{M}}, \tag{14}$$

with $l_\mathcal{M}$ the total length of $\mathcal{M}$. This corresponds to the measure suggested in [4].

3.3 Fixed number of Mmesh elements

The first study examines the influence of the interpolation order q and the size of $\mathcal{C}$ on the results of the curving process. The target curve $\mathcal{M}$ is fixed with $N^e_\mathcal{M} = 998$.

The curving procedures are performed for varying $N^e_\mathcal{C} \in \{6, 10, \ldots, 78\}$ and $q \in \{1, \ldots 8\}$ for the nodal approach and $N^e_\mathcal{C} \in \{10, 14, \ldots, 78\}$ and $q \in \{1, \ldots 15\}$ for the modal approach. Figures 3(b)–3(f) provide an example of the initial $\mathcal{C}$ and the final curved mesh with both methods for $N^e_\mathcal{C} = 10$ and $q = 4$,

Note that the different bounds of the parameter show the limitations of the two methods. The nodal curving algorithm failed to converge within iteration boundaries for some cases of $N^e_\mathcal{C}$ for $q = 8$. Calculations for $q = 9$ did not converge for any of the tested $N^e_\mathcal{C}$. The cases that did not converge are not included in the upcoming plots.

The modal curving approach on the other hand could be applied up to $q = 15$, with no indication that even higher orders would not be calculable. The limitation for the modal method

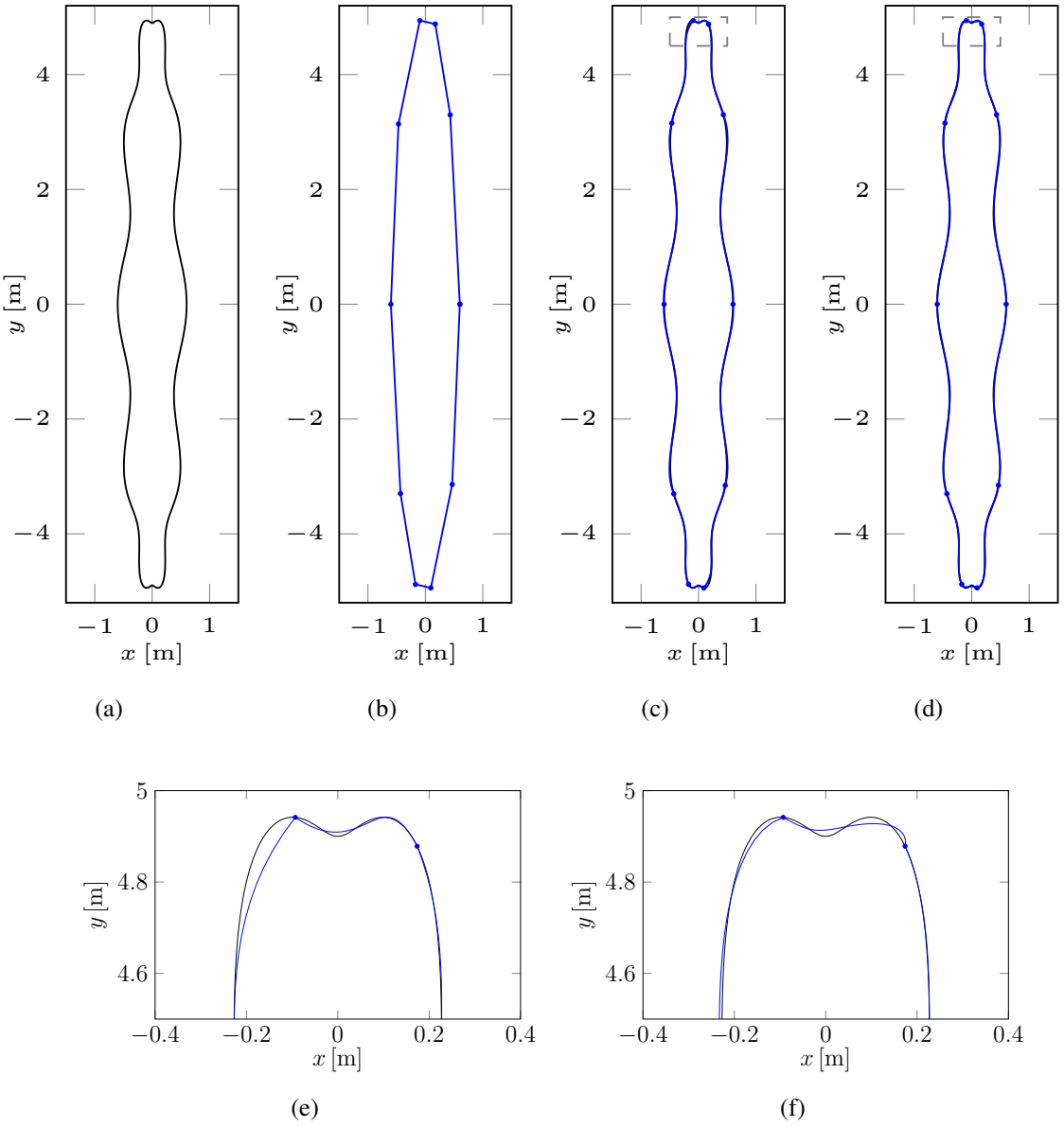
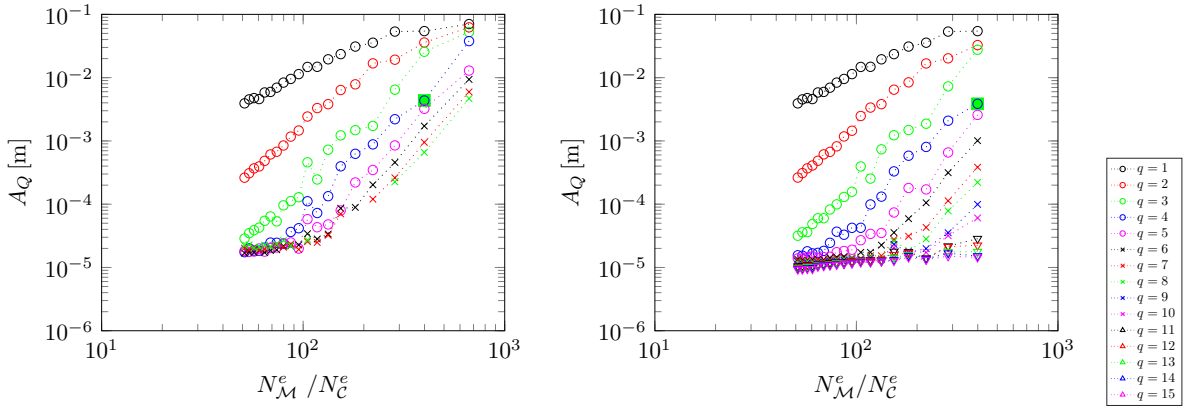
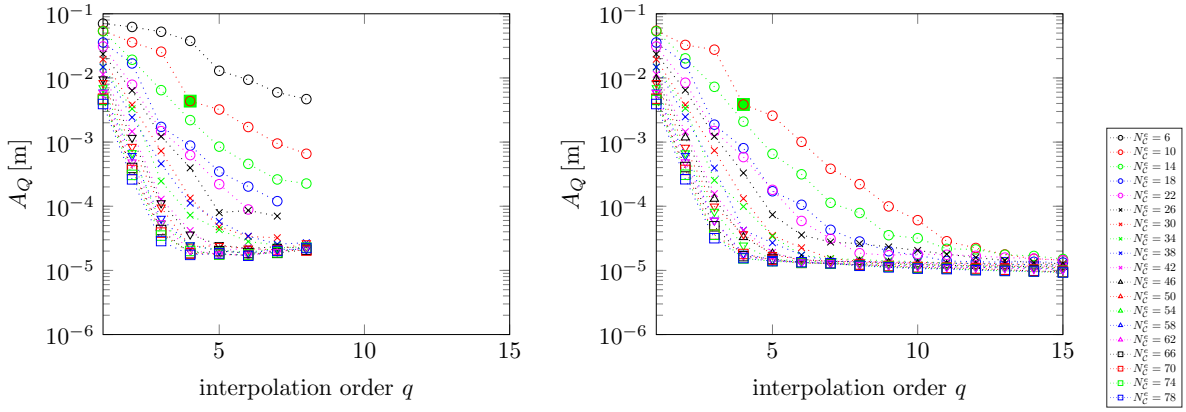


Figure 3: The distorted ellipse test case. (a) geometry, (b) Cmesh with $N_{\mathcal{C}}^{e} = 10$. The resulting curves for $q = 4$ after applying (c) the nodal curving approach or (d) the modal curving approach are shown in blue. The grey dashed boxes mark the zoom-in region that is shown in (e) for the nodal curving approach and in (f) for the modal curving approach.



(a) (b)

Figure 4: $N^e_{\mathcal{M}}/N^e_{\mathcal{C}}$-convergence for (a) the nodal method and (b) the modal method, with fixed $N^e_{\mathcal{M}} = 998$.



(a) (b)

Figure 5: q-convergence for (a) the nodal method and (b) the modal method, with fixed $N^e_{\mathcal{M}} = 998$.

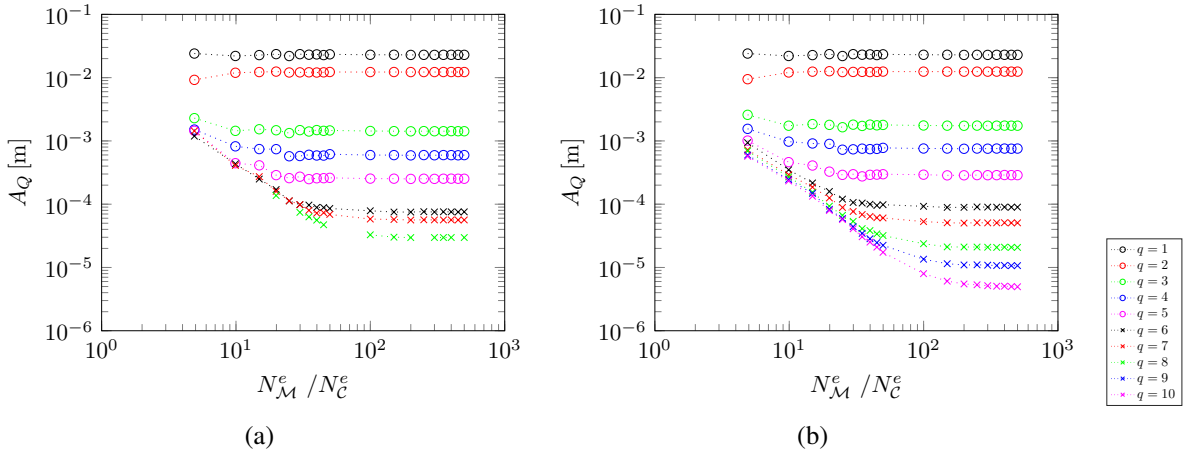
lies in the construction of the quadrangles $\mathcal{Q}_{e_i}$. For the case $N^e_{\mathcal{C}} = 6$, no set of extrusion lengths $\Lambda = \{\lambda_1, \ldots, \lambda_{N^e_{\mathcal{C}}}\}$ could be found such that the requirements $\mathcal{M} \in \bigcup_{e_i \in \mathcal{E}_{\mathcal{C}}} \mathcal{Q}_{e_i}$ and $\mathcal{Q}_{e_i} \cap \mathcal{Q}_{e_j} = \emptyset$ for $i \neq j$ were fulfilled.

In Figure 4, the $N^e_{\mathcal{M}}/N^e_{\mathcal{C}}$ convergence of the area based error measure is shown for different interpolation orders q. Each data point corresponds to one parameter set $(N^e_{\mathcal{C}}, q)$. The example case from Figures 3(b)–3(f) is highlighted by a green square.

The results indicate that both methods provide a similar behavior. For a fixed order q, the $\mathcal{C}$ approximation convergence increases with q following roughly $\mathcal{O}(q)$ up to $q = 3$. For higher orders q, the convergence order lies between 3 and 4. However, all approximations are limited in accuracy, and stagnate around $A_Q = 10^{-5}$.

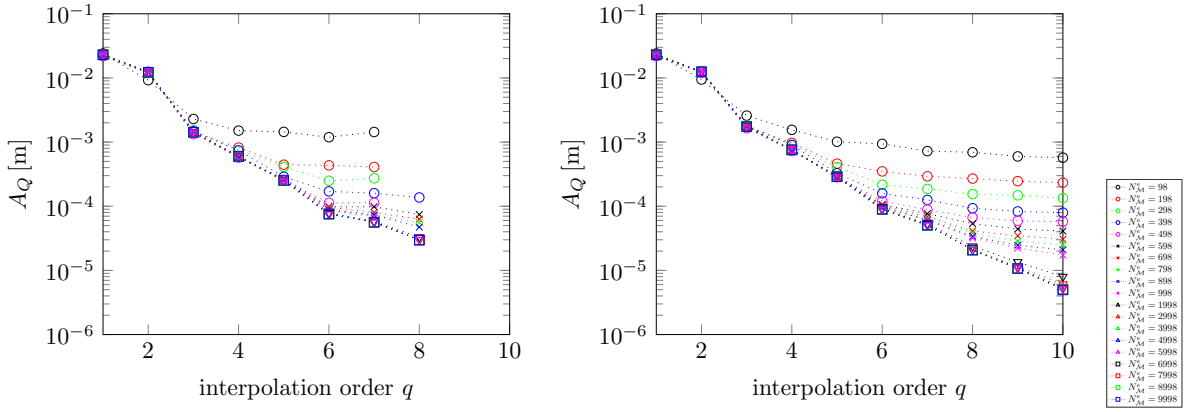
The corresponding q-convergence curves for different $N^e_{\mathcal{C}}$ are shown in Figure 5. As in the previous figure, the example case from Figures 3(b)–3(d) is highlighted by a green square.

With increasing q the error decreases until it stagnates. The observed bound is about the same for both methods. This seems to indicate that the limit in accuracy is caused by the choice of the $\mathcal{M}$ representation, more specifically by the fact that $\mathcal{M}$ is only C^0 continuous. This typically limits the accuracy up to which $\mathcal{M}$ can be polynomially interpolated.



(a) (b)

Figure 6: $^{N^e_{\mathcal{M}}}/_{N^e_{\mathcal{C}}}$-convergence for (a) the nodal method and (b) the modal method, with fixed $N^e_{\mathcal{C}} = 20$.



(a) (b)

Figure 7: q-convergence for (a) the nodal method and (b) the modal method, with fixed $N^e_{\mathcal{C}} = 20$.

3.4 Fixed number of Cmesh elements

A second study is performed in order to better understand the influence of $N^e_{\mathcal{M}}$ on the error of the approximation. The size of the Cmesh is now fixed with $N^e_{\mathcal{C}} = 20$, and the target Mmesh is varied. The parameter sets are $q \in \{1, \dots, 8\}$ for the nodal and $q \in \{1, \dots, 10\}$ for the modal algorithm, and $N^e_{\mathcal{M}} \in \{98, 198, \dots, 998, 1998, \dots, 9998\}$ for both methods. Again, the lower q limit for the nodal method as well as missing data points are caused by convergence issues.

Figure 6 shows the $^{N^e_{\mathcal{M}}}/_{N^e_{\mathcal{C}}}$-convergence for different interpolation orders q. Expectedly, as $\mathcal{M}$ is refined, the error first decreases and then stagnates. The stagnation levels decrease as q increases. It shows that with a fixed order q, $\mathcal{M}$ refinement is useful only up to a certain point.

Figure 7 shows the corresponding q-convergence curves for different $N^e_{\mathcal{M}}$. For low interpolation orders q, the $\mathcal{M}$-refinement has no impact on the performance of the mesh curving algorithm. But for larger values of q, the error stagnates for the coarser Mmeshes. The lower $N^e_{\mathcal{M}}$, the higher the stagnation level. This confirms the assumption from above that the stagnation is introduced by the C^0 continuity of $\mathcal{M}$. A finer $\mathcal{M}$ provides a better representation of the exact geometry and is in a colloquial sense smoother than a coarse $\mathcal{M}$. This highlights the fact that the order of approximation q should always be consistent with the quality of the Mmesh. After a certain point, since the information at hand on the target geometry is not satisfactory

enough, increasing the order and/or refining the Cmesh does not allow to improve geometric representation.

4 DISCUSSION AND CONCLUSIONS

In this paper, high-order mesh curving algorithms were considered in the case of limited knowledge of the original geometry, more precisely in the case where the original geometry is only given in the form of a refined linear mesh. Two approaches were proposed and tested on a distorted ellipse. The results for both methods indicate that, besides the number of edges $N_{\mathcal{C}}^e$ and the interpolation order q of the curved mesh, also the level of refinement of the target curve $\mathcal{M}$ provides a limit to the accuracy of the methods. The bound introduced by $N_{\mathcal{M}}^e$ is explained by the C^0-continuity of $\mathcal{M}$ that, even though it never becomes C^1 continuous, becomes smoother in a colloquial sense as $N_{\mathcal{M}}^e$ is increased.

In general, the two methods provided similar results, but both showed disadvantages as compared to the other. For the nodal curving algorithm, convergence issues were observed for high q values. Other node spacings such as Legendre or Chebyshev nodes are currently considered to test if these would provide better convergence behavior. The modal approach, on the other hand, is not applicable in some cases, due to the difficulty to construct a consistent reference frame. This might be overcome by small changes in $\mathcal{C}$ such as shifting of vertices or refinement of the affected elements. Concerning possible applications to more complex geometries as well as the extension to 3D, the modal curving algorithm seems more promising.

In general, the two discussed curving methods consider only boundary curves. Problems such as element tangling that arise when domain elements are included, were not examined in this study. Furthermore, geometric feature detection is required before the methods could be applied to real life geometries.

5 ACKNOWLEDGMENT

This work was performed as part of the CRANE project (Community and Ramp Aircraft NoisE, www.crane-eid.eu, GA: 606844) funded by the European Union under the Framework Programme 7. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

REFERENCES

[1] S. Dey, M.S. Shephard, J.E. Flaherty, Geometry representation issues associated with p-version finite element computations. *Computer Methods in Applied Mechanics and Engineering*, **150(1)**, 39–55, 1997.

[2] X.J. Luo, M.A. Shephard, J.-F. Remacle, The influence of geometric approximation on the accuracy of high order methods. *Rensselaer SCOREC report*, **1**, 2001.

[3] Z. Xie, R. Sevilla, O. Hassan, K. Morgen, The generation of arbitrary order curved meshes for 3D finite element analysis. *Computational Mechanics*, **51(3)**,361-374, 2013.

[4] J.-F. Remacle, J. Lambrechts, C. Geuzaine, and T. Toulorge, Optimizing the geometrical accuracy of 2D curvilinear meshes. *Procedia Engineering*, **82**, 228–239, 2014

[5] X. Jiao, D. Wang, Reconstructing high-order surfaces for meshing. *Engineering with Computers*, **28(4)**, 361–373, 2012.

[6] Y.V. Vassilevski, V.G. Dyadechko, K.N. Lipnikov, Hessian-based anisotropic mesh adaptation in domains with discrete boundaries. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **20(4)**, 391–402, 2005.

[7] K. Bock, J. Stiller, Generation of High-Order Polynomial Patches from Scattered Data.Spectral and High Order *Methods for Partial Differential Equations - ICOSAHOM 2012, Lecture Notes in Computational Science and Engineering*, **95**, 157–167, 2014.

[8] P. Volino, N.M. Thalmann, The SPHERIGON: a simple polygon patch for smoothing quickly your polygonal meshes. *Computer Animation 98. Proceedings*, 72–78, 1998.

[9] P. Šolín, K. Segeth, I. Doležel, *Higher order finite element methods*. Chapman & Hall / CRC, ISBN 1-58488-438-X, 2004.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

AUTOMATIC GENERATION OF 3D UNSTRUCTURED HIGH-ORDER CURVILINEAR MESHES

Michael Turner[1], David Moxey[1], Spencer J. Sherwin[1] and Joaquim Peiró[1]

[1]Department of Aeronautics, Imperial College London
Imperial College London, London, SW7 2AZ, UK
e-mail: m.turner14@imperial.ac.uk, d.moxey@imperial.ac.uk, j.peiro@imperial.ac.uk,
s.sherwin@imperial.ac.uk

Keywords: High-order, Mesh, Generation, Curvilinear, Unstructured 3D

Abstract. *The generation of suitable, good quality high-order meshes is a significant obstacle in the academic and industrial uptake of high-order CFD methods. These methods have a number of favourable characteristics such as low dispersion and dissipation and higher levels of numerical accuracy than their low-order counterparts, however the methods are highly susceptible to inaccuracies caused by low quality meshes. These meshes require significant curvature to accuratly describe the geometric surfaces, which presents a number of difficult challenges in their generation. As yet, research into the field has produced a number of interesting technologies that go some way towards achieving this goal, but are yet to provide a complete system that can systematically produce curved high-order meshes for arbitrary geometries for CFD analysis. This paper presents our efforts in that direction and introduces an open-source high-order mesh generator, NekMesh, which has been created to bring high-order meshing technologies into one coherent pipeline which aims to produce 3D high-order curvilinear meshes from CAD geometries in a robust and systematic way.*

1 INTRODUCTION

Unstructured low-order (linear) mesh generation of complex curvilinear domains has reached a level of sophistication and robustness that has seen the computational analysis of fluid flow become common place within numerous industrial applications. However, these linear meshes and their corresponding solvers are often limited to second-order accuracy. Recently, there has been an increased interest in CFD techniques which utilise high-order approximations as an alternative option to increase numerical accuracy within the bounds of reasonable computational cost. Furthermore, these methods can be used on complex domains with unstructured meshes [1] and exhibit higher rates of convergence on properly constructed meshes [2]. As such, high-order finite element analysis of fluid flow has been highlighted as a potential route for advancing CFD towards higher fidelity methods [3].

The development of high-order finite element solvers for unstructured meshes has allowed the development of high-order fluid flow solvers for a range of applications. These solvers have matured into a well-developed set of tools, such as, for example, the *Nektar++* spectral/hp element framework [4]. However, a significant bottleneck in the high-order CFD process is the generation of unstructured high-order meshes [3]. At present, this is one of the key factors preventing the wider uptake of high-order methods for CFD within academia and especially industry [1]. While unstructured high-order mesh generation is not a new area of academic research, the documented examples within academia seldom present the full and robust mesh generation frameworks which are required for non-expert users. The theoretical guarantees of mesh validity that exist for some methods of linear meshes (e.g. Delaunay [5]) do not exist for high-order meshes. This, among other reasons, makes high-order mesh generation a significantly tougher challenge than linear mesh generation.

High-order mesh generators are yet to reach a level of maturity where they can be used by anyone except expert users in a robust and reliable fashion. They have been shown to create valid suitable meshes on complex geometries [6, 7, 8]. However the issue of robustness is yet to be fully addressed. No research into high-order mesh generation has presented a full methodology which demonstrates the ability to systematically generate meshes on arbitrary geometries of interest to industry. In addition to this, the ability for a non-expert user to use the tools, input their own geometries and get a valid mesh without issues is rarely mentioned. Also the applicability of the meshes which can be created, with respect to the solvers and solutions which will be used, is not usually documented in the literature of high-order mesh generation. Before any form of industrial uptake can be truly considered, robustness must be addressed within all high-order mesh generation processes.

Currently, the standard approach adopted by almost all high-order meshing techniques is the *a posteriori* modification of a valid unstructured linear mesh [2, 6, 7, 8, 9, 10, 11, 12, 13]. Under this principle, the faces and edges lying on the surface of the geometry are curved by adding high-order nodes which also lie on the geometry surface. The issues with this approach are well documented [9, 14], as this simple idea presents a key obstacle to high-order mesh generation that must be overcome: in the case of high geometric curvature the curving process can frequently create elements which self-intersect or have near-tangent vertices, making the mesh unsuitable for computational use. The more recent research into high-order mesh generation focuses on the correction of these invalid elements and only for the correction stage. Little consideration is given to the meshing process as a whole. This work can be categorised into two types: those which use PDE solvers to prescribe some deformation onto the mesh nodes, which makes the mesh valid through the use of non-linear elasticity [15], linear elasticity [13],

thermo-linear elasticity [16] or the Winslow equations [7] and those which use numerical optimisation of quality based functionals [8, 6]. While these techniques may well be robust and highly capable, none present a high-order mesh generation framework geared towards end users that are looking to achieve high-order CFD in viscous simulations. The reality of CFD analysis is that, despite how critical meshes are to the resulting computation, mesh generation should ideally be mostly transparent and automatic for the end user. In linear mesh generation, efforts have been made to ensure the process as painless and quick as possible, but in high-order meshing the experience and technical ability of the user has never addressed. In this current situation, industrial uptake of high-order methods will unlikely be achieved in the near future because these users cannot produce the meshes easily and therefore find high-order methods quite inaccessible.

The purpose of this paper is to present our initial efforts to address the challenge of automatic mesh generation. We describe the philosophy behind the methods and the work carried out thus far on a new open-source program for generating 3D unstructured high-order meshes. Our goal is to use them for the CFD simulation of compressible, incompressible, viscous and inviscid flows, and to produce a tool with as much automation and robustness as possible, that will make it accessible to non-expert users. By automatic, we mean that the meshing process should be able to go from a CAD definition of the domain to the final valid high-order mesh, without requiring anything but the most basic of interactions with the user. This means that the program, as well as having the ability to generate high-order meshes needs the ability to handle CAD, generate coarse linear meshes and perform invalid element correction as part of one process. By robust, we mean that the system should work for as wide a range of geometries as possible and be able to produce valid high-order meshes with a good rate of success. This program is called *NekMesh*, which exists within the open-source high-order framework *Nektar++* and has entered a stage of beta testing before more development is conducted and more robust and capable technologies are added in the near future.

This paper begins by describing the program and outlines the core ideas behind the processes used currently at each stage of the pipeline. It also presents a number meshes and shows examples of the CFD results obtained on them and finishes by discussing our plans for future developments.

2 *NEKMESH*

NekMesh is a high-order mesh generation and modification program which can: convert meshes from one format to another; edit meshes whilst retaining high-order information; and generate high-order meshes from a CAD definition. It is set up in pipeline style; a series of modules are constructed and executed in order to arrive at a high-order mesh. A typical execution will involve an input stage, some processing (mesh editing) stages if required and an output. This arrangement means that the program can easily be used for a wide number of applications and can, with relative ease, be set up to read or write the high-order mesh in a number of different formats[1]. The focus of this paper is the generation of high-order meshes from a CAD description. We describe each stage of the pipeline in the following sections.

2.1 CAD INTERACTION

To interact with CAD geometries, functions including loading, modifying and querying the CAD are all required. *NekMesh* and *Nektar++* in general, have been interfaced with *OpenCas-*

[1]These currently include the *Nektar++*.xml, the *Gmsh* .msh and the *PyFR* .pyfrm formats.

cade [17] to provide this facility. *OpenCascade* is a large open-source library encapsulating a vast range of CAD operations and manipulations. Due to the large size of *OpenCascade*, and also to provide a facility to interact with alternative CAD engines, *NekMesh* uses a small wrapper layer between *Nektar++* and *OpenCascade*, thereby reducing the thousands of possible routines down to a few that are needed for mesh generation. Adopting this approach means that the *Nektar++* CAD engine shields developers from the complexity of *OpenCascade*. Consequently, the CAD handling within the program becomes more effective, robust and efficient.

2.2 AUTOMATIC MESH SPECIFICATION

One of the more fundamental challenges of high-order mesh generation is the creation of valid, suitably coarse linear meshes. It can be difficult for a user to define a set of mesh spacing which remains sufficiently coarse, sufficiently smooth, approximate the surface of the geometry well and still have the necessary resolution in the correct regions of the domain. This becomes extremely challenging in highly complex geometries and would usually require the user to go through many time-consuming iterations of generation to arrive at a suitable linear mesh. In *NekMesh*, this problem is alleviated by using a system to automatically obtain a full set of mesh spacings for the whole domain with very little input from the user.

NekMesh uses an octree structure for the spatial decomposition of the domain. Each subdomain, or octant, has a unique mesh specification parameter associated with it, which is queried during the meshing processes. To construct this representation of the domain the radius of curvature R is first sampled over the geometric surface of the domain and it is taken as

$$R = \min\left\{\frac{1}{k_1}, \frac{1}{k_2}\right\}, \tag{1}$$

where k_1 and k_2 are the principal curvatures at the point on a surface parametrised as $\boldsymbol{r}(u, v)$. A routine to ensure optimum sampling is used, such that any octant in the final octree which encompasses any region of geometric surface is guaranteed to contain more than one curvature sampling. The radius R is then related to a mesh sizing parameter δ as

$$\delta = 2R\sqrt{\varepsilon(2-\varepsilon)}. \tag{2}$$

The parameter ε can be viewed as a control of the sensitivity to curvature: decreasing this value increases the number of elements that will be created for a given curvature. Since R has a range from infinity, so too does δ, therefore sensible bounds must be placed on the value of δ such that $\delta_{\min} \leq \delta \leq \delta_{\max}$. In this setting then, $\delta_{\min}$, $\delta_{\max}$ and ε form the only set of user parameters required to obtain a set of mesh spacings.

Clearly, this curvature sampling only has information of the boundary surfaces. The octree is created by first encompassing the whole domain in a octant and then recursively and equally subdividing the octants until a stopping criteria is met. A octant will be subdivided if:

- The curvature samples within the octant differ in the value of δ by more than 10%; or
- The subdivision will not result in the new octants being smaller in size than $\delta_{\min}$; or
- If any of the neighbouring octants to a given octant are due to subdivide according to the two previous criteria, and their subdivision will cause a difference in size between neighbouring octants of more than one subdivision.

The first two points relate only to the curvature samples. Under these two conditions, only octants which contain a portion of the geometric surface will be split. The last point addresses this and ensures that the octree is smooth and therefore suitable to contain a smooth set of mesh spacings. Any octants which contain a number of curvature sampling points are then assigned a value of δ, which is taken as the minimum in the sample set. This only provides mesh spacings in octants which encompass some region of the surface. To complete the set of mesh spacings in the octree, this surface octant information is propagated outward using a mesh gradation criteria. To do this, each octant i is examined in turn. If the octant has any neighbours which have a value for δ, j, the relation

$$\delta_i = \delta_j + 0.1 r_{ij} \tag{3}$$

is used, where r_{ij} is the distance between the centroids of the two octants. This is repeated until all the octants in the domain have a value of δ. To ensure a smooth mesh specification, the values of the δ are then adjusted so that

$$\frac{|\delta_i - \delta_j|}{r_{ij}} \leq 0.1 \tag{4}$$

between an octant i and any of its neighbours j.

This system brings a very important feature to the mesh generation pipeline: automation. It permits a non-expert user to obtain a linear mesh which has suitable resolution in high curvature regions and that can be used for high-order meshing relatively quickly and without having to spend a lot of time setting up a large number of parameters. In regions of high curvature, any mesh generation pipeline is likely to create invalid elements if the linear mesh resolution is too low, but coarseness is the goal for linear mesh generation in the high-order meshing context. This fine balancing act is a challenge for the user. However within this system, very little effort is required and the number of invalid elements will be greatly reduced because the system will add higher resolution in regions of high curvature.

An additional advantage of this approach is that it is easy and computationally cheap to estimate the number of elements to be generated. With coarse meshes being the ideal and the computational cost of high-order methods being relatively high, the ability to predict the number of elements and alter the mesh before wasting any time generating is a powerful tool for the user. Figure 1 shows the accuracy of the element count prediction for three geometries up to 250 thousand elements. Furthermore, it is possible for the user to define refinement regions, with a smaller mesh sizing than would otherwise be specified, in the domain which will work in conjunction with the octree system to define a overall smooth mesh.

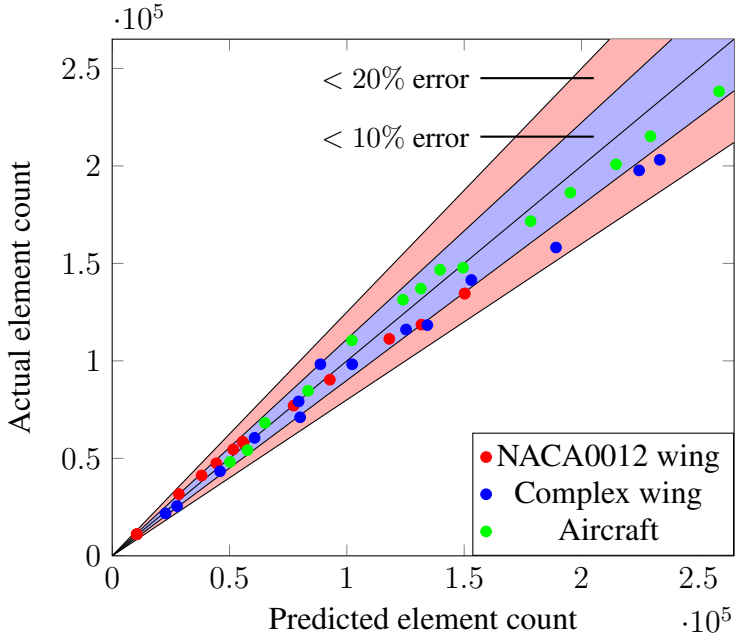


Figure 1: Predicted versus actual element count for various meshes using the automatic octree system

The use of a automatic mesh specification driven by surface topology using an octree structure is not completely new [18]. However, this reference noted that the computational cost and controllability of the method was poor, due to the use of the octree background grid, to the extent that the authors created a new system for automatic mesh definition which did not require a background grid [19]. In the context of generating coarse linear grids for high-order meshing, which inevitably contain far fewer elements, these conclusions are somewhat different. Indeed, we propose an efficient, robust and quality-controlled procedure. This is accomplished by actively keeping the algorithms in the construction of the octree as simple as possible.

Figure 2 shows an example high-order surface mesh of the DLR F6 geometry overlaid with a sliced view of the octree in the symmetry plane [20]. This figure shows that in regions of relatively low and constant curvature the octants are quite large; however in regions such as the tail and nose sections, the octree has smoothly subdivided numerous times to accommodate the very high curvature.

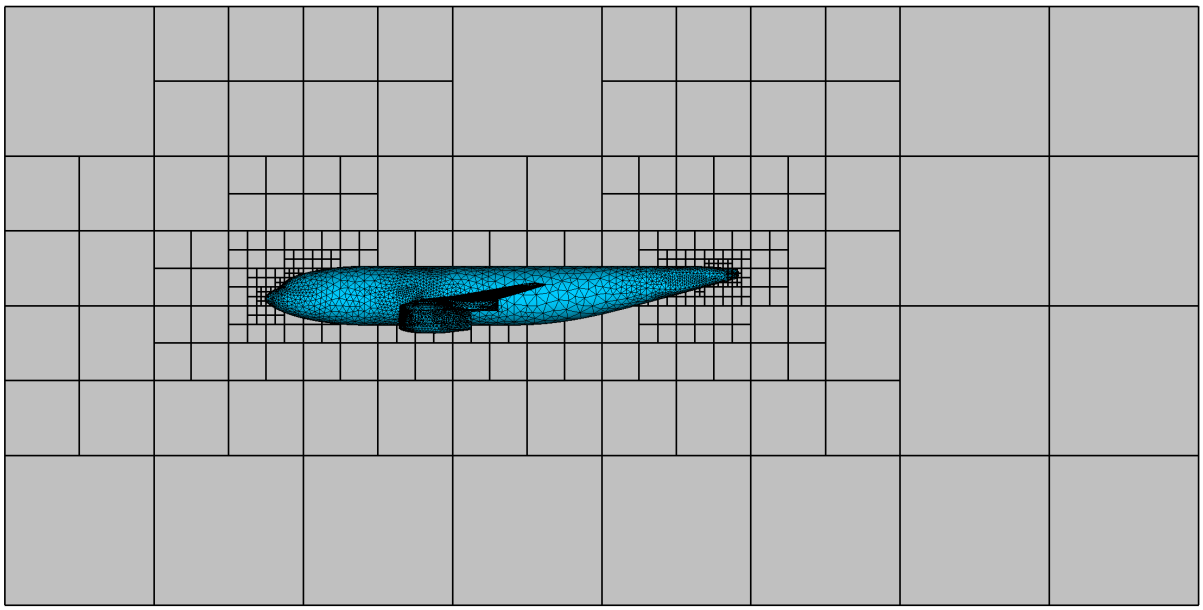


Figure 2: High-order mesh of the DLR F6 geometry with an overlay of the octree.

2.3 LINEAR MESH GENERATION

To automate high-order mesh generation, linear mesh generation must be part of the CAD to CFD pipeline. Using alternative linear mesh generators, and then importing the result, lowers robustness and quality because the routines are not typically designed for high-order use, and CAD information may not be preserved in the process. Some 3D linear mesh generators, such as those included in commercial packages will use a system for the mesh generation which first eliminates all the curvature information and reduces the domain to a linear approximation. This offers high levels of robustness but in linear meshes and the corresponding numerical methods the levels of inaccuracy in the geometric surface caused by this approach are, in most cases, too small to impact the final results. However in high-order methods this level of error is highly significant. Further to this, this approach can result in poor geometric representation in coarse meshes, especially in regions where surfaces are discontinuous. It also means that the surface mesh vertices cannot be located in the parameter spaces of the surfaces without using some form of reconstruction of this information, which can introduce errors and robustness issues in the generation of the high-order surface mesh.

To address these issues, *NekMesh* uses a bespoke approach to the generation of the linear mesh. Using the *NekMesh* CAD engine as the underlying source of information, the linear mesh is built in a bottom-up approach, thus ensuring geometric accuracy is maintained. In this linear mesh generation process, firstly, mesh vertices are fixed to the vertices of the CAD (0D). Curves are then meshed in their 1D parameter space using the end vertices as boundaries and the 3D real distances to decide the distribution in compliance with the mesh spacing within the octree. The 2D parameter spaces of the surfaces are then meshed, using the curve meshes as boundaries, and finally the interior volume is meshed using 3D constrained meshing. This bottom-up approach ensures that the CAD information associated with all the vertices, edges and faces in the surface mesh is kept and can be utilised easily in the high-order stages of the meshing pipeline.

To perform some of the more complex parts of the linear mesh generation, *NekMesh* uses third-party libraries, namely *Triangle* [21] to perform the 2D meshing in the parameter plane

of the surfaces and *TetGen* [22] to perform the interior tetrahedral meshing of the 3D volume. As might logically be expected, the meshing of the surfaces in the parameter plane using a Delaunay based method can produce highly irregular and low quality meshes once the surface mesh is projected into its 3D form, depending on how distorted the parametrisation of the CAD surface is. To overcome this, the linear surface mesh generator in *NekMesh* uses a optimisation step after the 3D surface mesh has been created. In this optimisation, both edge swapping and point relaxation are used to maximise the quality of the surface mesh. Clearly this approach does not preserve the Delaunay characteristic of the initial parameter plane mesh, but this is not our goal so long as the resulting mesh is of good quality. We find that this approach to linear mesh generation is robust and provides a solid foundation for high-order meshing, but because of the dependencies of the CAD, its ability to perform well is tightly linked to the distortion introduced by the CAD parametrisation.

2.4 HIGH-ORDER SURFACE MESHING

The key stage in the generation of a high-order mesh is the creation of the high-order surface. We know that small inaccuracies in the representation of the geometric boundary have a large impact on the flow solution. These inaccuracies include: highly distorted surface elements, i.e. very high curvature in the mesh entities; mesh nodes being a significant distance from the true CAD surface; and under-representation of the geometric curvature due to using a too low polynomial order with too large a element. If the vertex locations of the linear surface mesh are taken to be fixed, producing a high-order surface can be accomplished simply by using an affine mapping of the triangle in the 2D parameter plane to the reference triangle of a high-order element. This can then be used to locate the new high-order nodes in the parameter space, which are then projected into 3D using the CAD engine. However, this means that the high-order triangles will inherit the distortion of the CAD surface, lowering the quality of the mesh and in some cases causing invalid elements. To address this issue, *NekMesh* is equipped with a method to take the high-order surface mesh made using the affine mapping approach and optimise the location of the high-order nodes to reduce distortion. This is done by modelling the mesh entities as spring networks and minimising the spring energy, in a similar approach to the work of [12]. In mathematical notation, this can be expressed as

$$\min f, \tag{5}$$

$$f = \sum_{s} \frac{||\boldsymbol{r_1} - \boldsymbol{r_2}||^2}{w_s}, \tag{6}$$

which states that f, the spring energy, is the sum over all the springs in the system, where $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the 3D locations of the nodes at the ends of the springs and w_s is the inverse of the spring stiffness, which is calculated as a function of the nodal point distribution being targeted. Because the linear mesh vertices are held fixed during this procedure, the problem can be reduced to an entity-by-entity approach. First we optimise mesh edges that lie on curves; then edges that lie on surfaces; and finally interior triangle faces that lie on CAD surfaces. In the first case (edges on CAD curves), the problem is a 1D optimisation of spring system in the curve's parameter space t, so that

$$f = \sum_{i=1}^{P} \frac{||\boldsymbol{r}(t_{i+1}) - \boldsymbol{r}(t_i)||^2}{w_i}, \tag{7}$$

where i are the $P+1$ nodes along the high-order edge. Here, P is the polynomial order of the mesh being created and $w_i = z_{i+1} - z_i$, where z_i is the i-th entry in the distribution of nodal triangular points in the reference space of an edge, where $-1 \leq z \leq 1$. The initial values of t used come from the linear 1D mapping

$$t_i = t_1 \left(\frac{1 - z_i}{2} \right) + t_{P+1} \left(\frac{1 + z_i}{2} \right) \quad i = 1, ..., P+1. \tag{8}$$

where t_1 and t_{P+1} are the parametric coordinates of the end nodes of the edge, which are the vertices in the linear mesh and are considered to be fixed.

Performing the optimisation of the edges which lie on the CAD surfaces follows a very similar procedure but is formulated in the 2D parameter plane, i.e

$$f = \sum_{i=1}^{P} \frac{||\boldsymbol{r}(u_{i+1}, v_{i+1}) - \boldsymbol{r}(u_i, v_i)||^2}{w_i}. \tag{9}$$

This procedure reduces the distortion found in the high-order edges by minimising the length of the edge; that is, the optimised high-order edge will lie approximately on the geodesic between the two end points on the surface.

The procedure for optimising the location of face interior nodes requires a slightly alternative approach. The system is considered as a set of freely movable nodes, consisting of those nodes lying on the interior of the triangle, and a set of fixed nodes which lie on the edges. Each of the free nodes is connected to a system of six surrounding nodes by springs, and this is the system which is minimised. In a triangle of order P, there are $(P-2)(P-1)/2$ interior nodes. The function f is then written as

$$f = \sum_{i=1}^{(P-2)(P-1)/2} \sum_{s}^{6} \frac{||\boldsymbol{r}(u_i, v_i) - \boldsymbol{r}(u_s, v_s)||^2}{w_s}, \tag{10}$$

where w_s is calculated as the distance between the two nodes in a reference equilateral triangle, shown in the figure below along with the connectivity of the springs. The choice of a six spring system means that the method is applicable to any point distribution at any order. For example, Figure 3 shows a $P = 4$ triangle with a Gauss-Lobatto-Legendre distribution along the edges and a triangular Fekete distribution for the face interior points. To optimise the energy of the system a bounded version of the BFGS algorithm is used. This bounding is necessary due to the limits of the parameter space in the CAD entities [23]. One of the primary advantages of this approach is the availability and low cost of computing analytical gradients.

Figure 4 shows the effectiveness of this optimisation procedure. The left-hand figure shows the surface mesh before optimisation, and the right-hand figure after optimisation of the spring networks. In this case, the highly distorted CAD surface of the rounded leading edge of a wingtip causes suboptimal surface mesh generation. The figure clearly shows that the high-order triangles are deformed under the linear mapping. However, when this optimisation procedure is performed, the mesh edges approximate geodesic lines better and the resulting surface mesh is smoother.

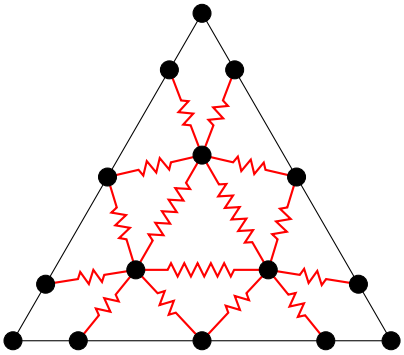
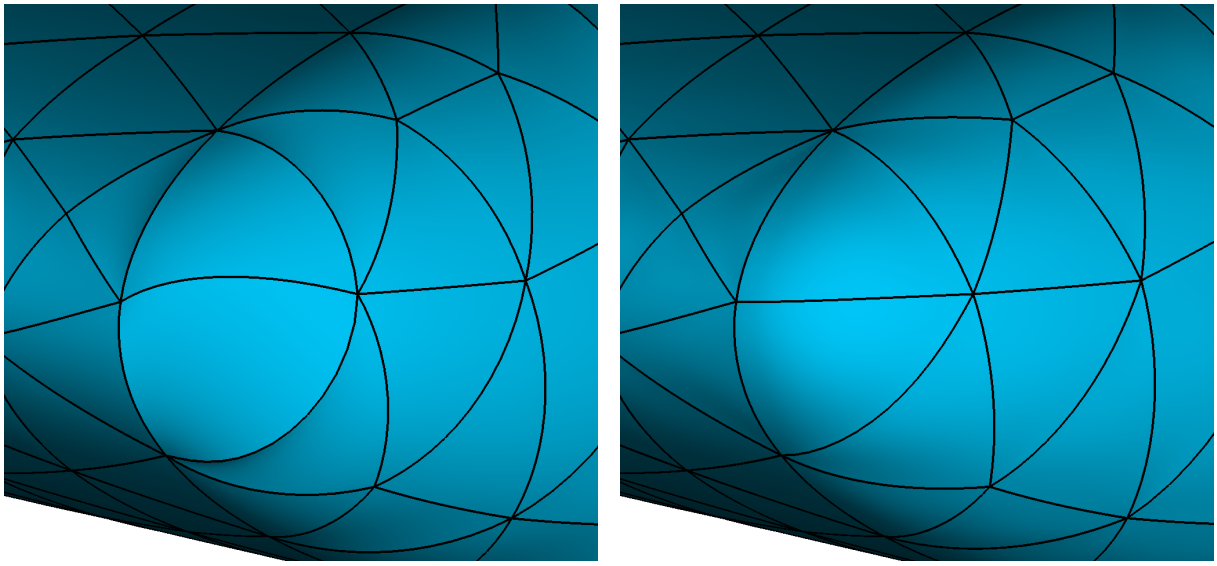


Figure 3: Reference Fekete distribution of points in a fourth-order triangle and the six spring system linking the free nodes.



(a) Unoptimised high-order surface.

(b) Optimised high-order surface.

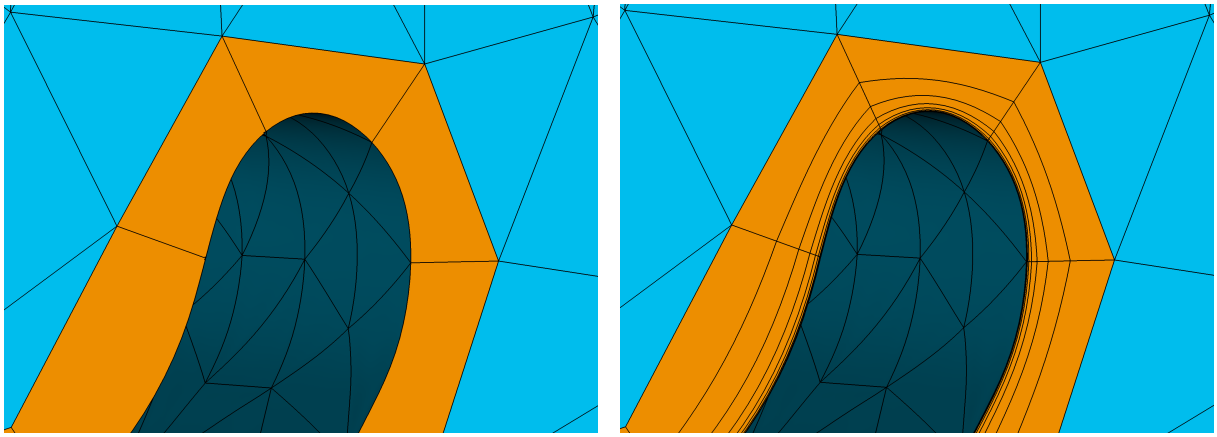
Figure 4: The high-order surface created using the affine mapping with (b) and without (a) optimisation in a region of high distortion in the CAD surface.

2.5 BOUNDARY LAYER GENERATION

The generation of anisotropic boundary layers to achieve low y^+ values has gone somewhat unnoticed in the recent research associated with creating robust high-order mesh generators, but is vital for the analysis of viscous flows. The sharp gradients in flow that are present in viscous simulations require highly stretched elements of a high aspect ratio of 100:1 and above in the wall-normal direction. This therefore poses a significant challenge for high-order mesh generation, since imposing surface curvature on thin elements will almost certainly yield self-intersecting elements in regions of high curvature.

To generate high-order boundary layers, *NekMesh* generates prismatic elements on the geometric boundary and tetrahedra in the rest of the mesh. The system uses a process whereby it inserts a linear 'macro' prismatic layer into the mesh of the desired thickness. This occurs after surface generation but before volume generation. The rest of the volume generation continues from that point using the top surface of the prisms as a pseudo-surface to mesh from. Once the high-order surface has been generated, which will curve one of the triangular faces

of the prism, the macro-prism is then split using an isoparametric approach [24]. This allows for the creation of very thin boundary layer elements without self intersection, since the height of the macro-layer can readily accommodate the curvature of the surface. The generation of the linear macro-prism layer represents one of the newest features of the program. Figure 5 shows a boundary layer region of macro-prisms split using the isoparametric approach to produce highly curved valid boundary-layer elements with very high aspect ratios, despite this they remain valid.



(a) (b)

Figure 5: Shows the splitting of a macro prism layer into 8 layers resulting in curved elements with high aspect ratios.

2.6 CORRECTION OF INVALID ELEMENTS

Up to this point, we have imposed the curvature of the CAD surfaces onto the surface triangulation and generated a volume mesh consisting of either tetrahedra (for inviscid meshes) or prisms and tetrahedra (for viscous boundary-layer meshes). However, this approach will invariably lead to the generation of volume elements where the curved faces impinge on another face, leading to a self-intersecting elements that are not suitable for computation. As such, a correction method is required in order to propagate the curvature into the interior of the domain, so as to correct invalid elements and generate a valid mesh.

As discussed in the introduction, there are a range of methods that can be used to fix these invalid elements. We have found that with the curvature targeting resolution from the automated mesh specification, high-order aware surface refinement and the use of prismatic boundary layers which accommodate more curvature than tetrahedra, the number of invalid elements generated is usually quite small. Complex cases with over 50,000 surface elements, providing the user parameters are sensible, tend to produce only a few invalid elements and with prismatic layers in a number of cases we find they can be prevented altogether. This could be alleviated further, or perhaps even eliminated, with high-order aware tetrahedral generation and this is a point for future work.

However, where invalid elements arise at the CAD surfaces, *NekMesh* can incorporate the use of a linear elastic solver, where the mesh is modelled as a solid body and the displacement at the CAD surface can be imposed to push curvature into the domain, as described in [16]. Since this operation is expensive, as it involves the solution of an elliptic PDE, we aim to

reduce the domain on which the deformation is imposed. We identify the invalid elements and their immediate neighbours by considering adjacent elements up to 10 levels deep, where one level is defined as connectivity through an element face. The linear elasticity equations are then solved on the subdomains, and the resulting displacements are imposed on the original mesh. Additionally, if there are still a large number of elements, the elastic solver can be run in parallel. Depending on the initial conditions of the linear mesh it may not be able to reach a valid configuration in all cases; however, we have found this approach to be quite effective in alleviating invalid elements.

3 EXAMPLES

This section describes two examples of complex three-dimensional geometries for which high-order meshes have been created using the *NekMesh* system. In each case we also show a flow solution to demonstrate the applicability and validity of the mesh. The first example is a high-angle of attack NACA0012 wing with a rounded wingtip, which is designed to run at high Reynolds number. Due to the convex nature of the geometry, the addition of a prismatic layer adjacent to the geometry meant that the high-order mesh was valid without needing to resort to the elasticity solver to correct any elements. The second is significantly more complex: a DLR F6 aeroplane geometry [20], an all tetrahedral mesh which did require the use of the linear elastic solver to correct invalid elements. In each case, the high-order mesh was created from the CAD definition using four user parameters: $\delta_{\min}$, $\delta_{\max}$, ε and P, where P is the desired order of the mesh.

3.1 NACA WING

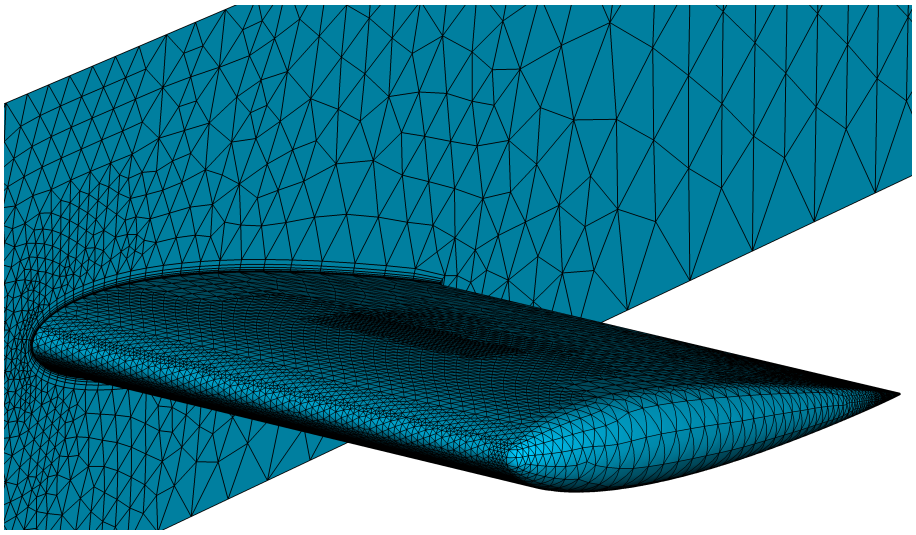
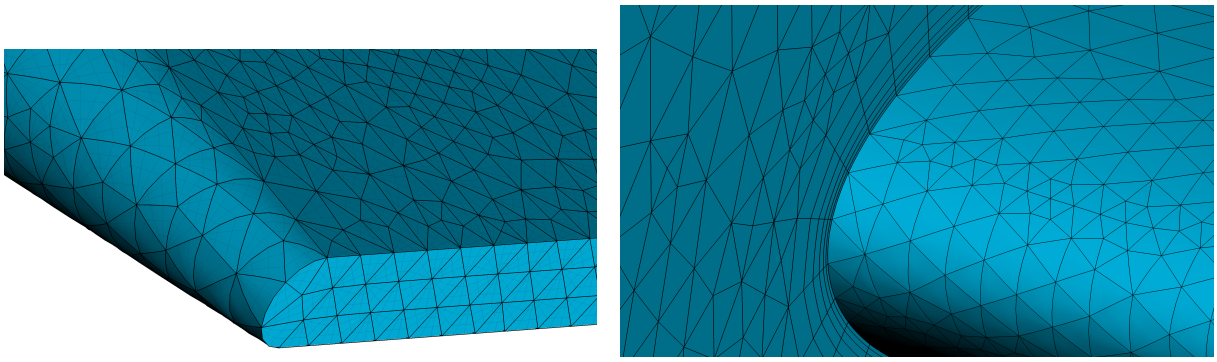


Figure 6: High-order mesh of the NACA wing.

Figure 6 shows a image of the surface of the NACA wing geometry. This mesh has a anisotropic boundary layer, as shown in figure 7. Regions of high curvature such as the leading edge of the wing and the curved wing tip, have increased resolutions compared to other parts of the mesh, as would be expected with the automated specification system. On the suction surface the resolution of the mesh has been manually increased to capture a separation region in the flow solution. Despite this modification, the octree system has ensured that the mesh

remained smooth, without large changes in element volume.



(a) Trailing edge of the mesh near the curved wing tip.

(b) High-order curved boundary in the symmetry plane.

Figure 7: Enlargements of areas of the NACA wing mesh.

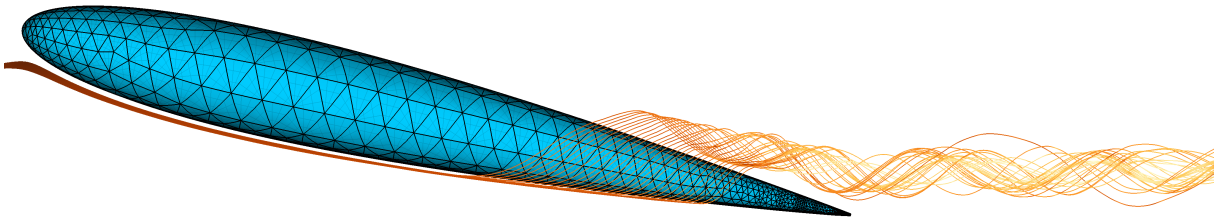


Figure 8: Particle trajectories following the tip vortex in the flow simulation at $Re = 10^5$ computed on a NACA wing mesh wiht $P = 4$.

To demonstrate the validity of the mesh, an incompressible flow simulation at $Re = 10^5$ was performed using *Nektar++*. Of particular interest for this case is the nature of the strong wingtip vortex produced [25]. Figure 8 shows that this vortex lies in the expected position, with the flow separating along approximately the correct location along the wing chord. We have found that in previous experiments, meshes which did not have suitable resolution or boundary layers were not able to properly capture these features.

3.2 DLR F6

The second example we consider is a significantly more complex DLR F6 jet geometry. A fully tetrahedral Euler mesh was created. Which required the use of the elasticity solver to correct invalid elements, but this number was small due to the resolution used near the surface and use of curvature-aware meshing. Figure 9 shows the mesh, with an inset highlighting the engine intake high curvature region. It can be observed that the automatic system for mesh specification has achieved its goal and added higher resolution to regions of the geometry which have high curvature, such as the engine intake and aircraft nose. Figure 10 shows a solution produced with this mesh for a low Reynolds number incompressible flow. While these results are not necessarily physically meaningful, they do demonstrate the suitability of the mesh. With a viscous boundary layer and significantly more computational time, this mesh could be capable of achieving much more realistic and relevant results, but this is outside the scope of this paper.

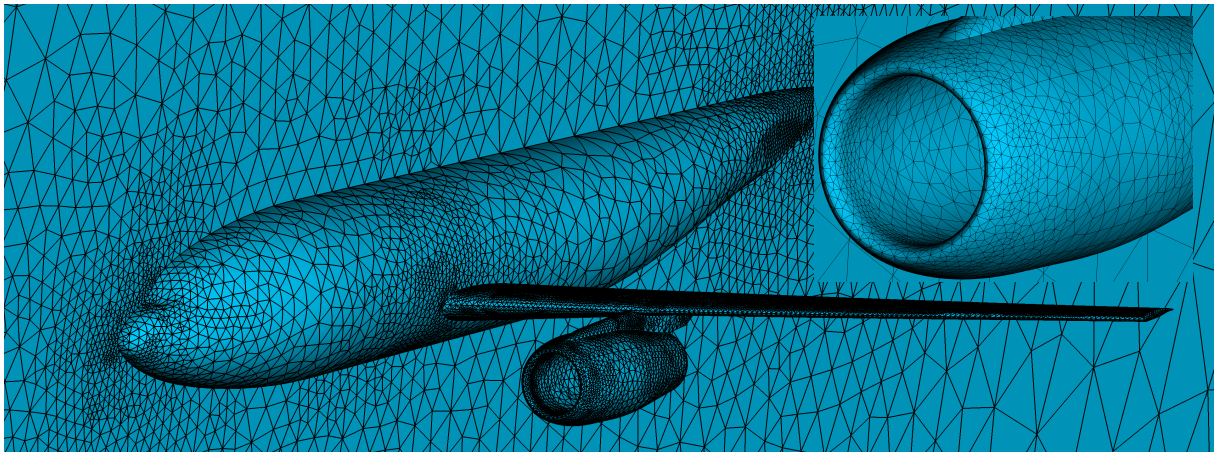


Figure 9: High-order mesh of the DLR F6 geometry. The inset shows an enlargement of the mesh around the front of the engine.

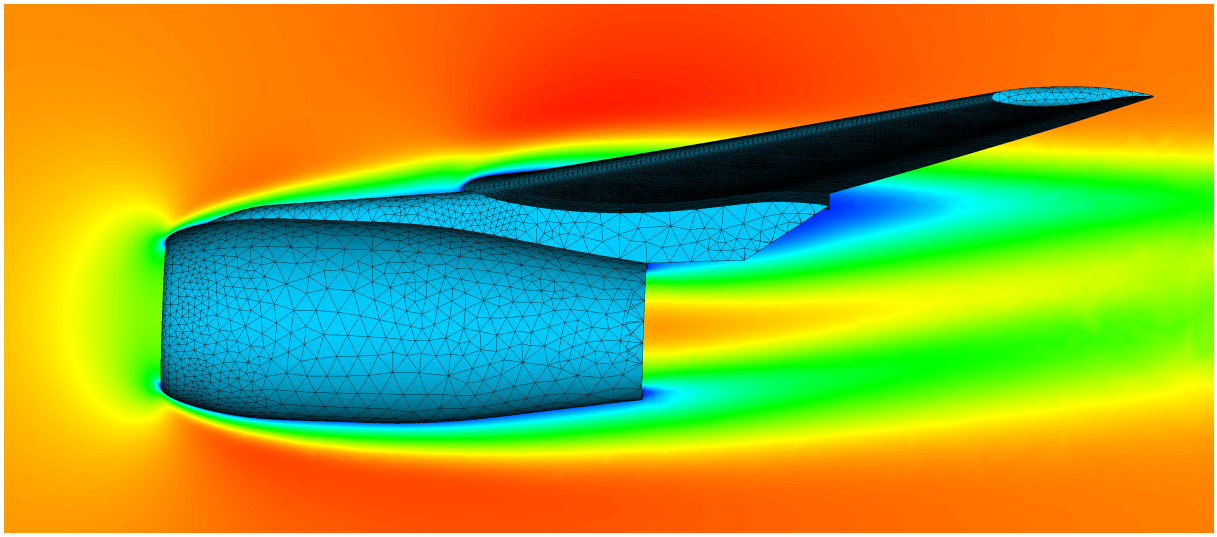


Figure 10: Low Reynolds number flow solution through the engine section.

4 CONCLUSIONS

We have introduced *NekMesh*, a system specifically designed to tackle robustness and complexity issues in the generation of high-order curvilinear 3D meshes for complex geometric cases. The methodology for each stage of the mesh generation pipeline has been outlined and two examples have demonstrated the capabilities of the system. *NekMesh* is currently under development; this article outlines the capabilities of the system as it is today, and marks the beta testing of the preliminary framework. We expect that in the future the structure and methodology of the program will advance and change becoming a much more capable system. The work in the near future will focus around more complex boundary layer generation techniques and more advanced robust methods for *a posteriori* invalid element corrections.

REFERENCES

[1] PE Vincent and A Jameson. Facilitating the adoption of unstructured high-order methods amongst a wider community of fluid dynamicists. *Mathematical Modelling of Natural*

Phenomena, 6(3):97–140, 2011.

[2] O Sahni, XJ Luo, KE Jansen, and MS Shephard. Curved boundary layer meshing for adaptive viscous flow simulations. *Finite Elements in Analysis and Design*, 46(1-2):132–139, 2010.

[3] J Slotnick, A Khodadoust, and J Alonso. CFD Vision 2030 study: A path to revolutionary computational aerosciences. Technical report, NASA, 2014.

[4] C Cantwell, D Moxey, A Comerford, A Bolis, G Rocco, G Mengaldo, D Grazia, S Yakovlev, J Lombard, D Ekelschot, B Jordi, H Xu, Y Mohamied, C Eskilsson, B Nelson, P Vos, C Biotto, M Kirby, and SJ Sherwin. Nektar++: An open-source spectral/hp element framework. *Computer Physics Communications*, 192:205–219, 2015.

[5] D. T. Lee and B. J. Schachter. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer and Information Sciences*, 9(3):219–242, 1980.

[6] A Gargallo-Peiró, X Roca, and J Sarrate. Optimization of a regularized distortion measure to generate curved high-order unstructured tetrahedral meshes. *International Journal for Numerical Methods in Engineering*, 103:342–363, 2015.

[7] M Fortunato and P-O Persson. High-order unstructured curved mesh generation using the Winslow equations. *Journal of Computational Physics*, 307:1–14, 2016.

[8] T Toulorge, C Geuzaine, JF Remacle, and J Lambrechts. Robust untangling of curvilinear meshes. *Journal of Computational Physics*, 254:8–26, 2013.

[9] S Dey, RM O'Bara, and MS Shephard. Towards curvilinear meshing in 3D: the case of quadratic simplices. *Computer-Aided Design*, 33(3):199–209, 2001.

[10] XJ Luo, MS Shephard, RM O'Bara, R Nastasia, and MW Beall. Automatic p-version mesh generation for curved domains. *Engineering with Computers*, 20(3):273–285, 2004.

[11] MS Shephard, JE Flaherty, KE Jansen, X Li, XJ Luo, N Chevaugeon, JF Remacle, MW Beall, and RM O'Bara. Adaptive mesh generation for curved domains. *Applied Numerical Mathematics*, 52:251–271, 2005.

[12] SJ Sherwin and J Peiró. Mesh generation in curvilinear domains using high-order elements. *International Journal for Numerical Methods in Engineering*, 53(1):207–223, 2001.

[13] ZQ Xie, R Sevilla, O Hassan, and K Morgan. The generation of arbitrary order curved meshes for 3D finite element analysis. *Computational Mechanics*, 51(3):361–374, 2012.

[14] SJ Sherwin and J Peiró. High-order automatic mesh generation: optimal p-type surface generation. In *16th IMACS World Congress*, Lausanne, 2000.

[15] P-O Persson and J Peraire. Curved mesh generation and mesh refinement using Lagrangian solid mechanics. In *47th AIAA Aerospace Sciences Meeting*, pages 1–11, Hampton, Virgina, 2009.

[16] D Moxey, D Ekelschot, U Keskin, SJ Sherwin, and J Peiró. A thermo-elastic analogy for high-order curvilinear meshing with control of mesh validity and quality. *Computer-Aided Design*, 72:130–139, 2016.

[17] OpenCascade, 2016.

[18] L.A. Kania and S.Z. Pirzadeh. A geometrically-derived background function for automated unstructured mesh generation. In *17th AIAA Computational Fluid Dynamics Conference*, page 5240, Toronto, 2005.

[19] S Z. Pirzadeh. Advanced unstructured grid generation for complex aerodynamic applications. *AIAA Journal*, 48(5):904–915, 2010.

[20] O. Brodersen and A. Stuermer. Drag prediction of engine-airframe interference effects using unstructured Navier-Stokes calculations. In *19th AIAA Applied Aerodynamics Conference*, number 2001-2414, Anaheim, 2001.

[21] J Shewchuk. Triangle: Engineering a 2D quality mesh generator and delaunay triangulator. *Applied Computational Geometry: Towards Geometric Engineering*, 1148:203–222, 1996.

[22] H Si. TetGen, a Delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software*, 41(2):11:1–11:36, 2015.

[23] RH Byrd, P Lu, J Nocedal, and C Zhu. A limited memory algorithm for bound constrained optimisation. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[24] D Moxey, M Green, SJ Sherwin, and J Peiró. An isoparametric approach to high-order curvilinear boundary-layer meshing. *Computer Methods in Applied Mechanics and Engineering*, 283:636–650, 2015.

[25] J-E. W. Lombard, D. Moxey, S. J. Sherwin, J. F. A. Hoessler, S. Dhandapani, and M. J. Taylor. Implicit large-eddy simulation of a wingtip vortex. *AIAA Journal*, pages 1–13, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MESH GENERATION FOR THE 2D NURBS-ENHANCED FINITE ELEMENT METHOD

Ruben Sevilla, Luke Rees and Oubay Hassan

Zienkiewicz Centre for Computational Engineering
College of Engineering, Swansea University, Swansea, SA1 8EN, Wales, UK.
e-mail: R.Sevilla@swansea.ac.uk

Keywords: triangular mesh, NEFEM, geometric features, high-order finite elements, NURBS.

Abstract. *The first method that enables the fully automatic generation of triangular meshes suitable for the so-called NURBS-enhanced finite element method (NEFEM) is presented. The methodology is able to produce meshes of the desired element size irrespectively of the geometric complexity and always encompassing the exact boundary representation of the domain given by NURBS curves. The strategy enables to completely circumvent the time consuming process of de-featuring complex geometric models before a finite element mesh suitable for the analysis can be produced. Numerical examples demonstrate the applicability of the proposed approach and the advantages compared to traditional finite element mesh generators.*

1 INTRODUCTION

The last decade has seen an increasing interest in high-order accurate methods due to their ability to reduce the dissipation and/or dispersion associated to traditional low-order methods. This is particularly attractive in computational electromagnetics [1, 2, 3] and computational fluid dynamics [4, 5, 6, 7] due to the need to propagate waves or vortices over long distances.

The need of curved elements to fully exploit the benefits of high-order methods has prompted a great interest in the research community on the generation of high-order curvilinear meshes [8, 9, 10, 11, 12, 13]. In terms of computational efficiency, very coarse curvilinear meshes and very high-order approximations are preferred, but the effect of geometric inaccuracies becomes evident. In [14] it was showed that the error induced by the isoparametric formulation can be up to one order of magnitude higher than a formulation with an exact boundary representation. In addition, complex geometries of large scale objects often contain very small geometric features, making necessary to produce extremely refined meshes in some regions. In many occasions, it is necessary to invest a non-negligible amount of time removing these small features present in the computer aided design (CAD) model before attempting to produce a mesh suitable for finite element analysis.

CAD and numerical simulations are still far from being integrated in a seamless manner. NURBS-enhanced finite element method (NEFEM) was proposed to bridge this gap between [15] by defining curved elements in contact with the CAD boundary in terms of the exact boundary description. The advantages of NEFEM have been studied in detail in [16, 14] and the extension of NEFEM to three dimensional domains was presented in [17], see also the review in [18]. Despite all these benefits, the lack of an automatic mesh generator has hampered the widespread application of NEFEM. In fact, this lack of such an automatic mesh generator for NEFEM has motivated the development of methods that do not require the generation of fitted NEFEM meshes but still use the NEFEM rationale, see for instance [19, 20, 21].

A novel mesh generation technique that allows to produce triangular meshes where the elements account for the exact boundary representation of the domain irrespectively of the desired element size and the geometrical complexity is presented. The produced meshes allow to extend the NEFEM formalism introduced in [15], where it was assumed that an edge of an element must be defined by at most one curve. A technique to extend the meshes to higher order by using a solid mechanics analogy is also proposed. Several examples demonstrate the applicability and potential of the proposed mesh generation technique.

2 NURBS-enhanced FEM fundamentals

An two dimensional domain Ω is considered, whose boundary is represented by non-uniform B-spline (NURBS) [22] curves $\boldsymbol{C}^k$ for $k = 1, \ldots, M$, with M the total number of boundary curves and a regular partition of the domain $\overline{\Omega} = \bigcup_e \overline{\Omega}_e$ in triangles is assumed.

An element edge Γ in contact with the NURBS boundary is defined by

$$\Gamma := \bigcup_{j=1}^{N_C} \boldsymbol{C}_j([\lambda_1^j, \lambda_2^j]) \tag{1}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the edge vertices, N_C the total number of curves defining the edge Γ, $\boldsymbol{C}_1(\lambda_1^1) = \boldsymbol{x}_1$ and $\boldsymbol{C}_{N_C}(\lambda_2^{N_C}) = \boldsymbol{x}_2$.

Assuming that the boundary of the element $\partial\Omega_e$ is a closed piecewise curve formed by trimmed NURBS curves and straight segments, the element Ω_e can be defined as

$$\Omega_e := \{\boldsymbol{x} \in \Omega \ \mid \ \mathrm{ind}_{\partial\Omega_e}(\boldsymbol{x}) \text{ is odd}\}$$

where $\mathrm{ind}_{\partial\Omega_e}$ is the index number of a point $\boldsymbol{x}$ with respect to the closed curve $\partial\Omega_e$.

Two properties characterise NEFEM compared to a classical isoparametric finite element formulation [14]. Firstly, the polynomial approximation is defined in the physical space, rather than in the reference element. Secondly, NEFEM uses the exact boundary representation of the domain to compute the integrals appearing in a weak variational formulation. For instance, a boundary integral along an edge on the boundary given by (1) is computed as

$$\int_\Gamma f(x,y)\, d\ell = \sum_{j=1}^{N_C} \int_{\lambda_1^j}^{\lambda_2^j} f\big(\boldsymbol{C}_j(\lambda)\big)\, \|\boldsymbol{C}_j'(\lambda)\|\, d\lambda$$

for a generic function f, see [23] for further details.

3 Generation of NEFEM meshes

3.1 Boundary discretisation

A set of sampling points is first defined on each NURBS curve on the boundary by utilising the distribution function described in [24]. The distribution of sampling point is then refined in regions where the distance from the straight sided segment connecting two sampling points and the true CAD boundary is more than a specified tolerance. Using the spacing distribution function, the desired mesh spacing at the sampling points is computed. Finally, when the spacing induces smaller elements that a specified tolerance, the spacing of the sampling points is checked and corrected.

Once the sampling points are generated, boundary curves are combined into *loops*. A loop is a collection of boundary curves $\boldsymbol{L} := \{\boldsymbol{C}_k\}_{k=1,\dots,N_L}$, where N_L is the total number of curves forming the loop, $\boldsymbol{C}_1(0) = \boldsymbol{C}_{N_L}(1)$ and $\boldsymbol{C}_k(1) = \boldsymbol{C}_{k+1}(0)$ for $k = 1, \dots, N_L - 1$.

The curves within a loop are ordered such that the first curve, $\boldsymbol{C}_1$, is of minimum length, where the length of a curve $\boldsymbol{C}_k$ trimmed to the subspace of the parametric space $[\lambda_i, \lambda_j]$ is

$$L_k([\lambda_i,\lambda_j]) := \int_{\boldsymbol{C}_k([\lambda_i,\lambda_j])} d\ell = \int_{\lambda_i}^{\lambda_j} \|\boldsymbol{C}_k'(\lambda)\|\, d\lambda$$

and the total length of the loop is by $L = \sum_{k=1}^{N_L} L_k([0,1])$.

For a given vertex $\boldsymbol{x}_{i-1}$, a candidate boundary vertex $\boldsymbol{x}_i$ at a distance $h(\boldsymbol{x}_{i-1})$ is first identified. As the objective is to enable the creation of edges across different curves, it is necessary to find the curve where the candidate vertex belongs, $\boldsymbol{C}_s$, and the associated parametric coordinate, λ_i.

In the next step, the mesh front formed by $\boldsymbol{x}_{i-1}$ and the candidate $\boldsymbol{x}_i = \boldsymbol{C}_s(\lambda_i)$ is checked, to guarantee that a valid element exists and can be built in the next stage of the proposed meshing technique. For a given edge Γ, the *horizon* of a vertex $\boldsymbol{x}_k$ of Γ is defined as the most distant point $\boldsymbol{x}_k^\star \in \Gamma$ that can be connected to $\boldsymbol{x}_k$ without intersecting Γ, as shown in Figure 1.

If the distance from the intersection of the lines connecting the boundary vertices and their corresponding horizons, $\boldsymbol{P}$, to the boundary vertices is less than the desired spacing, an interior node to form a triangle with the required spacing can be found. In this case, the second boundary vertex $\boldsymbol{x}_i$ is accepted and, otherwise, the second candidate is defined as the most distant sampling points that generates a valid front.

3.2 Domain discretisation

Given two vertices of a valid front, $\boldsymbol{x}_1^\star$ and $\boldsymbol{x}_2^\star$, and the intersection of the lines connecting the vertices and their horizons, a line of search is defined as the angle bisector of the two lines

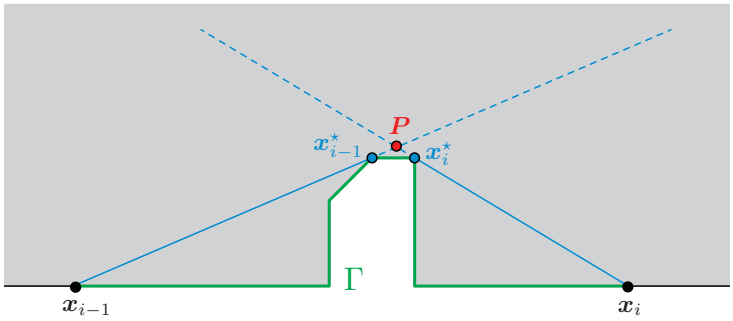


Figure 1: Illustration of the procedure to check the validity of a mesh front.

connecting the boundary vertices and their respective horizons, illustrated by a dashed red line in Figure 2. The third vertex of the element is defined as

$$\boldsymbol{x}_3 := \boldsymbol{P} + t\boldsymbol{n}_{\boldsymbol{P}} \tag{2}$$

where t takes the maximum value that guarantees that the edges $\overline{\boldsymbol{x}_1\boldsymbol{x}_3}$ and $\overline{\boldsymbol{x}_2\boldsymbol{x}_3}$ do not intersect the boundary and have length not greater than the desired element size h.

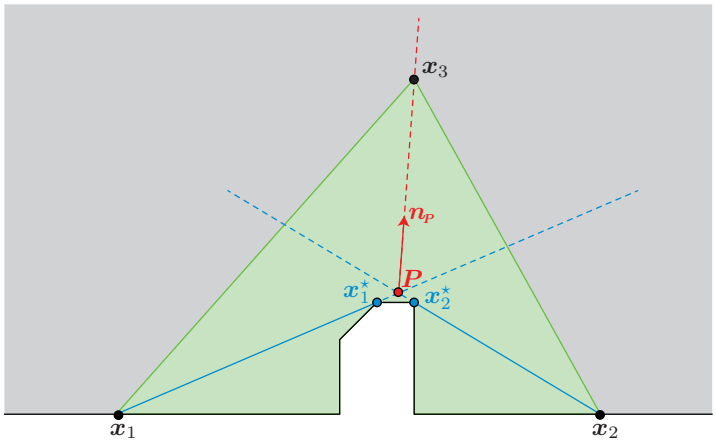


Figure 2: Illustration of the technique to build a triangular element.

3.3 Numerical integration cells

To perform the numerical integration on a NEFEM element, a recursive approach to subdivide an element in integration cells with, at most, one edge defined by a NURBS curve is devised. The integrations cells with an edge on the boundary are defined as a convex linear combination of the curved edge and the interior node as

$$\boldsymbol{\psi}_{\boldsymbol{x}_3}(\lambda, \vartheta) := (1 - \vartheta)\boldsymbol{C}(\lambda) + \vartheta \boldsymbol{x}_3 \tag{3}$$

where $\boldsymbol{x}_3$ denotes the internal vertex and $\boldsymbol{C}$ is the curve defining the boundary. This enables to build a composite two dimensional quadrature for elements in contact with the NURBS boundary.

The proposed algorithm uses a point interior to the domain, $\boldsymbol{Q}$ to check the visibility of the sampling points. The point $\boldsymbol{Q}$ is selected to be the interior node if the element has only one edge defined by NURBS as illustrated in the example of Figure 3. Otherwise, the point $\boldsymbol{Q}$ is selected to be the mid point of the only internal edge.

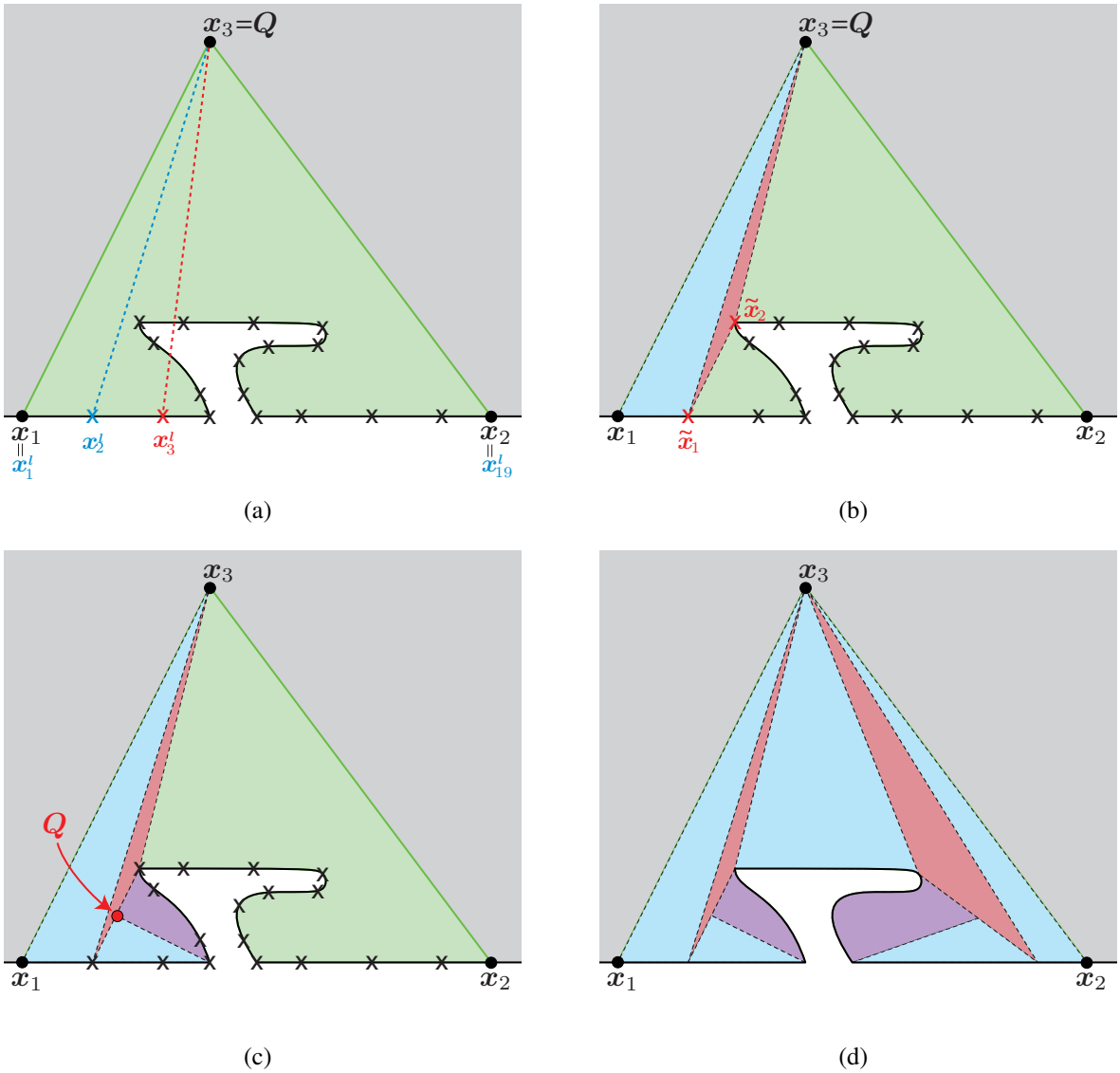


Figure 3: Illustration of the procedure to build the integration cells of an element. (a) Checking the visibility of sampling points, (b) formation of the the first two cells, (c) recursive application of the algorithm and (d) final partition of integration cells.

The visibility of each sampling points is checked until a non-visible point is found, the third sampling point in the example of Figure 3(a). Using the last visible point, two integration cells are built (red and blue cells in Figure 3(b)). The definition of the interior $\boldsymbol{Q}$ is then changed and the algorithm is applied to the triangle formed by $\tilde{\boldsymbol{x}}_1$, $\tilde{\boldsymbol{x}}_2$ and $\boldsymbol{Q}$, producing two new cells as illustrated in Figure 3(c). The procedure continues until the element is completely partitioned in integration cells as shown in Figure 3(d).

3.4 Extension to higher order elements

The strategy to extend the produced mesh to high-order is based on a linear elastic analogy [10] but the solution of the solid mechanics problem is performed element by element to ensure that the interior edges are kept as straight segments.

For an element Ω_e with at least one edge on the NURBS boundary, the triangle formed by its vertices is denoted by T_e. A high-order nodal distribution of the desired degree p is placed in T_e and a linear elastic problem is solved, assuming that the domain is an elastic medium and the Dirichlet boundary conditions are the displacements required to place the high-order edge nodes over the true (CAD) boundary. After solving the elastic problem, a high-order nodal distribution is obtained in the element Ω_e containing the exact description of the geometry.

4 Examples

4.1 Low-order mesh of a plate with a thin rectangular inclusion

The first example involves the generation of a NEFEM mesh of a square plate with a thin rectangular inclusion. The geometry of the inclusion is described using four NURBS curves (straight segments in this case) and the outer boundary is treated in the usual hierarchical manner. The total number of elements and nodes is 1,353 and 732 respectively, with only 11 elements with an edge on the NURBS boundary. It can be observed that the element size of the produced mesh is almost uniform, despite the discrepancy between the desired element size and the thickness of the inclusion.

A close view of the mesh around the inclusion is depicted in Figure 4, where it can be clearly observed that the element size is completely independent on the thickness of the inclusion. Two elements in contact with the boundary have one edge defined by two different NURBS curves. The integration cells, obtained by following the strategy presented in Section 3.3, are also shown in Figure 4.

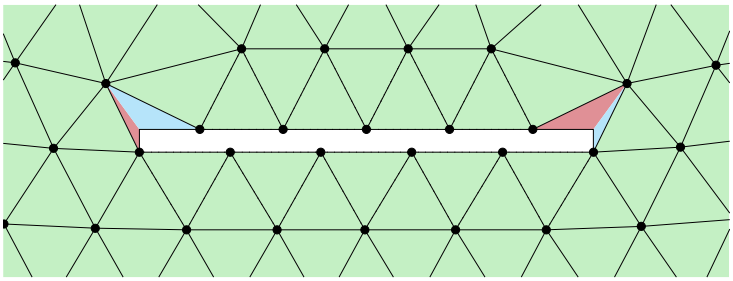


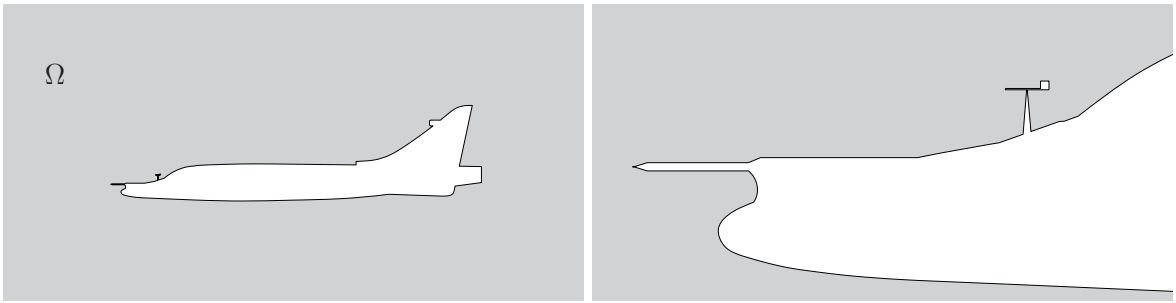
Figure 4: Detail of the NEFEM mesh near the inclusion

The minimum and maximum element edge length on the boundary is approximately 1.8 and 2.0 respectively, illustrating the ability of the proposed approach to discretise the boundary with a desired element size, irrespectively of small geometric features.

In this example, the main advantage of the proposed meshing technique is not to reduce the number of elements compared to a traditional FE mesh but to avoid the small elements introduced by a standard FE mesh generator to capture the thickness of the inclusion.

4.2 High-order mesh around an aircraft profile

The second example involves the generation of a NEFEM mesh around the aircraft profile represented in Figure 5. The detailed view of the front part shows a number of very small features compared to the size of the aircraft.



(a) (b)

Figure 5: (a) Domain Ω representing the exterior of an aircraft profile and (b) detailed view of the domain showing a number of small geometric features.

The aircraft profile is described by 53 NURBS curves, where the length of these curves differ in more than six orders of magnitude. A NEFEM mesh with uniform desired element size 2000 times larger than the shortest curve is shown in Figure 6. Two detailed view of the mesh near

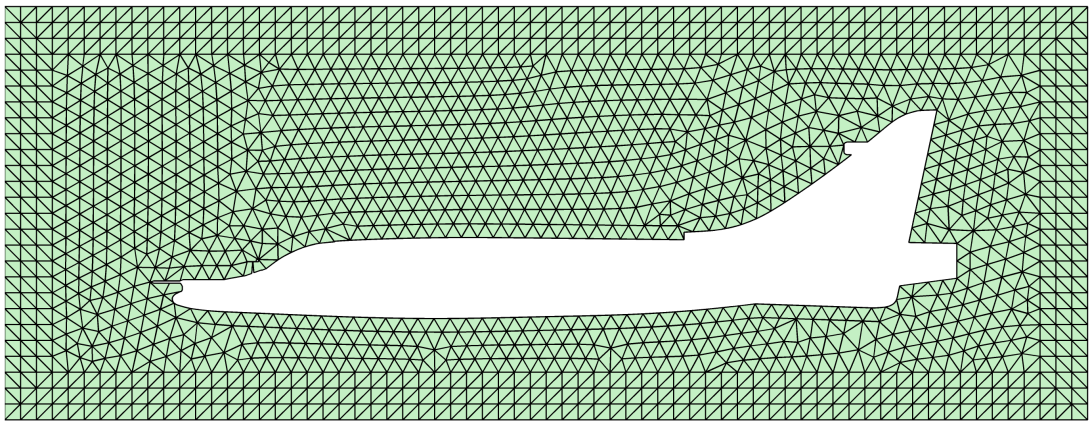
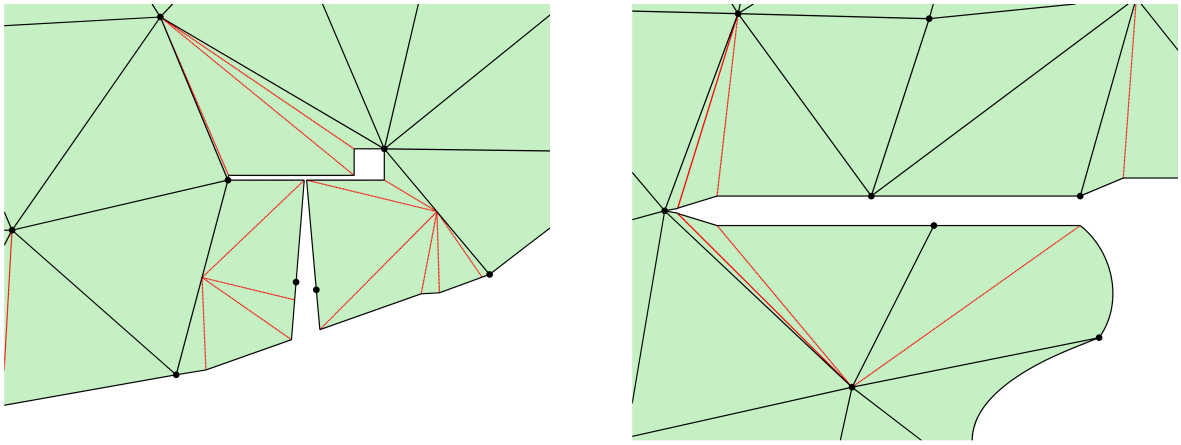


Figure 6: NEFEM mesh around a complex aircraft profile.

the most critical zones are shown in Figure 7, illustrating the integration cells used in some elements that contain edges formed by several NURBS curves.

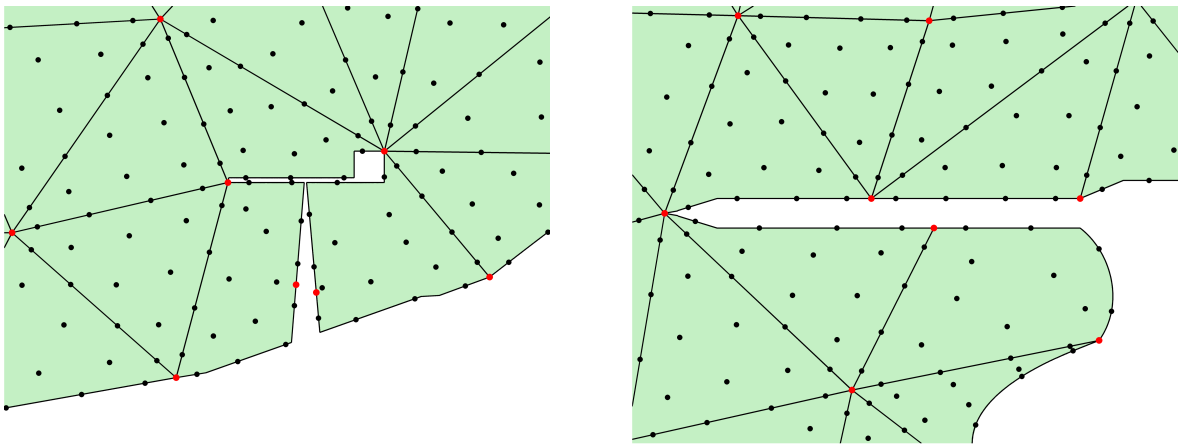
The minimum and maximum length of the edges on the NURBS boundary generated using the proposed approach are 0.03 and 0.05 respectively, illustrating the ability to maintain the desired element size independently on the complexity of the NURBS boundary representation of the domain.

Next, the mesh is extended to high-order using the proposed solid mechanics analogy only for elements with one or more edges on the boundary. The generated mesh for a degree of approximation p=5 is shown in Figure 8.



(a) (b)

Figure 7: Details of the NEFEM mesh of Figure 6 near the regions containing the two shortest NURBS curves. The black dots denote the element vertices and the red lines are the integration cells.



(a) (b)

Figure 8: Detail of the NEFEM mesh and the high-order nodal distribution adapted to the exact geometry. Red dots denote the element vertices.

5 Concluding remarks

A technique to produce meshes that account for the CAD boundary representation of the domain given by NURBS curves and with an element size independent on the geometric complexity of the computational domain has been presented. A methodology to extend the generated meshes to an arbitrary order is proposed and several examples demonstrate the applicability and potential of the proposed mesh generation technique.

REFERENCES

[1] J. S. Hesthaven and T. Warburton, "Nodal high-order methods on unstructured grids I. time-domain solution of Maxwell's equations," *Journal of Computational Physics*, vol. 181, no. 1, pp. 186–221, 2002.

[2] M. Ainsworth, "Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods," *Journal of Computational Physics*, vol. 198, no. 1, pp. 106–130, 2004.

[3] R. Sevilla, O. Hassan, and K. Morgan, "The use of hybrid meshes to improve the efficiency of a discontinuous Galerkin method for the solution of Maxwells equations," *Computers & Structures*, vol. 137, pp. 2–13, 2014.

[4] H. Luo, J. D. Baum, and R. Löhner, "A fast, p-multigrid Discontinuous Galerkin method forcompressible flows at all speeds," in *Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit*, (Reno, Nevada), AIAA, January 2006.

[5] R. Sevilla, O. Hassan, and K. Morgan, "An analysis of the performance of a high-order stabilised finite element method for simulating compressible flows," *Computer Methods in Applied Mechanics and Engineering*, vol. 253, pp. 15–27, 2013.

[6] Z. Wang, K. Fidkowski, R. Abgrall, F. Bassi, D. Caraeni, A. Cary, H. Deconinck, R. Hartmann, K. Hillewaert, H. Huynh, *et al.*, "High-order CFD methods: current status and perspective," *International Journal for Numerical Methods in Fluids*, vol. 72, no. 8, pp. 811–845, 2013.

[7] H. Huynh, Z. J. Wang, and P. Vincent, "High-order methods for computational fluid dynamics: a brief review of compact differential formulations on unstructured grids," *Computers & Fluids*, vol. 98, pp. 209–220, 2014.

[8] M. S. Shephard, J. E. Flaherty, K. E. Jansen, X. Li, X. Luo, N. Chevaugeon, J.-F. Remacle, M. W. Beall, and R. M. O'Bara, "Adaptive mesh generation for curved domains," *Applied Numerical Mathematics*, vol. 52, no. 2-3, pp. 251 – 271, 2005.

[9] P.-O. Persson and J. Peraire, "Curved mesh generation and mesh refinement using lagrangian solid mechanics," in *Proceedings of the 47th AIAA Aerospace Sciences Meeting and Exhibit*, AIAA, 2009.

[10] Z. Q. Xie, R. Sevilla, O. Hassan, and K. Morgan, "The generation of arbitrary order curved meshes for 3D finite element analysis," *Computational Mechanics*, vol. 51, no. 3, pp. 361–374, 2013.

[11] T. Toulorge, C. Geuzaine, J.-F. Remacle, and J. Lambrechts, "Robust untangling of curvilinear meshes," *Journal of Computational Physics*, vol. 254, pp. 8–26, 2013.

[12] A. Gargallo-Peiró, X. Roca, J. Peraire, and J. Sarrate, "Distortion and quality measures for validating and generating high-order tetrahedral meshes," *Engineering with Computers*, vol. 31, no. 3, pp. 1–15, 2014.

[13] D. Moxey, D. Ekelschot, U. Keskin, S. Sherwin, and J. Peiró, "A thermo-elastic analogy for high-order curvilinear meshing with control of mesh validity and quality," *Procedia Engineering*, vol. 82, no. 0, pp. 127 – 135, 2014. 23rd International Meshing Roundtable (IMR23).

[14] R. Sevilla, S. Fernández-Méndez, and A. Huerta, "Comparison of high-order curved finite elements," *International Journal for Numerical Methods in Engineering*, vol. 87, no. 8, pp. 719–734, 2011.

[15] R. Sevilla, S. Fernández-Méndez, and A. Huerta, "NURBS-enhanced finite element method (NEFEM)," *International Journal for Numerical Methods in Engineering*, vol. 76, no. 1, pp. 56–83, 2008.

[16] R. Sevilla, S. Fernández-Méndez, and A. Huerta, "NURBS-enhanced finite element method for Euler equations," *International Journal for Numerical Methods in Fluids*, vol. 57, no. 9, pp. 1051–1069, 2008.

[17] R. Sevilla, S. Fernández-Méndez, and A. Huerta, "3D-NURBS-enhanced finite element method (NEFEM)," *International Journal for Numerical Methods in Engineering*, vol. 88, no. 2, pp. 103–125, 2011.

[18] R. Sevilla, S. Fernández-Méndez, and A. Huerta, "NURBS-enhanced finite element method (NEFEM): A seamless bridge between CAD and FEM," *Archives of Computational Methods in Engineering*, vol. 18, no. 4, pp. 441–484, 2011.

[19] G. Legrain, "A NURBS enhanced extended finite element approach for unfitted CAD analysis," *Computational Mechanics*, vol. 52, no. 4, pp. 913–929, 2013.

[20] R. Sevilla and E. Barbieri, "NURBS distance fields for extremely curved cracks," *Computational Mechanics*, vol. 54, no. 6, pp. 1431–1446, 2014.

[21] O. Marco, R. Sevilla, Y. Zhang, J. J. Ródenas, and M. Tur, "Exact 3D boundary representation in finite element analysis based on Cartesian grids independent of the geometry," *International Journal for Numerical Methods in Engineering*, vol. 103, no. 6, pp. 445–468, 2015.

[22] L. Piegl and W. Tiller, *The NURBS Book*. London: Springer-Verlag, 1995.

[23] R. Sevilla and S. Fernández-Méndez, "Numerical integration over 2D NURBS shaped domains with applications to NURBS-enhanced FEM," *Finite Elements in Analysis and Design*, vol. 47, no. 10, pp. 1209–1220, 2011.

[24] J. Peiró, "Surface grid generation," in *Handbook of grid generation* (J. F. Thompson, B. K. Soni, and N. P. Weatherill, eds.), ch. 19, CRC Press, 1999.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

IMAGE SEGMENTATION TECHNIQUES FOR BIOMEDICAL MODELING: ELECTROPHYSIOLOGY AND HEMODYNAMICS

Alexander A. Danilov[1,2], **Roman A. Pryamonosov**[1,2], **and Alexandra S. Yurova**[1,2,3]

[1]Russian Academy of Sciences, Institute of Numerical Mathematics
Gubkina 8, Moscow, 119333, Russia
e-mail: a.a.danilov@gmail.com, breaksound@yandex.ru, sanyka@gmail.com

[2] Moscow Institute of Physics and Technology
Institutskii 9, Dolgoprudny, 141700, Russia

[3] Lomonosov Moscow State University
Leninskie Gory 1, Moscow, 119991, Russia

Keywords: Image Segmentation, Mesh Generation, Medical Images, Hemodynamics, Electrophysiology.

Abstract. *The work addresses segmentation techniques for generation of individualized computational domains on the basis of medical imaging dataset. The computational domains will be used in 3D electrophysiology models and 3D-1D coupled hemodynamics models. Several techniques for user-guided and automated segmentation of soft tissues, segmentation of vascular and tubular structures, generation of centerlines, 1D network reconstruction, correction and local adaptation are examined. We propose two algorithms for automatic vascular network segmentation and user-guided cardiac segmentation.*

1 INTRODUCTION

In this paper we present and develop methods and algorithms for construction of patient-specific discrete geometric models for two important medical applications. Each application imposes specific restrictions on both the input medical images and the output patient-specific discrete model, and, therefore, calls for a specific class of 3D reconstruction methods. The first application considers numerical modelling of hemodynamics. The second application considers electrophysiology modelling.

Individualized regional network of blood vessels is built from patient-specific CT-scans by our fully automatic algorithm for arteries identification. It starts from the fast variant of the isoperimetric distance trees algorithm [1] for aorta identification. The coronary and cerebral networks are reconstructed by the use of Frangi vesselness filter [2]. The filter is based on Hessian 3D analysis of the CT-image and is applicable for all tubular structures in the vascular data set.

Modelling of cardiac electrophysiology may be formalized as the full-scale study of the heart electrical activity from inner-cellular level to the cardiac tissues level. Various mathematical models and numerical methods have been developed for modelling of cardiac electrophysiology [3]. Generally the full-scale electrophysiology modelling involves simulations of electric potential propagation on three different scales: the single-cell models represented by ODE are used on cellular level, the bidomain model represented by PDE system is used on cardiac tissue level, the simplified quasi-static version of Maxwell's equations is used on the whole-body level.

The reconstruction of personalized anatomical model of the pathological heart is one of the crucial steps in electrophysiology modelling. The bidomain model requires an accurate anatomical model of patient heart accounting myocardium anisotropy structure. Our preliminary simulations indicate that electrophysiological processes have low sensitivity to segmentation errors outside the thorax region. Thus, patient-specific segmentation should be focused on the cardiac region, and a reference human model may be used in the remaining part of the body.

The corner stone for medical image processing is segmentation process when each voxel of the 3D medical image is assigned with the particular tissue or internal organ label. Various medical image segmentation techniques have been developed [4, 5, 6]. The most promising fully automatic segmentation methods belong to atlas-based segmentation techniques. The patient-specific segmentation is obtained from the atlas of presegmented images of other individuals. This atlas should contain enough different cases for accurate mapping of the new patient data. Thus atlas-based approach requires huge amount of segmentation expert work for preparation of atlases and the development of algorithms dealing with big data. Semi-automatic segmentation technologies require interaction with the operator. They are used primarily for precise local segmentations, where only one organ or tissue is processed. In our previous work we used several techniques for adaptation of the once segmented reference human model to different individuals. This technique relies on anthropometric scaling, control points mapping and supervised segmentation [7, 8].

In the following two sections we discuss and present several methods for 3D reconstruction of vascular network and cardiac model from medical images.

2 VASCULAR SEGMENTATION

In the current work we consider two regions of interest: cerebral and coronary vessels. Input data is DICOM datasets obtained with contrast enhanced Computed Tomography Angiography

(ceCTA). For simplicity of presentation we assume that quasi-isotropic voxel sizes are used in this work.

Our automatic vessel segmentation methodology for coronary arteries was proposed in [9]. In the current work it is extended to process cerebral vessels. Essential steps of this method consist of aorta segmentation, computation of vesselness values, searching branches of aorta arch or ostia points, and removing segmentation errors near aorta boundary. Important additional step in the cerebral case is bone intensity decreasing by multiscale Match Mask Bone Elimintaion (multiscale MMBE) algorithm [10]. For every patient the last step requires not only ceCTA dataset but also corresponding CTA dataset not enhanced by contrast. All steps are shown in Figure 1.

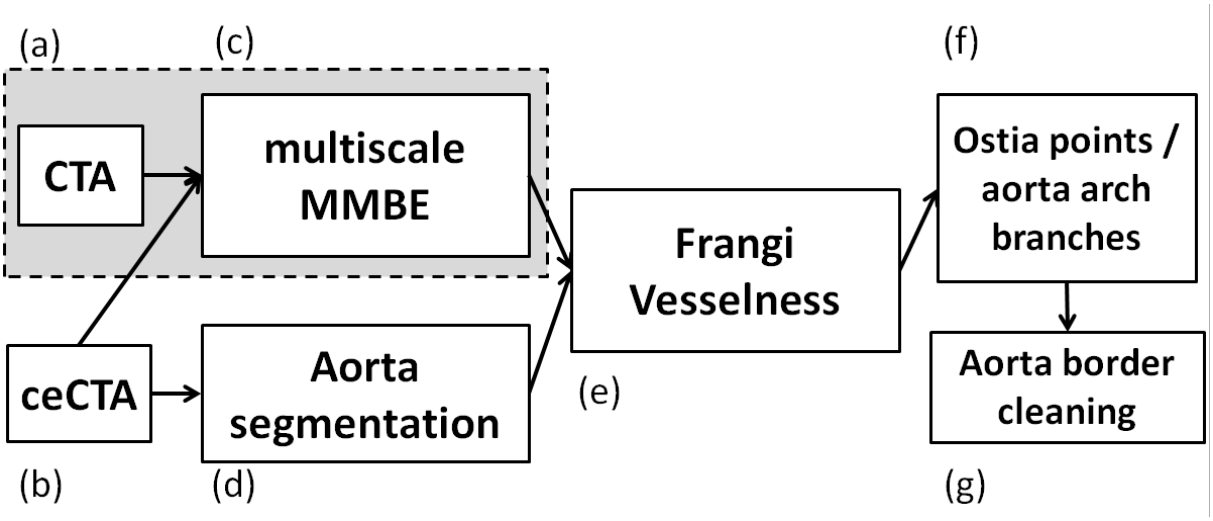


Figure 1: Segmentation pipeline. Gray rectangle highlights cerebral case only steps. (a) and (b) are algorithm inputs, (c)-(g) essential algorithm steps.

Large variation of vessel radii results in numerically expensive Hessian-based filtering. Since aorta is several times larger than arteries studied in this work it is natural to use specific algorithm for its segmentation. In both coronary and cerebral cases we perform the following steps. First, the Hough Circleness Transformation [11] is used to detect largest bright disk D corresponding to aorta cross-section. Second, connected mask is obtained by region growing, starting at the center of D, with lowest intensity inside of D as a threshold parameter. Third, IDT algorithm [1] cuts mask in bottlenecks and, finally, mathematical morphology operations are used to remove vessels branching from aorta.

Once aorta is segmented, the multiscale MMBE method is applied to decrease intensities of bone tissues in cerebral case. This approach allows us to apply standard segmentation methodologies to get anatomically correct segmentations.

The next step is computation of Frangi Vesselness [2], which results in bigger values inside bright tubular structures. In coronary datasets only coronary arteries and aorta represent these structures. In cerebral case such structures are bones and arteries, and once bones are darkened, Frangi Vesselness is applicable. Segmentation of arteries is produced by thresholding of vesselness values. Then coronary arteries and cerebral aorta arch branches are found as voxel connectivity components near the aorta boundary.

Final step is removal of segmentation errors. Thresholding of Frangi Vesselness values may produce “leaks” near the aorta boundary. Let us assume we have segmented object represented

by set of voxels A, but segmentation S contains set of extra voxels B, i.e. $S = A \cup B$. In case S is a connected component we call B as leak. In terms of definition leaks cannot be removed by "remove islands" procedure. Also leaks may have different shapes, so it is hard to find appropriate mathematical morphology operation to remove leaks automatically.

We propose automatic method for leaks removal from aorta boundary. Let us denote mask of cerebral vessels by V_{mask}, mask of aorta by A_{mask} and define voxel layer $L_i = \{v \in V_{\text{mask}} \mid \text{dist}(A_{\text{mask}}, v) = i\}$. For $i = i_{\max}, \ldots, 0$ the following procedure is taken: remove all voxels from layer L_i having no neighbouring voxels in L_{i+1}. Experiments has shown that $i_{\max}$ can be set to 15 voxels. In theory $i_{\max}$ must be set bigger than the thickness of a leak.

Several examples of vessel structure reconstruction are presented in [8, 9].

3 CARDIAC IMAGES SEGMENTATION

Mathematical modelling of electrophysiology requires an anatomical 3D cardiac model. Our group created a reference 3D cardiac model based of Visible Human Project (VHP) data [12]. This model is used for evaluation of modelling methods and algorithms. The reference model may be used for patient-specific adaptation in case medical imaging data is insufficient for complete segmentation.

Several segmentation techniques were examined for automatic segmentation of heart ventricles. The preprocessing steps include optional noise reduction using non-local means filtering and histogram equalization processing. The first used approach was based on SLIC supervoxel clustering [13] and recursive region adjacency graph partitioning.

The second approach was based on user-guided active contour segmentation with supervised random forest classification from ITK-SNAP segmentation platform [14]. The results of user-guided segmentation highly depend on user expertise.

Based on our experience we propose to combine both approaches: the supervised random forest classification is used for supervoxel grouping. Several post-processing steps are used to improve the segmentation.

VHP dataset provides CT, MRI and photo images of human cadaver. Transversal photo images are digitized with resolution 0.33×0.33 mm, distance between images is 1 mm. Photo images were aligned, rescaled and combined in a 3D volume dataset. ITK-SNAP segmentation software was used for visualization and segmentation. Only thorax region was investigated resulting in cropped image $573 \times 330 \times 170$ voxels with resolution 1 mm (Figure 2).

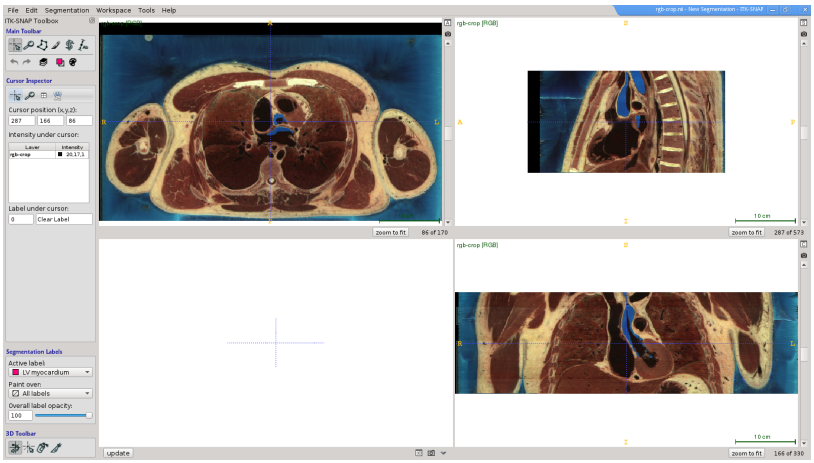


Figure 2: ITK-SNAP platform snapshot with VHP photo images of cardiac region.

Heart cavities and blood vessels were segmented using active contours method and random forest classifier. These regions are filled with blood and usually have a distinct color and/or intensity on photo images and CT/MRI images. The segmented region was split into parts corresponding to right atrium (RA), left atrium (LA), right ventricle (RV), left ventricle (LV), aorta, pulmonary trunk, superior vena cava and inferior vena cava (Figure 3).

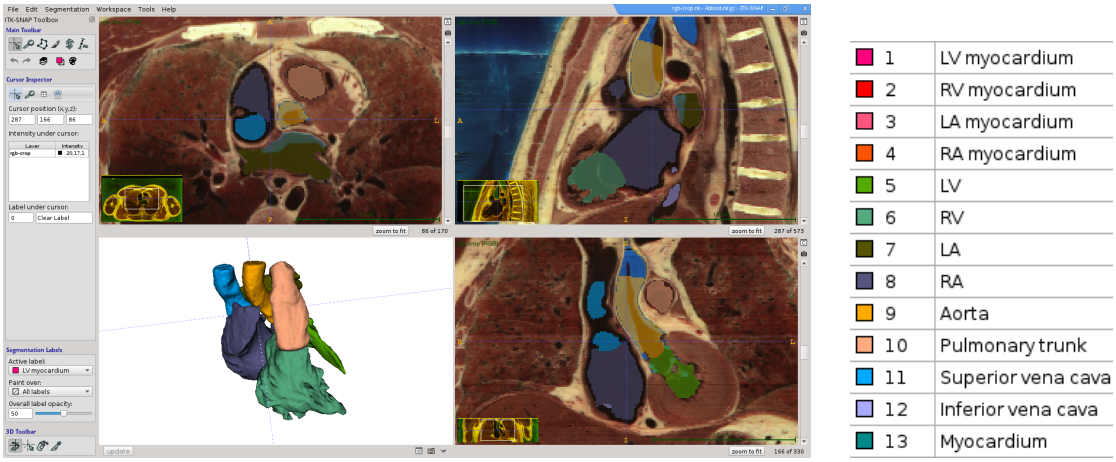


Figure 3: Segmentation of cardiac cavities and blood vessels.

Myocardium segmentation is performed in the vicinity of segmented heart cavities. Myocardium tissue differs in texture and intensity compared with neighbouring fat tissue, bones and lungs (Figure 4). This segmentation step is also performed using active contours method and random forest classifier.

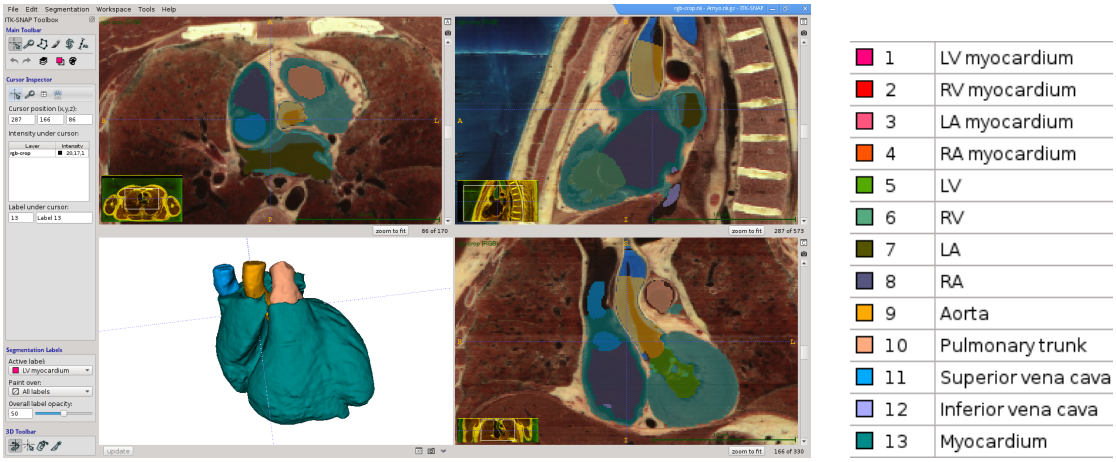


Figure 4: Myocardium segmentation.

The final step involves segmentation smoothing. The myocardium tissue was divided into four parts using mathematical morphology operations: right atrium, left atrium, right ventricle, and left ventricle (Figure 5). This division will be used in modelling algorithms for synthetic computation of myocardium fiber orientation.

The segmented model is a 3D structure with $128 \times 119 \times 129$ voxels with resolution 1 mm and 12 tissue labels corresponding to different heart cavities, blood vessels and four myocardium regions. The segmented model may be used for estimation of myocardium mass and ventricles

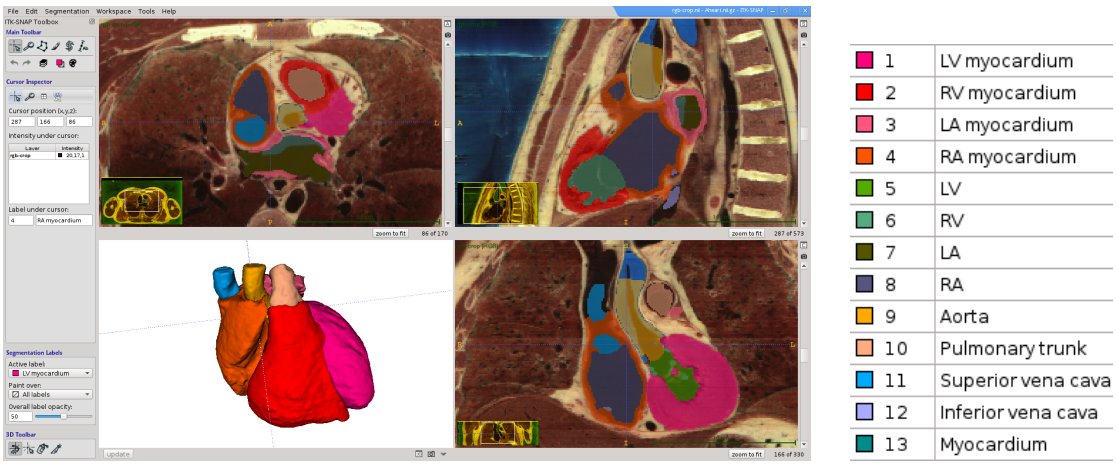


Figure 5: Final VHP segmented model.

volume. Assuming relative density of myocardium tissue is equal to 1.05 g/ml we obtain the mass of right ventricle 95 g, and left ventricle – 187 g. The volume of right ventricle is 77 ml, the volume of the left ventricle – 21 ml. The computed myocardium mass lies in reference limits for a male aged 38 years [15, 16]. We should note, that the volume of the left ventricle lies bellow the minimum reference values [16]. This inconsistency in ventricle volumes between VHP data and *in vivo* reference studies may be due to the stopped heart in VHP dataset. An underestimated volume of the left ventricle should be accounted in further investigations. The left ventricle volume may be extended by thickening the walls of left ventricle.

The adaptive unstructured tetrahedral mesh was constructed using Delaunay triangulation algorithm from CGAL Mesh library [17]. The maximum mesh size is 3 mm, the minimum mesh size in the vicinity of heart boundaries and material interfaces is 1 mm. The computational mesh consists of 367 318 tetrahedra and 77 953 vertices (Figure 6). Upscaling of multi-labeled segmented image was used to improve the resolution of input data.

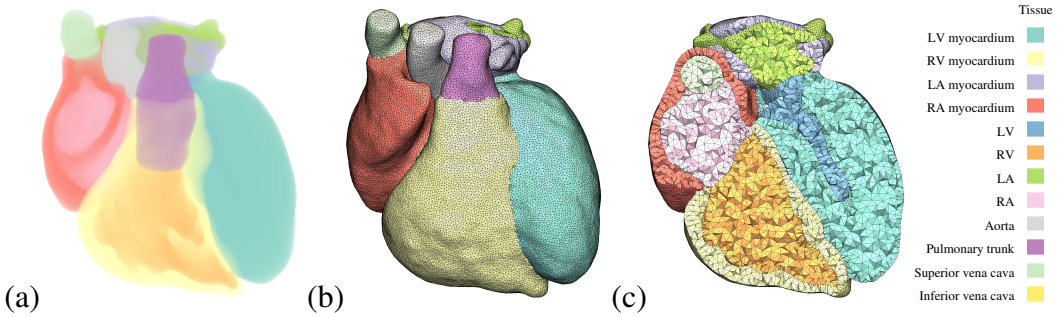


Figure 6: Unstructured mesh for VHP heart: (a) translucent 3D model; (b) triangular surface mesh; (c) volume cut of the tetrahedral mesh.

The cardiac mesh may be used in electrophysiology modelling. The proposed segmentation and mesh generation pipeline may be applied to patient's CT/MRI images.

4 CONCLUSIONS

The work addresses several segmentation techniques for generation of individualized computational domains on the basis of medical imaging datasets. Two algorithms were proposed for automatic vascular network segmentation and user-guided cardiac segmentation. Several

examples of numerical modeling applications are presented in [8, 9].

ACKNOWLEDGMENTS

The work was supported by the Russian Science Foundation (RSF) grant 14-31-00024.

REFERENCES

[1] L. Grady, Fast, quality, segmentation of large volumes – Isoperimetric distance trees. A. Leonardis, H. Bischof, A. Pinz eds. *Computer Vision – ECCV 2006*, Springer-Verlag, Berlin, Heidelberg, **3953**, 449–462, 2006.

[2] A.F. Frangi, W.J.Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering. W.M. Wells, A. Colchester, S.L. Delp eds. *Medical Image Computing and Computer-Assisted Interventation – MICCAI'98*, Springer Verlag, Berlin, Germany, **1496**, 130–137, 1998.

[3] G. Lines, M. Buist, P. Grottum, et al., Mathematical models and numerical methods for the forward problem in cardiac electrophysiology. *Computing and Visualization in Science*, **5**, 215–239, 2002.

[4] M. Holtzman-Gazit, R. Kimmel, N. Peled, D. Goldsher, Segmentation of thin structures in volumetric medical images. *IEEE Transactions on Image Processing*, **15**, 354–363, 2006.

[5] A.G. Radaelli, J. Peiró, On the segmentation of vascular geometries from medical images. *International Journal for Numerical Methods in Biomedical Engineering*, **26**, 3–34, 2010.

[6] S.Y. Yeo, X. Xie, I. Sazonov, P. Nithiarasu, Segmentation of biomedical images using active contour model with robust image feature and shape prior. *International Journal for Numerical Methods in Biomedical Engineering*, **30**, 232–248, 2014.

[7] A.A. Danilov, D.V. Nikolaev, S.G. Rudnev, et al., Modelling of bioimpedance measurements: unstructured mesh application to real human anatomy. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **27**, 431–440, 2012.

[8] Yu.V. Vassilevski, A.A. Danilov, T.M. Gamilov, et al., Patient-specific anatomical models in human physiology. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **30**, 185–201, 2015.

[9] A. Danilov, Yu. Ivanov, R. Pryamonosov, Yu. Vassilevski, Methods of graph network reconstruction in personalized medicine. *International Journal for Numerical Methods in Biomedical Engineering*, doi: 10.1002/cnm.2754, 2015.

[10] H.A. Gratama van Andel, H.W. Venema, G.J. Streekstra, et al., Removal of bone in CT angiography by multiscale matched mask bone elimination. *Medical Physics*, **34**, 3711–3723, 2007.

[11] T.S. Yoo, M.J. Ackerman, W.E. Lorensen, et al., Engineering and algorithm design for an image processing API: A technical report on ITK – The Insight Toolkit. J.D. Westwood, et al. eds. *Proceedings of Medicine Meets Virtual Reality 02/10*, Amsterdam: IOS Press, 586–592, 2002.

[12] The National Library of Medicine's Visible Human Project, http://www.nlm.nih.gov/research/visible/.

[13] R. Achanta, A. Shaji, K. Smith, et al., SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 2274–2282, 2012.

[14] P.A. Yushkevich, J. Piven, H.C. Hazlett, et al., User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, **31**, 1116–1128, 2006.

[15] A.M. Maceira, S.K. Prasad, M.Khan, D.J. Pennell, Reference right ventricular systolic and diastolic function normalized to age, gender and body surface area from steady-state free precession cardiovascular magnetic resonance. *European Heart Journal*, **27**, 2879–2888, 2006.

[16] P.A. Cain, R. Ahl, E. Hedstrom, et al., Age and gender specific normal values of left ventricular mass, volume and function for gradient echo magnetic resonance imaging: a cross sectional study. *BMC Medical Imaging*, **9**, doi: 10.1186/1471-2342-9-2, 2009.

[17] L. Rineau, M. Yvinec, A generic software design for delaunay refinement meshing. *Computational Geometry*, **38**, 100–110, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A CONSISTENTLY LINEARISED SOLID MECHANICS-BASED MESH DEFORMATION TECHNIQUE FOR HIGH ORDER CURVED ELEMENTS

Roman Poya[1], **Ruben Sevilla**[1], **and Antonio J. Gil**[1]

[1]Zienkiewicz Centre for Computational Engineering, College of Engineering, Swansea University Bay Campus, Crymlyn Burrows, Swansea SA1 8EN, UK

e-mail: {R.Poya, R.Sevilla, A.J.Gil}@swansea.ac.uk

Keywords: Mesh deformation, curved elements, high order methods, distortion measure, solid mechanics analogy.

Abstract. *In this work, a novel solid mechanics-based mesh deformation technique for high order curved elements is presented. The technique falls under the a posteriori curved mesh generation category, where higher order nodes are placed on a linear mesh and the geometry is then deformed to conform to the exact CAD boundary. In contrast to the existing a posteriori approaches in the literature such as the techniques based on the inclusion of residual stresses, parametrised and varying material constants, regularisation and smoothing of curved meshes, in this work, a rather consistent solid mechanics approach is followed. This implies, that the underlying Euler-Lagrange equations that need to be solved for, emerge from an energy principle, with well-defined internal energies constructed for an hyperelastic system, which are subsequently, consistently linearised. Depending on the geometrical parameterisation, the approach guarantees better mesh quality and lower condition number for the system of equations. Furthermore, due to the introduction of independent invariants emanating from fibre, surface and volume mappings, the essential mesh distortion measures are encoded in the formulation. The paper proves that for for two-dimensional elements such as triangles and quadrilaterals, not all the distortion measures can be independent. An example of materially instable internal energy is provided to pinpoint the importance of a consistent formulation in the context of highly stretched boundary layer meshes.*

1 INTRODUCTION

The use of curved elements is nowadays accepted to be crucial in order to fully exploit the advantages of high-order discretisation methods, but until relatively recently, the challenge of automatically generating high-order curvilinear meshes has been an obstacle for the widespread application of high-order methods [Vincent and Jameson(2011)]. Methods to produce high-order curvilinear meshes are traditionally classified into *direct* methods and *a posteriori* methods [Dey, O'Bara, and Shephard]. Direct methods build the curvilinear high-order mesh directly from the CAD boundary representation of the domain whereas a posteriori approaches rely on mature low-order mesh generation algorithms to produce an initial mesh that is subsequently curved using different techniques, such as local modification of geometric entities, solid mechanics analogies or optimisation.

Within the category of a posteriori approaches, the solid mechanics analogy first proposed in [Persson and Peraire] has become increasingly popular. The main idea is to consider the initial, low-order, mesh as the undeformed configuration of an elastic solid. High-order nodal distributions are then inserted into all of the elements and then the nodes over element edges/faces in contact with the curved parts of the boundary are projected onto the true CAD boundary. The displacement required to move the nodes onto the true boundary is interpreted as an essential boundary condition within the solid mechanics analogy. The solution of the elastic problem provides the desired curvilinear mesh as the deformed configuration. The initial approach proposed in [Persson and Peraire] used a non-linear neo-Hookean constitutive model. Several attempts to reduce the computational cost of this approach have been proposed based on a linear elastic analogy, see [Xie, Sevilla, Hassan, and Morgan]. It is clear that when large deformations are induced to produce the deformed curvilinear high-order mesh, a linear elastic model can result in non-valid elements due to the violation of the hypothesis of small deformations. In order to alleviate this problem, it is possible to split the desired (potentially large) displacement of boundary nodes into smaller load increments. Other approaches to increase the robustness of the linear elastic analogy have been recently introduced, see for instance [Moxey, Ekelschot, Keskin, Sherwin, and Peiró], where *pseudo* thermal effects are introduced.

In this work, a novel a posteriori solid mechanics-based mesh deformation technique for high order curved elements is presented. In contrast to the existing a posteriori approaches in the literature such as the techniques based on the inclusion of residual stresses, parametrised and varying material constants, regularisation and smoothing of curved meshes, in this work, a rather consistent solid mechanics approach is followed. This implies, that the underlying Euler-Lagrange equations that need to be solved for, emerge from an energy principle, with well-defined internal energies constructed for an hyperelastic system, which are subsequently, consistently linearised. The theoretical and computational framework is presented in the next two sections, respectively and finally followed by the numerical examples sections.

2 NO-LINEAR CONTINUUM MECHANICS

Let us consider the motion of a continuum from its initial undeformed (or material) configuration $\Omega_0 \subset \mathbb{R}^d$, with boundary $\partial\Omega_0$ and outward unit normal $\boldsymbol{n}_0$, into its final deformed (or spatial) configuration $\Omega \subset \mathbb{R}^d$, with boundary $\partial\Omega$ and outward unit normal $\boldsymbol{n}$, where d represents the number of spatial dimensions. In the context of curved mesh generation, the initial (undeformed) configuration Ω_0 represents a linear mesh with planar faces (edges in two dimensions) and the final (deformed) configuration Ω represents the final curved high-order mesh, as illustrated in Figure 1. The motion is described by a mapping $\boldsymbol{\phi}$ which links a material particle

from material configuration $\boldsymbol{X}$ to spatial configuration $\boldsymbol{x}$ according to $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{X})$.

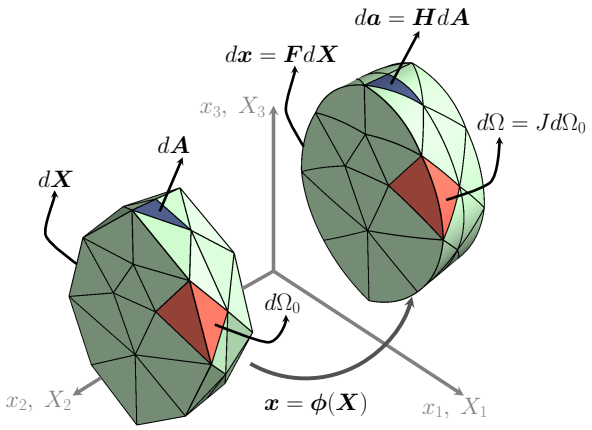


Figure 1: Deformation map of a continuum and illustration of the strain measures $\boldsymbol{F}$, $\boldsymbol{H}$ and J.

The following well-known strain measures can be introduced, namely the two-point deformation gradient tensor or fibre-map $\boldsymbol{F}$, the two-point co-factor or adjoint tensor or area map $\boldsymbol{H}$ and the Jacobian J or volume-map

$$\boldsymbol{F} = \boldsymbol{\nabla}_0 \boldsymbol{\phi} = \frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{X}}, \qquad \boldsymbol{H} = J \boldsymbol{F}^{-T}, \qquad J = \frac{1}{3} \boldsymbol{H} : \boldsymbol{F} \tag{1}$$

where $\boldsymbol{\nabla}_0$ denotes the gradient with respect to material coordinates. The fundamental strain measures $\{\boldsymbol{F}, \boldsymbol{H}, J\}$, also illustrated in Figure 1, encode the essential modes of deformation, based on which a complete set of indepdendent mesh quality/distortion measure can be defined. Kinematically any other quality measure would be a combination of these three measures.

3 A CONSISTENT INCREMENTALLY LINEARISED APPROACH

To guarantee and/or maintain essential mathematical requirements such as objectivity and polyconvexity for the linearised strain energy density, a linearised solid mechanics approach must emanate from an underlying non-linear variational principle, as the notion of objectivity and polyconvexity cannot be invoked in small strains. This is typically achieved by consistent linearisation of the non-linear total potential energy through a Taylor series expansion. To illustrate this, let us consider the total potential energy in non-linear system, cast in the form of an iterative (Newton-Raphson) scheme

$$\Pi(\boldsymbol{\phi}^{\star}_{n+1}) = \inf_{\boldsymbol{\phi}_{n+1} \in \boldsymbol{\mathcal{V}}_{n+1}} \left\{ \int_{\Omega_0} \Psi(\boldsymbol{C}_{n+1}) \, d\Omega_0 \right\} \tag{2}$$

where $\boldsymbol{\mathcal{V}}_{n+1} = \left\{ \boldsymbol{\phi}_{n+1} \in [\mathcal{H}^1(\Omega_0)]^d \ : \ \boldsymbol{\phi}_{n+1}(\boldsymbol{X}) = \bar{\boldsymbol{x}}_{n+1} \text{ on } \partial\Omega_0 \right\}$ and $\boldsymbol{x}_{n+1} = \boldsymbol{\phi}_{n+1}(\boldsymbol{X})$ is the position vector of the material points at increment $n+1$, which can be evaluated through an incremental displacement $\boldsymbol{u}$ superimposed on the deformed configuration at increment n, i.e. $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{u}$. At increment n, the current position vector $\boldsymbol{x}_n$, the state of deformation gradient $\boldsymbol{F}_n$ and, subsequently, the Cauchy-Green strain $\boldsymbol{C}_n$ are fully known. In a non-linear regime, the motion of the continuum from n to $n+1$ is solved iteratively, as the amount of displacements,

the state of deformation gradient $\boldsymbol{F}_{n+1}$ and the Cauchy-Green strain $\boldsymbol{C}_{n+1}$ cannot be determined explicitly.

However, in the context of high-order curved mesh generation, it is convenient to approximate (2) through a Taylor series expansion of the form

$$\Pi_u(\boldsymbol{u}^\star) = \inf_{\boldsymbol{u}\in\boldsymbol{\mathcal{U}}} \left\{ \int_{\Omega_0} \left(D\Psi(\boldsymbol{C}_n)[\boldsymbol{u}] + \frac{1}{2} D^2\Psi(\boldsymbol{C}_n)[\boldsymbol{u};\boldsymbol{u}] \right) d\Omega_0 \right\}. \tag{3}$$

where $\boldsymbol{\mathcal{U}} = \left\{ \boldsymbol{u} \in [\mathcal{H}^1(\Omega_n)]^d \ : \ \boldsymbol{u} = \bar{\boldsymbol{u}} \text{ on } \partial\Omega_n \right\}$. Notice that the first term in Taylor series expansion i.e. $\Psi(\boldsymbol{C})$ would be a constant term describing the state of strain energy density at increment n, which vanishes at the moment of computing the stationary point of (3). Certainly, embedded in the definition of the new total potential energy (Π_{lin}) in (3) are the first and second directional derivatives of the non-linear total potential energy. The spatial form of the linearised total potential energy can now be obtained as

$$\Pi_u(\boldsymbol{u}^\star) = \inf_{\boldsymbol{u}\in\boldsymbol{\mathcal{U}}} \left\{ \int_{\Omega_n} \left(\boldsymbol{\sigma}_n : \boldsymbol{\varepsilon}_n(\boldsymbol{u}) + \frac{1}{2}\, \boldsymbol{\varepsilon}_n(\boldsymbol{u}) : \boldsymbol{c}_n : \boldsymbol{\varepsilon}_n(\boldsymbol{u}) + \frac{1}{2}\, \boldsymbol{\sigma}_n : \left((\nabla_n \boldsymbol{u})^T (\nabla_n \boldsymbol{u}) \right) \right) d\Omega_n \right\} \tag{4}$$

where the subscript n denotes the state of deformation, stresses, tangent elasticity and the volume at increment n, namely $\boldsymbol{\varepsilon}_n$, $\boldsymbol{\sigma}_n$, $\boldsymbol{c}_n$ and Ω_n. In addition, ∇_n represents the spatial gradient operator at increment n. The stationary condition of (4), obtained after the linearisation with respect to the virtual displacement $\boldsymbol{v}$, leads to the principle of virtual work

$$D\Pi_u(\boldsymbol{u}^\star)[\boldsymbol{v}] = \int_{\Omega_n} \left(\underbrace{\boldsymbol{\sigma}_n : \boldsymbol{\varepsilon}_n(\boldsymbol{v})}_{\mathcal{R}_n} + \underbrace{\boldsymbol{\varepsilon}_n(\boldsymbol{u}) : \boldsymbol{c}_n : \boldsymbol{\varepsilon}_n(\boldsymbol{v})}_{\mathcal{C}_n} + \underbrace{\boldsymbol{\sigma}_n : ((\nabla_n \boldsymbol{u})^T (\nabla_n \boldsymbol{v}))}_{\mathcal{G}_n} \right) d\Omega_n = 0. \tag{5}$$

It is worth noting that, in the right hand side of (5), the first term $\mathcal{R}_n$ corresponds to the residual stresses, the second term $\mathcal{C}_n$ to the linearised constitutive stiffness term and the last term $\mathcal{G}_n$ to the geometric stiffness term. The emergence of the geometric stiffness term is due to consistent linearisation of the non-linear total potential energy, which would not have appeared, had the starting point not been chosen to correspond to a non-linear total potential energy. As will be seen in the numerical examples, in the context of high-order curved mesh generation, the geometric stiffness term, stiffens the interior elements of the computational mesh against severe distortion, hence producing meshes with better quality. Note that unlike in the non-linear analysis, since (5) is linear in $\boldsymbol{u}$, a further linearisation is not required. It is evident that, if a single increment is used to reach the final configuration, i.e. when $n = 0$, the equations of classical linear elasticity are recovered.

4 QUALITY MEASURES

In a standard high-order finite element formulation, measures involving the Jacobian of the isoparametric mapping have been extensively used, in particular the so-called scaled Jacobian. This measure only quantifies volumetric deformations and alternative measures that exploit different modes of deformation and account for shape, skewness and degeneracy of elements can also be considered. However, it is worth noting that not all of these quality measures can be regarded as independent quantities. From a solid mechanics point of view, only three independent isotropic and two independent transversely isotropic quality measures for a generic

element e can be introduced

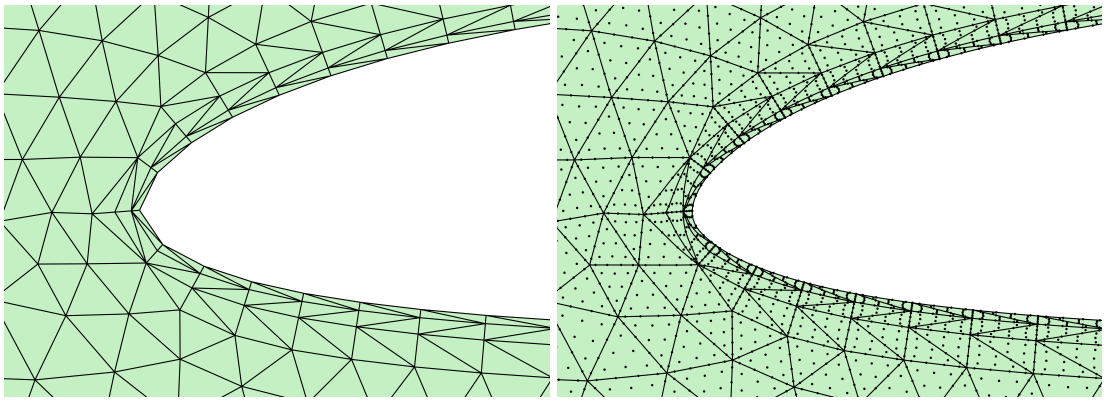
$$Q_j^e = \sqrt{\frac{\min_{\boldsymbol{\xi} \in R} \{I_j\}}{\max_{\boldsymbol{\xi} \in R} \{I_j\}}} \qquad \text{for } j = 1, \dots, 5, \tag{6}$$

where R denotes the reference element employed in the isoparametric formulation, with local coordinates $\boldsymbol{\xi}$. If necessary, further quality measures can be obtained through a linear combination of the invariants I_j which will be independent of the geometrical parametrisation.

5 NUMERICAL EXAMPLES

This section presents a few selected numerical examples. For all the examples, material parameters are chosen as $E = 10^5$ for Young's modulus, $E_A = \frac{5E}{2}$ for transversely isotropic bending modulus and $G_A = \frac{E}{2}$ for transversely isotropic shear modulus and the Poisson's ratio is varied within the interval [0.001,0.495].

The first example considers an anisotropic boundary layer mesh around the SD7003 aerofoil with a stretching level of 25 in the boundary layer. The detailed view near the leading edge of the initial linear triangular and the high order mesh with a degree approximation of $p = 5$ is shown in Figure 2 having 27,410 nodes. The curvilinear mesh is obtained using 20 load increments of a consistently linearised version of Neo-Hookean model. The minimum quality measures characterising fibre, surface and volume deformation are, $Q_1 = 0.991$, $Q_2 = 0.991$ and $Q_3 = 0.982$, respectively, numerically confirming that the first two quality measures (related to fibre and surface deformation) are identical in two-dimension.

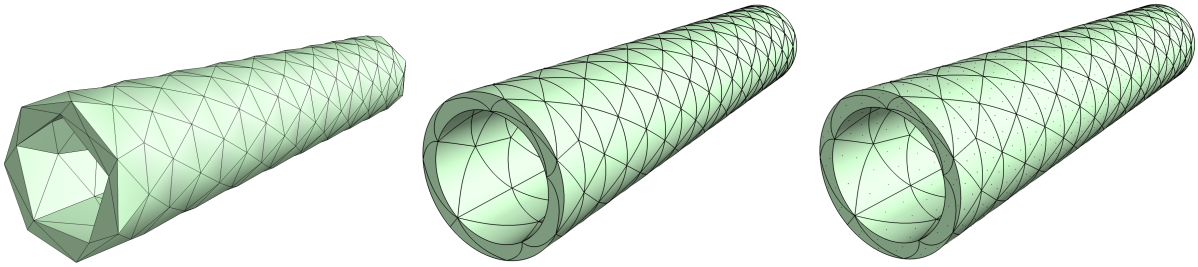


(a) Linear mesh. (b) High-order mesh with p=5.

Figure 2: Boundary layer mesh around an aerofoil.

The next example considers an isotropic tetrahedral mesh around a cylinder. The initial linear and the high order mesh with a degree approximation of $p = 4$ is shown in Figure 3 having 38,355 nodes. The curvilinear mesh is obtained using 30 load increments of a consistently linearised version of nearly incompressible Mooney-Rivlin type material model. The minimum quality measures characterising fibre, surface and volume deformation are, $Q_1 = 0.965$, $Q_2 = 0.928$ and $Q_3 = 0.891$, respectively. Note that nodes lying on the planar faces of cylinder require in-plane translation.

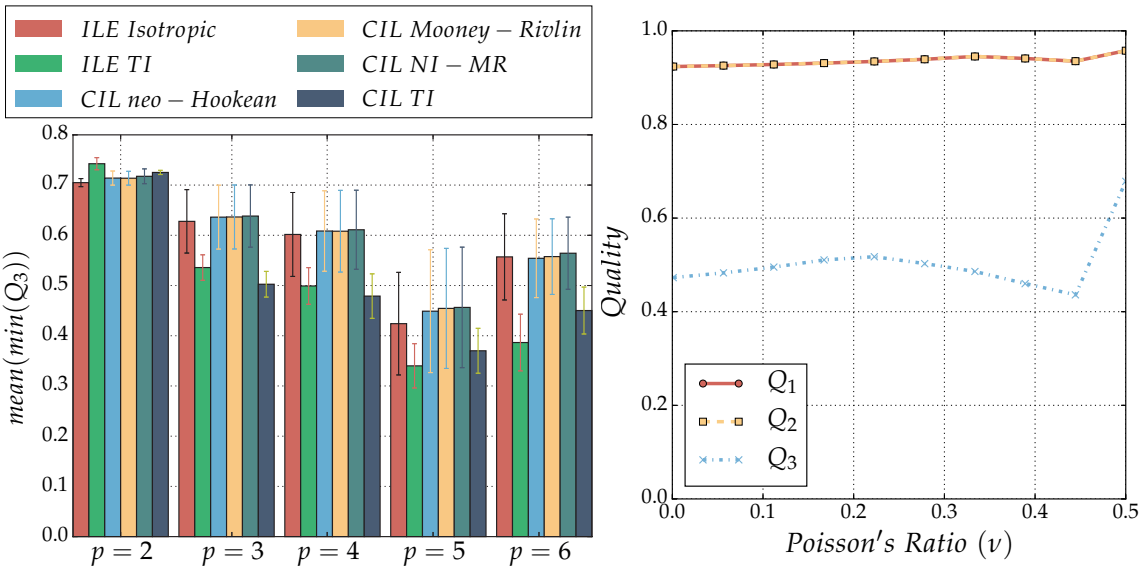
Finally 5 pin-points two critical aspects of solid mechanics based curved mesh generation approaches, namely, the importance of a well-defined (i.e. polyconvex) material model depicted in 5a (note that all material models apart form ILE TI and CIL TI models are well defined;



(a) Linear surface mesh. (b) $p = 4$ surface mesh (c) $p = 4$ mesh surface showing nodes

Figure 3: Isotropic mesh around cylinder.

notice the drop in mesh quality associated with these material models), and the equivalence of fibre quality and surface quality i.e. $Q_1 = Q_2$ for two-dimensional problems shown in 5b; both for anisotropic boundary layer mesh. For the explicit forms of the material models refer to [Poya, Sevilla, and Gil].



(a) Mean value and standard deviation of the minimum scaled Jacobian of the generated meshes as a function of the Poisson's ratio for different materials and degrees of approximation.

(b) Different quality measures of the generated meshes with p=4 as a function of the Poisson's ratio for the ILE isotropic approach and using five load increments.

6 CONCLUSIONS

A consistently linearised solid mechanics based framework for the generation of high-order curvilinear meshes has been presented. The derivation of the approach, based on energy principles, is used to propose mesh quality measures based on independent invariants of the strain energy density. Of the three isotropic quality measures proposed, Q_3 is the most impactful indicator, which corresponds to the so-called scaled Jacobian traditionally used by the high-order mesh generation community.

In terms of the material parameters, the use of a Poisson's ratio near the incompressible limit is generally advised in order to maximise the quality of the resulting mesh. For isotropic meshes, a low number of increments (e.g. five increments) is typically sufficient to obtain the

maximum possible quality, whereas for highly stretched meshes and for high-orders of approximation (i.e. $p > 4$) a higher number (e.g. 40 increments) is needed to obtain good quality meshes.

ACKNOWLEDGMENTS

The first author acknowledges the financial support received through The Erasmus Mundus SEED program. The second author gratefully acknowledges the financial support provided by the Sér Cymru National Research Network in Advanced Engineering and Materials. The third author acknowledges the financial support received through The Leverhulme Prize awarded by The Leverhulme Trust, UK.

REFERENCES

[Dey, O'Bara, and Shephard] Dey S, O'Bara RM, Shephard MS (2001) Towards curvilinear meshing in 3D: the case of quadratic simplices. Computer-Aided Design 33(3):199 – 209

[Moxey, Ekelschot, Keskin, Sherwin, and Peiró] Moxey D, Ekelschot D, Keskin U, Sherwin S, Peiró J (2014) A thermo-elastic analogy for high-order curvilinear meshing with control of mesh validity and quality. Procedia Engineering 82(0):127 – 135, 23rd International Meshing Roundtable (IMR23)

[Persson and Peraire] Persson PO, Peraire J (2009) Curved mesh generation and mesh refinement using lagrangian solid mechanics. In: Proceedings of the 47th AIAA Aerospace Sciences Meeting and Exhibit, AIAA

[Poya, Sevilla, and Gil] Poya R, Sevilla R, Gil A (2016) A unified framework for a posteriori high-order curved mesh generation using solid mechanics. Under Review

[Vincent and Jameson(2011)] Vincent PE, Jameson A (2011) Facilitating the adoption of unstructured high-order methods amongst a wider community of fluid dynamicists. Math Model Nat Phenom 6(3):97–140

[Xie, Sevilla, Hassan, and Morgan] Xie ZQ, Sevilla R, Hassan O, Morgan K (2013) The generation of arbitrary order curved meshes for 3D finite element analysis. Computational Mechanics 51(3):361–374

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

VALIDATION OF MODELING ASSUMPTIONS FOR THE BUILDUP SIMULATION OF LASER BEAM MELTING ON THE BASIS OF THE RESIDUAL STRESS DISTRIBUTION

F. Bayerlein[1], C. Zeller[1], M. Wunderer[1], J. Weirather[1], M. Schmid[1], C. Seidel[1], M. F. Zaeh[1], R. Hessert[2], G. Schlick[2], T. Uihlein[2], and M. Hofmann[3]

[1]Technical University Munich
Institute for Machine Tools and Industrial Management
Boltzmannstraße 15, 85748 Garching, Germany
e-mail: fabian.bayerlein@iwb.tum.de

[2] MTU Aero Engines AG
Dachauer Straße 665, 80995 Munich, Germany
e-mail: roland.hessert@mtu.de

[3] Heinz Maier-Leibnitz Zentrum (MLZ)
Technical University Munich
Lichtenbergstraße 1, 85748 Garching, Germany
e-mail: michael.hofmann@frm2.tum.de

Keywords: Laser beam melting, computational methods, engineering sciences, neutron diffraction, X-ray diffraction, residual stresses

Abstract. *The growing market for additive manufacturing processes such as laser beam melting (LBM) poses new challenges. With more difficult product requirements, the need for a deeper understanding of the resulting stress states and the underlying physical principles increases. To gain a deeper understanding of the residual stress states of LBM-manufactured parts, simulations of the build-up process were carried out and validated by means of neutron diffraction (ND), X-ray diffraction (XRD) as well as incremental hole-drilling (IHD). The gathered data is also intended to serve as a validation case for other simulation models and tools.*

1 Introduction

Laser beam melting is an additive manufacturing technique to produce dense metal parts through layer-wise powder melting and re-solidification by means of a focused laser beam. This leads to significant temperature gradients both spatially and temporally, which in turn cause residual stresses through thermal strains. Today, the knowledge about the correlation between part geometry, resulting temperature gradients during the manufacturing process and, ultimately, residual stress states after the end of the build process is not sufficient and only partly validated [1, 2]. This issue is a considerable challenge for the component design of dimensionally accurate parts [3, 4]. Currently, a lot of process know-how and expert knowledge is necessary to pre-warp the geometry in such a way that the manufactured result matches the target geometry. An iterative approach to meet this challenge before the manufacturing stage by pre-warping the geometry based on simulation results is described in [5]. In recent time, further simulation models have been increasingly developed to predict residual stresses and distortions of additively manufactured parts [6, 7, 8, 9, 10].
This paper presents the validation results of the simulation model described in [5, 11]. Moreover, to create a general understanding of the correlation between temperature gradients and residual stresses in LBM, a simple cuboid geometry was selected. The presented investigations focus on the nickel-base superalloy Inconel 718, which is highly relevant to e. g. the aerospace industry [12, 13]. Validation of the simulation model was already carried out for both a tool steel 1.2709 [14] and the AlSi12 material system [15].

Neutron diffraction measurements of a massive cuboid geometry were used in order to understand the stress state within the bulk volume. In addition, incremental hole-drilling as well as X-ray diffraction measurements were carried out to validate residual stresses close to the surface. An alternative approach to estimate residual stresses in the interior of additively manufactured parts using digital image correlation is presented in [2]. A more in-depth analysis of the microstructure and mechanical behavior of Inconel 718 fabricated by selective laser melting is described in [16, 17]. Additional details about the experiments are summarized in Subsection 2.1.

For the simulation presented in this manuscript, the LBM process is simulated by a sequentially coupled finite element simulation of the thermal and thermomechanical processes. Due to the complexity of the real process, simplifications are necessary for the simulation to calculate residual stresses and distortions on the scale of industry relevant parts within a reasonable time frame. While real parts comprise thousands of layers consisting of thousands of hatches, several of those layers are combined into so-called layer compounds for the simulation [18]. Furthermore, a simultaneous heat input to the whole layer is used, thereby neglecting the distortions and residual stresses caused by the single weld seams [11]. Subsection 2.2 contains additional information on the used simulation model.

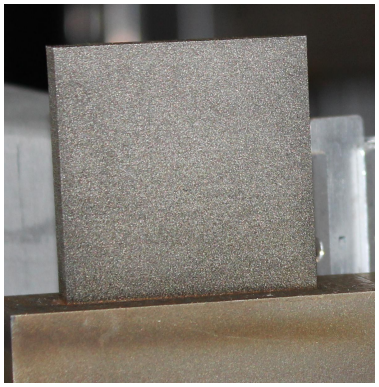
2 Methodology

The presented results are intended to help to understand the distribution of residual stresses in a simple additively manufactured cuboid structure in order to improve the understanding and to serve as a way of validating the results of corresponding simulations. The necessary experimental methods and analyses are presented in this section as well as the simulation model that is to be validated.

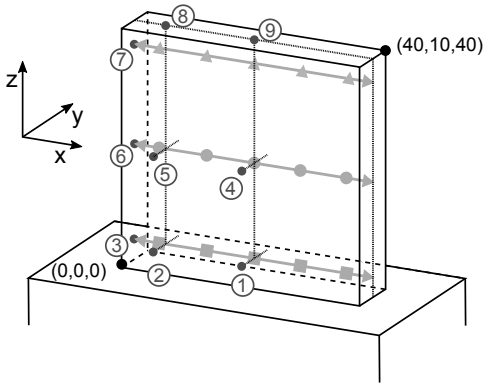
2.1 Experimental setup and methods

The geometry used for the validation is shown in Figure 1. Specimen manufacturing was carried out on an EOS Eosint M280 LBM machine calibrated and maintained for aerospace production use located at MTU Aero Engines AG in Munich. The machine uses a $400\,\mathrm{W}$ Nd:YAG fiber laser with a spot size of approximately $80\,\mu\mathrm{m}$. Laser focusing and positioning is achieved by an F-Theta lens and two Galvanometer scanning mirrors. Argon was used as a shielding gas and the oxygen concentration within the build chamber is controlled to a maximum of $0.1\,\%$.

The powder material for the experiments was Inconel 718 with a particle diameter between $15\,\mu\mathrm{m}$ and $45\,\mu\mathrm{m}$ provided by Oerlikon Metco. The EOS standard parameter set for this material and $20\,\mu\mathrm{m}$ layer thickness originally developed for the Eosint M270 generation was used. This includes, among others, a stripe pattern for the exposure strategy.



(a) Specimen on build plate within the experimental setup



(b) Schematic depiction of the cuboid geometry with the build plate. The lines with markers represent the location for the neutron diffraction measurements, while the numbered dots indicate the location of X-ray and hole-drilling measurements.

Figure 1: The used specimen, a cuboid geometry ($40 \times 40 \times 10\,\mathrm{mm}^3$). The directions are referred to in the following as x: longitudinal, y: transversal and z: normal.

IHD and XRD techniques were applied in side and top surfaces of an LBM manufactured cuboid sample to characterize the residual stress depth distributions. For the IHD technique, high-speed drilling equipment was used to perform a cylindrical blind hole. This technique is recommended since minor drilling-induced plastic deformations occur and a satisfying hole shape can be achieved. Surface strain relief was measured by three-element strain gage rosettes (CEA-06-062UM-120, Measurements Group) in depth increments of 0.02 to $0.06\,\mathrm{mm}$, up to about $1.2\,\mathrm{mm}$ below the surface. Smaller steps were carried out close to the surface to access stress gradients more precisely. The typical hole diameter was about $1.9\,\mathrm{mm}$. Residual stress evaluation was carried out according to the differential method, using the elastic constants $E = 190\,\mathrm{GPa}$ and $\nu = 0.34$ for the top surface and $E = 175\,\mathrm{GPa}$ and $\nu = 0.34$ for the side surfaces of the rectangular sample.

XRD residual stress analysis was combined with local electrolytic layer removal to obtain residual stress profiles. Stress relaxation due to layer removal was neglected since the affected region was small and no significant relaxation effects could be expected. Lattice deformations of the Ni planes were measured on a modified ψ-diffractometer (Xstress 3000 G2, Stresstech Group) for 13 ψ-angles between $-45°$ and $45°$ using Mn Kα radiation. Residual stresses were calculated for the plain stress condition using the X-ray elastic constants $E\,\{311\} = 199\,\mathrm{GPa}$

and $\nu\{311\} = 0.33$ for IN718 matrix material [19]. In addition to the macroscopic residual stresses, the full width at half maximum (FWHM) values of the X-ray diffraction peaks were also recorded; they indicate near surface microstructural changes in the material due to the LBM process.

The neutron diffraction measurements were performed at the STRESS-SPEC instrument. For the neutron diffraction experiment a wavelength of $\lambda = 1.55\,\text{Å}$ was selected using the Si (400) monochromator, which allowed to measure the Nickel $\{311\}$ reflection at a scattering angle of about $2\theta_S \approx 91°$. The gauge volume was defined using a $2 \times 2\text{mm}^2$ slit in the incoming beam and a radial collimator having a $2\,\text{mm}$ field of view in the diffracted beam. In total three lines at half width ($y = 5\,\text{mm}$) of the block and at different positions (e. g. at $z = 5\,\text{mm}$, $z = 20\,\text{mm}$ and $z = 38\,\text{mm}$) along the build direction z were measured (cf. Figure 1b on the preceding page). Each line consisted of five measurement positions along the long axis of the IN718 block with the exact positions in the x-direction at $2\,\text{mm}$, $12\,\text{mm}$, $20\,\text{mm}$, $28\,\text{mm}$ and $38\,\text{mm}$, respectively.

The strain is measured in the direction of the scattering vector, which bisects the incident and the diffracted beam. Since stress is a tensor, measurements are generally required in six orientations to completely determine the stress state at any point in the sample. However, in case of the built IN718 cuboid, we inferred that the principal stress axes roughly coincide with the principal axes x, y and z of the cuboid. Additionally, the effect of assuming the principal stresses is not relevant for the comparisons carried out in this study, as the stresses in three perpendicular directions are correctly gathered, they may just not account for all of the stress in the sample [20]. As the main goal of this study is the validation of a simulation model and more exactly, the stresses along the principal axes of the cuboid geometry, no systematical error is introduced. Thus, the stresses can be calculated from the measured strains using diffraction elastic constants $E\{311\}$ and $\nu\{311\}$.

As in our case no stress free reference sample was available, the stresses were calculated using the mechanical boundary condition that the near surface stresses in x-direction should decline to zero. The possible error of this assumption relates to a shift in the y-axis of the plots in Figure 3 on page 7 but is not depicted in the plots.

2.2 Simulation

The used simulation model was developed at the Institute for Machine Tools and Industrial Management (*iwb*) and contains several modeling assumptions necessary to allow the simulation of the highly complex laser beam melting process (cf. [5, 11] for additional details).
The position of the nodes of the computational mesh, for example, is required to be restricted to discrete levels in build-up direction, similar to the real layer wise manufacturing process. With real parts comprising thousands of layers, abstractions are made by summarizing multiple real layers in so-called layer compounds. Modeling every single layer in the simulation would currently result in an unacceptably long calculation time. It is however not considered necessary due to the fact that the layers close to each other share the same temperature history and are therefore subject to similar thermal loading and thus expected to exhibit similar mechanical loading. The presented results were generated with a layer compound height of $0.5\,\text{mm}$ and hexahedral elements of similar size.

The simulation uses a simultaneous heat input to the whole layer and considers the effect of the heat conduction through the part, neglecting stresses introduced by the single weld seams. While the selective solidification leads to high initial stresses, considering the established subsequent manufacturing process, where multiple layers are remelted and hatch patterns are rotated

between each layer, the effect is assumed to be characterizable by a homogeneous heat input as a first step.

The heat input is realized by adjusting the temperature of the nodes on the upside of the top layer to solidus temperature. The load time corresponds to the time a point in the real process is subject to temperatures above solidus and is based on Rosenthal's solution for a moving heat source on a semi-infinite body [21].

The thermal and mechanical simulation are sequentially coupled by solving the transient heat equation and applying the calculated temperature distribution per time step as distinct load steps within the mechanical simulation. The feedback, the resulting warpage, is deemed negligible for the thermal simulation due to its limited effect size in the range of $10^{-4}\,\mathrm{m}$. As distortions of the building platform are observable in the real manufacturing process, it is necessary to take the building platform into account. The model thus not only comprises the part but also an exact representation of the used building platform.

In order to validate the simulation, a comparison of values of residual stresses was carried out.

3 Results

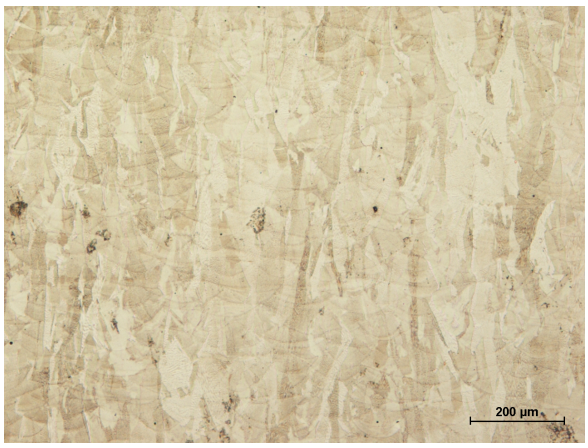
Post-process light and scanning electron microscopic images reveal the expected dense structure with columnar grains in the x-y-plane [16, 17] (cf. Fig. 2). In contrast, the x-z-plane of the part shows no predominant direction in terms of grain orientation (cf. Figure 2b on the following page).

The results of the different stress evaluations are depicted in Figure 3 on page 7 and Figure 4 on page 8. The first three plots show a comparison between the simulated and the measured stresses in the three principal directions of the cuboid geometry, along the measurement paths of the neutron diffraction measurements with the hollow markers indicating experimentally gathered values. The given error is limited to the propagation of the fitting error inherent to diffraction measurements, i. e. the error from choosing the correct reference sample for the stress-free lattice strain is not included. All errors are acceptable, with the transversal stresses exhibiting the highest uncertainty.

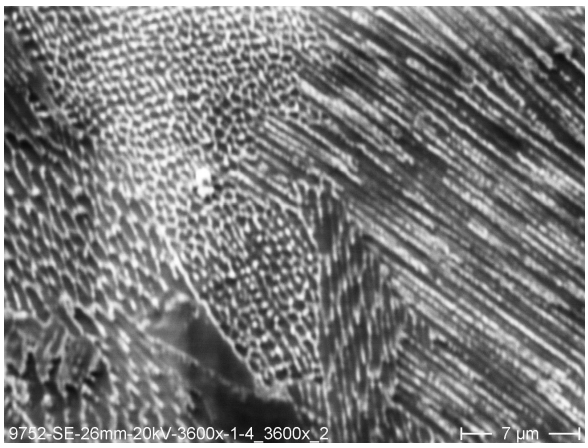
The plot for the normal stresses shows similar tendencies for all measurement paths, the respective magnitude is however overestimated by the simulation for the inner lower area of the part. In the border areas, tensile stresses near yield are both simulated and measured while the middle of the part is exhibiting compressive stresses that decline with the distance from the build plate. Both the simulation and the measurement suggest an almost neutral stress state along the upper path with the simulation predicting tensile stresses near the surface.

The longitudinal stresses exhibit the best agreement between simulation and measurements. With the exception of the central point of the middle path, the simulation results are within or close to the given experimental accuracy. The stress distribution is characterized by a low level of tensile stresses near the bottom of the part, a similar level of compressive stresses along the path at $z = 20\,\mathrm{mm}$ and a maximum at the top and in the middle of the part, exhibiting tensile stress. The boundary condition, that longitudinal stresses should disappear at the edge of the geometry in x-direction, was used to set the stress-free reference angle. Thus, the respective values naturally decline towards $x = 0$ and $x = 40\,\mathrm{mm}$.

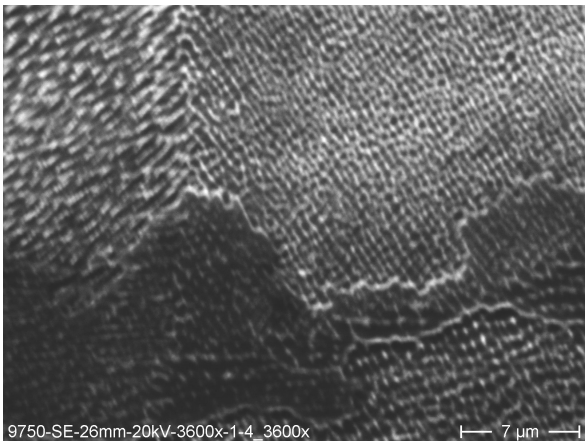
In the transversal, i. e. the y-direction, the predicted stresses follow a similar trend as the measurements, the agreement is however not sufficient, especially for the lowest scan path in the middle of the part. Additionally, the missing symmetry in combination with the high expected gradients suggests a positioning error of the gauge volume for one of the extreme



(a) Light microscope image of the x-z-plane, etched and 100 times magnified



(b) Scanning electron microscopic image of the x-z-plane, etched and 3600 times magnified



(c) Scanning electron microscopic image of the x-y-plane, etched and 3600 times magnified

Figure 2: Microstructure of the used specimen

positions ($x = 2$ or $x = 38\,\mathrm{mm}$). The stresses are rather negligible in the middle of the part but exhibit a considerable maximum of $300\,\mathrm{MPa}$ near the edge and bottom of the part.

The contour plots in Figure 4 on page 8 combine all determined values for an easier comparison. The plots are stretched to show a quadratic cross section to increase readability, with only Figure 4f having the original aspect ratio. The contour lines represent the results of the simulation described in Subsection 2.2, the values marked with a diamond indicate neutron diffraction results and the alphanumerically labeled data, marked by circles and rectangles, was provided by the IHD and XRD experiments. For most measuring points, the results of the last two methods (obtained at corresponding positions on opposite sides of the cuboid) coincide, but especially for holes close to edges of the cuboid, IHD measurements are less reliable and diverge from those of the XRD trials due to plastic deformation. In these cases, the results obtained by X-ray diffraction are used and the corresponding points are marked with a rectangle instead of a circle. Additionally, if the cross sections do not coincide with actual measurement points, a linear interpolation between the neighboring points is used. For example, no neutron diffraction result is available for the cross section at $x = 5$ and thus it is inferred from the values at $x = 2$ and $x = 11$.

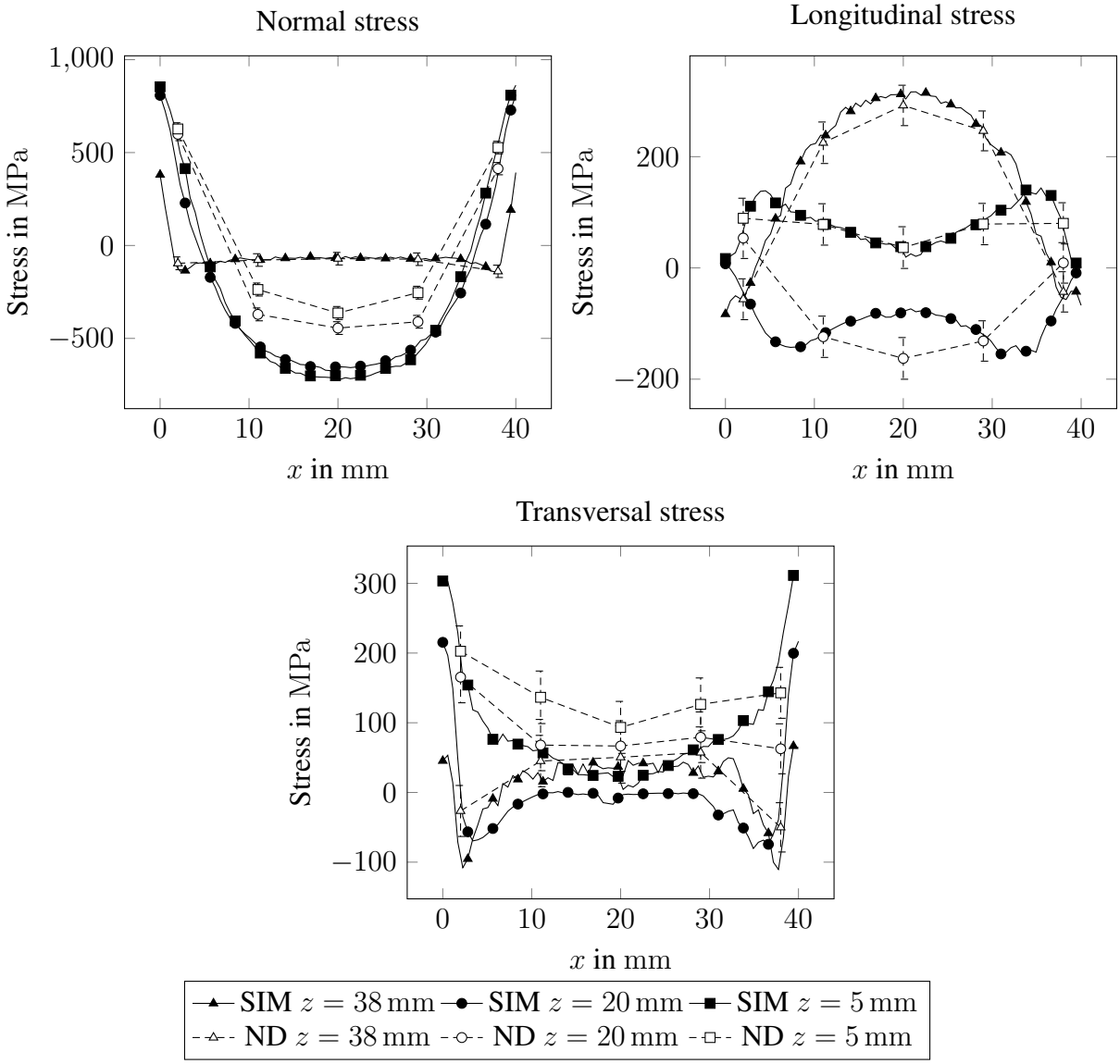
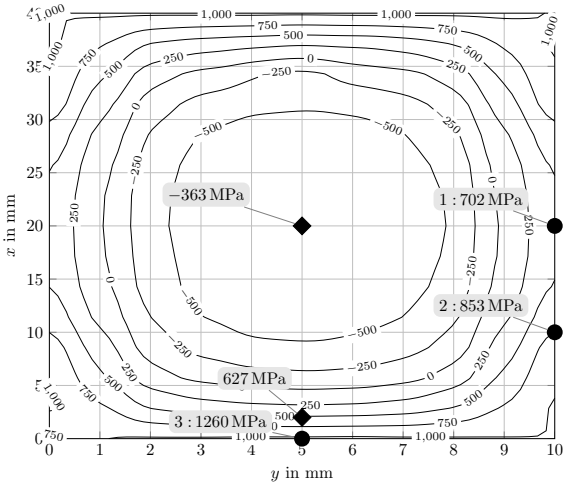
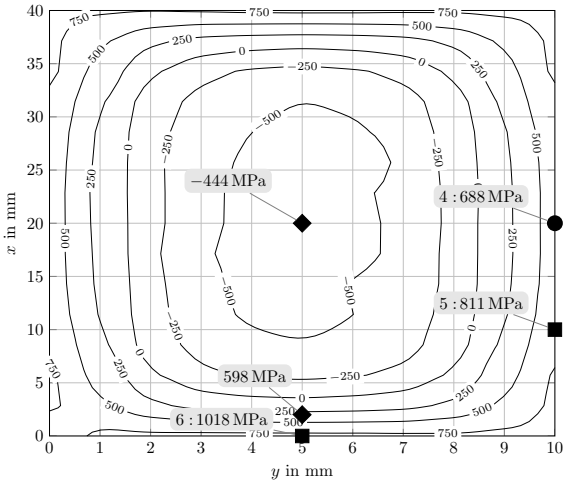


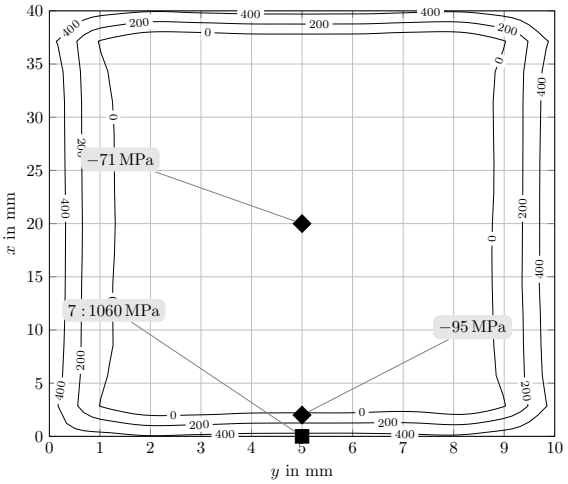
Figure 3: Comparison of simulated (SIM) and measured stresses by neutron diffraction (ND) for the three principal directions of the geometry in different build-up (z) heights



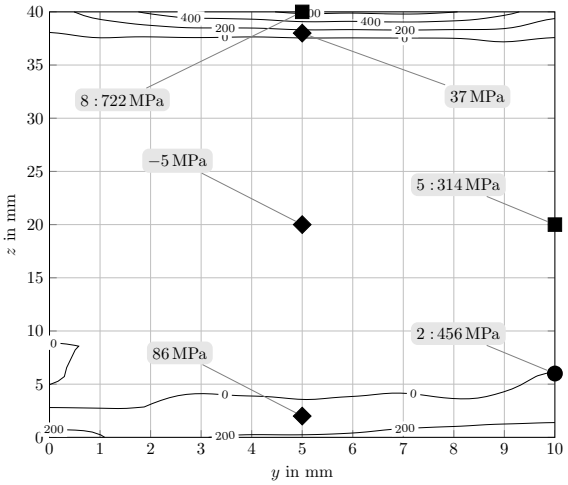
(a) Normal stresses at $z = 5\,\mathrm{mm}$



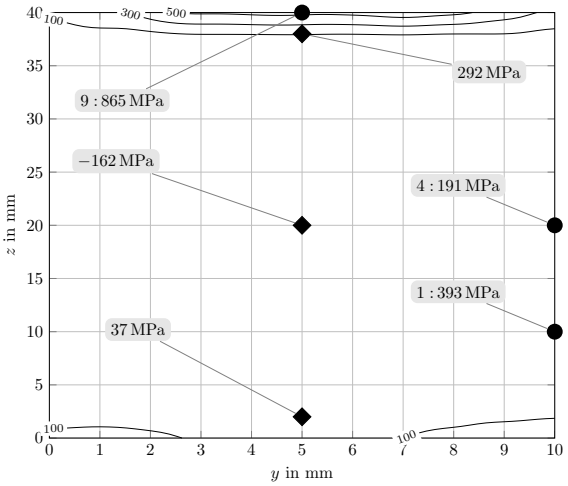
(b) Normal stresses at $z = 20\,\mathrm{mm}$



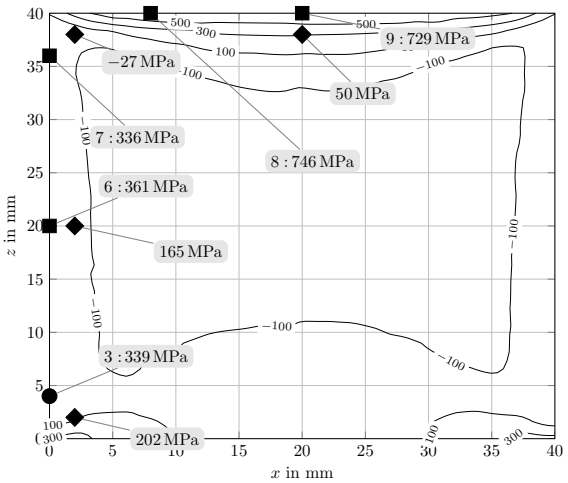
(c) Normal stresses at $z = 38\,\mathrm{mm}$



(d) Longitudinal stresses at $x = 5\,\mathrm{mm}$



(e) Longitudinal stresses at $x = 20\,\mathrm{mm}$



(f) Transversal stresses at $y = 5\,\mathrm{mm}$

Figure 4: Comparison of simulated (contour plot), neutron diffraction (diamonds), IHD (circle) as well as XRD (rectangle) measured results for the residual stresses in the presented cuboid geometry. The lines in the contour plots are equidistant.

The stress plots for the z-direction show the mentioned characteristic of high tensile stresses in the areas close to the surface and a compressive region in the middle of the part that declines with increasing z-height (cf. Figures 4a to c). The experimental results match the trend and show acceptable agreement with the simulation. However, the magnitude of tensile stresses found by experiments is not reproduced by the simulation, e. g. Figure 4c, measurement point 7.

In the longitudinal x-direction, the plots continue to show the agreement between ND and simulation. However, the high tensile stresses in the regions close to the surface are not predicted by the simulation. Additionally, the plot exhibits a certain level of asymmetry. This is attributable to the non-symmetric placement of the cuboid on the build plate.

For the transversal stresses, the agreement is less prominent. While the high stresses at the top of the part are correctly predicted, the simulation results do not match the experiment in the side surfaces of the cuboid. The stress state is characterized by tensile stresses at the bottom and top of the part and a compressive region in the middle of the upper half of the geometry.

4 Discussion and Conclusion

The prediction of residual stresses shows sufficient agreement with the measured values. However, especially in the border areas, the prediction and experiment differ from one another. This is assumed to be attributable to the missing modeling of the contour exposure.

The prediction of normal stresses shows an overestimation of the actual result. This is related to the modeling of the heat input that, in contrast to the real process, only accounts for a heat flow in z-direction as the consecutive hatches are summarized, eliminating temperature gradients in the x-y-area of the topmost layer. The further development should aim to incorporate the effects of the hatching into the simulation model and solve this issue.

However, the level of agreement between simulation and neutron diffraction results is highly encouraging and suggests that the chosen modeling approach is suitable for the simulation of the process. The conclusions are in particular:

1. The stresses occurring in an LBM-manufactured cuboid range from high tensile areas near yield close to the surface of the part to medium compressive stresses in the middle.
2. The presented simulation tool is capable of predicting the trend of residual stresses for the cuboid geometry.
3. The main point of deviation is the area close to the surface that exhibits very high tensile stresses not predicted by the simulation.

5 Outlook

Further investigations are necessary that include measurements of the stresses within the build plate in order to get a complete picture of the stress distribution. The statistical power of the work carried out is very low as only one sample part manufactured on one LBM system with one scan strategy was examined. The representativeness of the results is to be examined. It is however expected, that the results of other manufacturing systems are comparable. Additionally, the stripe pattern is one example of a heat input that is as homogeneous as possible, comparable to the also well-established chessboard strategy, and may thus serve as a valid reference point for this kind of strategy. Last, further studies should combine results of warpage and residual stresses so that the full mechanical response to the manufacturing process becomes available.

6 Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Program (FP7/2007-2013) for the Clean Sky Joint Technology Initiative under grant agreement No. 287087. However, explanations made reflect only the authors' views. Hence, the JU and the Union are not liable for any use that may be made of the information contained therein. The IGF-Project No 17.911 N / DVS-No. 13.007 was promoted by the AiF under the program for the promotion of industrial research and development (IGF) by the Federal Ministry of Economics and Technology. This work is based upon experiments performed at the STRESS-SPEC instrument operated by FRM II at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany.





References

[1] Terry T. Wohlers. *Wohlers report 2014. 3D printing and additive manufacturing state of the industry annual worldwide progress report*. Fort Collins, Col.: Wohlers Associates, 2014. 276 pp. ISBN: 9780991333202.

[2] Amanda S. Wu et al. "An Experimental Investigation into Additive Manufacturing-Induced Residual Stresses in 316L Stainless Steel". In: *Metallurgical and Materials Transactions A* 45 (13 2014), pp. 6260–6270. ISSN: 1073-5623. DOI: 10.1007/s11661-014-2549-x.

[3] William E. Frazier. "Metal Additive Manufacturing: A Review". English. In: *Journal of Materials Engineering and Performance* 23 (6 2014), pp. 1917–1928. ISSN: 1059-9495. DOI: 10.1007/s11665-014-0958-z. URL: http://dx.doi.org/10.1007/s11665-014-0958-z.

[4] J.-P. Kruth et al. "Assessing and comparing influencing factors of residual stresses in selective laser melting using a novel analysis method". In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 226 (6 2012), pp. 980–991. ISSN: 0954-4054. DOI: 10.1177/0954405412437085.

[5] F. Bayerlein et al. "Improving cost effectiveness in additive manufacturing – Increasing dimensional accuracy in laser beam melting by means of a simulation-supported process chain". In: *Conference proceedings / Die Fachkonferenz zur Numerischen Simulation*. CADFEM GmbH et al. Grafing bei München, Darmstadt: CADFEM GmbH, 2015, pp. 1–9. ISBN: 393752312X.

[6] Nils Keller and Vasily Ploshikhin. "New Method for fast predictions of residual stress and distortion of AM parts". In: *Proceedings of the 25th Solid Freeform Fabrication Symposium*. Ed. by D. L. Bourell et al. Vol. 25. The University of Texas at Austin. Austin, TX and USA, 2014, pp. 1229–1237.

[7] Richard Martukanitz et al. "Toward an integrated computational system for describing the additive manufacturing process for metallic materials". In: *Additive Manufacturing* 1-4 (2014), pp. 52–63. ISSN: 22148604. DOI: 10.1016/j.addma.2014.09.002.

[8] Loucas Papadakis et al. "A computational reduction model for appraising structural effects in selective laser melting manufacturing". In: *Virtual and Physical Prototyping* 9 (1 2014), pp. 17–25. ISSN: 1745-2759. DOI: 10.1080/17452759.2013.868005.

[9] Nachiket Patil. *A novel numerical framework for simulation of multiscale spatio-temporally non-linear systems in additive manufacturing processes*. 2014. DOI: 10.18297/etd/1099.

[10] W. King et al. "Overview of modelling and simulation of metal powder bed fusion process at Lawrence Livermore National Laboratory". In: *Materials Science and Technology* 31 (8 2015), pp. 957–968. ISSN: 0267-0836. DOI: 10.1179/1743284714Y.0000000728.

[11] C. Seidel et al. "Simulation of the Laser Beam Melting Process – Approaches for an Efficient Modelling of the Beam-material Interaction". In: *Procedia CIRP* 25 (2014), pp. 146–153. ISSN: 22128271. DOI: 10.1016/j.procir.2014.10.023.

[12] A. Thomas et al. "High temperature deformation of Inconel 718". In: *Journal of Materials Processing Technology* 177 (1-3 2006), pp. 469–472. ISSN: 09240136. DOI: 10.1016/j.jmatprotec.2006.04.072.

[13] Fritz Klocke et al. "Turbomachinery component manufacture by application of electrochemical, electro-physical and photonic processes". In: *CIRP Annals - Manufacturing Technology* 63 (2 2014), pp. 703–726. ISSN: 0007-8506. DOI: 10.1016/j.cirp.2014.05.004.

[14] Michael F. Zaeh and Gregor Branner. "Investigations on residual stresses and deformations in selective laser melting". In: *Production Engineering* 4 (1 2010), pp. 35–45. ISSN: 0944-6524. DOI: 10.1007/s11740-009-0192-y.

[15] T. A. Krol et al. "Verification of Structural Simulation Results of Metal-based Additive Manufacturing by Means of Neutron Diffraction". In: *Physics Procedia* 41 (2013), pp. 849–857. ISSN: 18753892. DOI: 10.1016/j.phpro.2013.03.158.

[16] K. N. Amato et al. "Microstructures and mechanical behavior of Inconel 718 fabricated by selective laser melting". In: *Acta Materialia* 60 (5 2012), pp. 2229–2239. ISSN: 13596454. DOI: 10.1016/j.actamat.2011.12.032.

[17] Johannes Strößner, Michael Terock, and Uwe Glatzel. "Mechanical and Microstructural Investigation of Nickel-Based Superalloy IN718 Manufactured by Selective Laser Melting (SLM)". In: *Advanced Engineering Materials* 17 (8 2015), pp. 1099–1105. ISSN: 1527-2648. DOI: 10.1002/adem.201500158.

[18] G. Branner. *Modellierung transienter Effekte in der Struktursimulation von Schichtbauverfahren*. Vol. 246. Forschungsberichte / IWB. München: Utz, Herbert, 2011. 230 S. ISBN: 3831640718.

[19] Julia N. Repper. "Einfluss mikroskopischer Eigenspannungen auf die makroskopische Eigenspannungsanalyse mittels Neutronenbeugung". Dissertation. München: Technische Universität München, 2010-01-01.

[20] R. A. Winholtz and A. D. Krawitz. "The effect of assuming the principal directions in neutron diffraction measurement of stress tensors". In: *Materials Science and Engineering: A* 205 (1–2 1996), pp. 257–258. ISSN: 09215093. DOI: 10.1016/0921-5093(95)10040-7.

[21] D. Rosenthal. "Mathematical theory of heat distribution during welding and cutting". In: *Welding Journal* 5 (20 1941), pp. 220–234.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MESHING OCEAN DOMAINS FOR COASTAL ENGINEERING APPLICATIONS

Alexandros Avdis[1], Christian T. Jacobs[2], Simon L. Mouradian[1], Jon Hill[3] and Matthew D. Piggott[1,4]

[1]Department of Earth Science and Engineering
Imperial College London, United Kingdom
e-mail: {a.avdis, simon.mouradian06, m.d.piggott}@imperial.ac.uk

[2]Faculty of Engineering and the Environment
University of Southampton, United Kingdom
email: C.T.Jacobs@soton.ac.uk

[3]Environment Department
University of York, United Kingdom
email: jon.hill@york.ac.uk

[4]Grantham Institute
Imperial College London, United Kingdom

Keywords: Mesh Generation, Geographical Information Systems (GIS), Computer Aided Design (CAD), Tidal Renewable Power, Severn Estuary, Tidal Lagoon.

Abstract. *As we continue to exploit and alter the coastal environment, the quantification of the potential impacts from planned coastal engineering projects, as well as the minimisation of any detrimental effects through design optimisation, are receiving increasing attention. Geophysical fluid dynamics simulations can provide valuable insight towards the mitigation and prevention of negative outcomes, and as such are routinely used for planning, operational and regulatory reasons. The ability to readily create high-quality computational meshes is critical to such modelling studies as it impacts on the accuracy, efficiency and reproducibility of the numerical results. To that end, most (coastal) ocean modelling packages offer tailored mesh generation utilities. Geographical Information Systems (GIS) offer an ideal framework within which to process data for use in the meshing of coastal regions. GIS have been designed specifically for the processing and analysis of geophysical data and are a popular tool in both the academic and industrial sectors. On the other hand Computer Aided Design (CAD) is the most appropriate tool for designing coastal structures and is usually the user interface to generic three–dimensional mesh generation frameworks. In this paper we combine GIS and CAD with a view towards mesh generation for an impact study of the proposed Swansea Bay Tidal Lagoon project within the Bristol Channel and Severn Estuary. We demonstrate in this work that GIS and CAD can be used in a complementary way to deliver unstructured mesh generation capabilities for coastal engineering applications.*

1 INTRODUCTION

Geophysical domains are geometrically complex, with shorelines the most commonly known example of fractal geometry in nature [1]. Many transient features in oceanic flow emerge due to the multi–scale nature of the domain geometry. Therefore, the predictive accuracy of geophysical fluid dynamics simulations will depend on an appropriately accurate and efficient representation of a wide range of features present in the geometry. In addition to natural features, in the area of coastal engineering a further critical small–scale geometrical feature is that of the infrastructure: pipes, piers, tidal dams and wave defence structures being typical examples. The minimisation of environmental impact as well as the resilience of infrastructure are of such importance that policy frameworks increasingly regulate these aspects of coastal infrastructure projects. This has led to the adoption of hydrodynamic coastal modelling as a necessary tool in the infrastructure design process. Resolving coastal structures in a computational mesh, or at least accurately parameterising their effect on the flow, is necessary to achieve high predictive accuracy in simulations. However, the impact of many coastal structures can be seen at much larger scales, especially when linked to tidal power generation [2, 3, 4]. Therefore, the simulation domain must extend over much greater scales. In fact, the desire for high resolution in parts of a large multi–scale domain, is common in coastal as well as ocean modelling [5, 6, 7].

Unstructured meshes are well suited to the efficient tessellation of complex, multi–scale coastal domains, by allowing the element size to vary by orders of magnitude. Unstructured mesh generation methods have matured and can produce high–quality meshes for domains typically encountered in coastal ocean modelling. However, most generic mesh generation tools have been developed within the wider realm of *Computational Fluid Dynamics (CFD)* and are not geared towards ocean and coastal modelling, as typically a *Computer Aided Design (CAD)* interface is available to the user. CAD is ubiquitous in coastal engineering and an essential component in industrial fluid dynamics applications. The native ability of CAD to accurately specify coastal structures or other infrastructure, often with reference to set standards, is necessary in the design as well as construction processes. However, CAD is not the best framework for the description of geophysical data, where *Geographical Information Systems (GIS)* are far better suited. The flexibility, extensibility and robustness of GIS, and in particular the ability to process and analyse complex and multi–scale geophysical information, has led to their integration with data visualisation, mesh generation [8, 9] and data management [10] frameworks.

In this paper we focus on the generation of multi–scale unstructured meshes in the context of coastal engineering applications and present a framework to achieve this, combining GIS and CAD. In the following sections we outline the developed framework, and discuss how it relates to mesh generation in coastal engineering. A preliminary application to the Bristol Channel and Severn Estuary follows, where numerous projects on renewable tidal power generation have been proposed. Most such projects require significant coastal infrastructure design and development. We close with a discussion of our key findings and outlook on future work.

2 MESHING FOR COASTAL ENGINEERING APPLICATIONS

The domain in simulations of coastal and ocean flows is usually best described in a topologically two–dimensional space, where the domain is bounded by shorelines and open boundaries. Mesh generation in two dimensions is the natural and obvious choice in cases where a two–dimensional framework, such as the depth–averaged shallow water equations, can be used to express the phenomena under study. In cases where a three–dimensional framework must be used (e.g. when non–hydrostatic effects are important) a topologically two–dimensional

description of the domain will typically suffice. Most three–dimensional models will only require a two–dimensional mesh, over a reference surface, and construct a vertically structured three–dimensional mesh by projecting the surface mesh vertices towards the ocean floor. GIS is therefore ideal for describing domains in realistic geophysical simulations [11], since one of the primary purposes of GIS is to store, manipulate and analyse geographical, topologically two–dimensional information. In order to meet the demand for robust mesh generation of geophysical domains several utilities have been developed, adopting GIS methodologies. For example, *BlueKenue*, *MIKE Zero* and *RGF-Grid* are mesh generation and post–processing utilities used by the *Telemac*, *MIKE* and *Delft3D (Flexible Mesh)* models respectively. In this paper we use the utilities described by Avdis et al. [9] which interface the QGIS suite [12] and the Gmsh mesh generator [13], by linking similar abstractions used in both GIS and mesh generators: *vector* data structures are used to describe domain boundaries and *raster* data structures are used to express fields such as the spatial distribution of the desired element edge size.

Unlike GIS, CAD is predominantly aimed at designing three–dimensional entities, and an explicit reference to a geographical coordinate reference system need not be given. Particular emphasis is put on construction or manufacture processes, where determination of materials and construction methods are often aided by the more general *Computer Aided Engineering (CAE)* framework which includes interfaces to industrial fluid dynamics solvers and mesh generation suites. However, the use of both CAD and GIS in planning and construction of infrastructure, as well as the recent trends of increasing accuracy demands from geo–location services has led both GIS and CAD to adopt features and capabilities from one another. Mesh generation for coastal engineering is an example of this complementary relationship, where both CAD and GIS must be used; detailed descriptions of coastal structures and their construction can only be made in a CAD framework, while only a two–dimensional description of the surrounding environment is necessary. In this study we used CAD to extract broad features of the coastal infrastructure and define the relevant shoreline boundaries, identifying an acceptable approximation to the wetting line. In general, the definition of shoreline boundaries can only be made on a case by case basis, accounting for tidal or other inundation variations. These boundaries were then imported into GIS and were used in the domain definition and mesh generation.

3 TIDAL LAGOONS IN THE BRISTOL CHANNEL

The large tidal range in the Bristol Channel and Severn Estuary has inspired many proposals for renewable energy extraction [17]. Several plans have proposed building dams spanning the estuary (for example [18])–proposals collectively known as the Severn Barrage. Sections of the structure house sluice gates and hydro–turbines, in order to control the flow of the water and in particular to form and harness a potential energy difference across the dam wall. However, construction cost and concerns over potential environmental impacts [19] have been major failing points. Tidal lagoons are similar in construction and operation, but at a smaller scale. A tidal lagoon proposal currently being considered by the UK Government is the Swansea Bay Tidal Lagoon [20, 21]. Figures 1(d,e) and 2 show its location and outline the proposed structure. As shown, a sea wall with appropriate turbine and sluice gate housings encloses a large area within the estuary. Since the structure of a tidal lagoon is local and does not cross the estuary, tidal lagoons are thought to result in smaller environmental impacts [22]. The appraisal of tidal lagoons must include environmental impact studies and the design must limit these. Although the level of commercial fishing activity within the Severn Estuary and Inner Bristol Channel is low when compared to other British inshore fisheries [23], the channel is an important nursing ground for commercially sensitive fishery species such as sole and bass. These juvenile fish populations

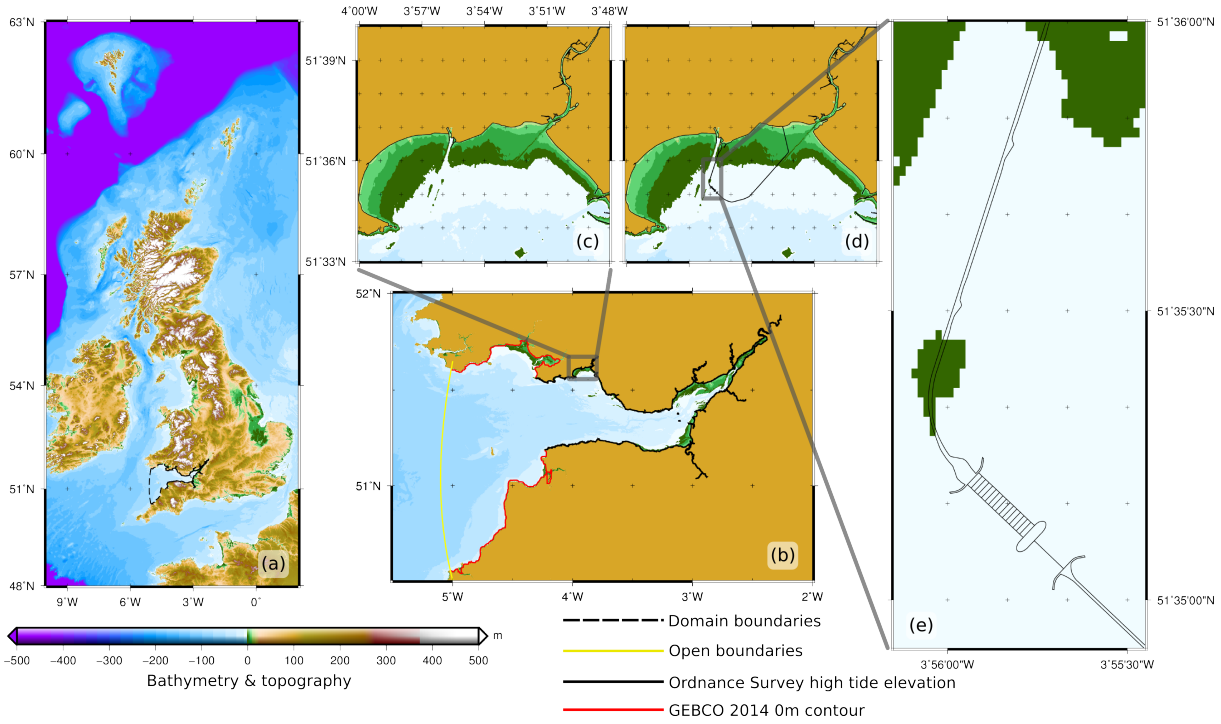


Figure 1: The domain boundaries and bathymetry for simulations of the Severn Estuary and Bristol Channel, with and without the Swansea Bay Tidal Lagoon structure: (a) Map of GEBCO 2014 bathymetry [14] over the UK and Ireland, with the simulation domain boundaries; (b) The simulation domain boundaries, without the tidal lagoon, identifying the source of the boundaries; (c) Detail of the Swansea Bay area; (d) Detail of the Swansea Bay area with the tidal lagoon boundaries, derived from [15]; (e) Detail of the tidal lagoon boundaries showing the housing of the tidal turbines and sluice gates. The Digimap marine bathymetry [16] is shown in panels (b,c,d,e).

must be preserved. The estuary is also notably rich in fish species, being home to more than 100 variations [24]. In addition to the environmental concerns, building tidal lagoons requires significant capital, leading to investors (including the state) requiring estimates on return with sufficient accuracy and ample supporting evidence [25]. Accurate optimisation studies of tidal lagoon structures, in order to minimise environmental impacts and maximise return, are therefore important. Numerical simulations of the flow over the wider region resolving the structure and its effects can be a valuable tool in appraisal and design (as carried out, for example, in [26]). Here we present the construction of a mesh as well as simulation results over the Severn Estuary and Bristol Channel, resolving the tidal lagoon structure.

3.1 Mesh Generation

Panels (a,b,c) in figure 1 show the domain boundaries superimposed on bathymetry maps of the simulation domain without the tidal lagoon structure. Panel (a) shows an overview of the whole domain in relation to the United Kingdom, while panel (b) identifies the sources of the domain boundaries with differing fidelities. The domain boundaries have been selected to be suitable for simulations aimed at identifying the effects of the Swansea Bay tidal lagoon on the Severn Estuary, so high mesh and shoreline geometry resolution was used in the inner part of the Bristol Channel and the Severn Estuary. The Western part of the Bristol Channel is included in the simulation, in order to place the open boundary away from the region of interest. As shown in figure 1(b), the UK Ordnance Survey shoreline data [27], and in particular the high-tide

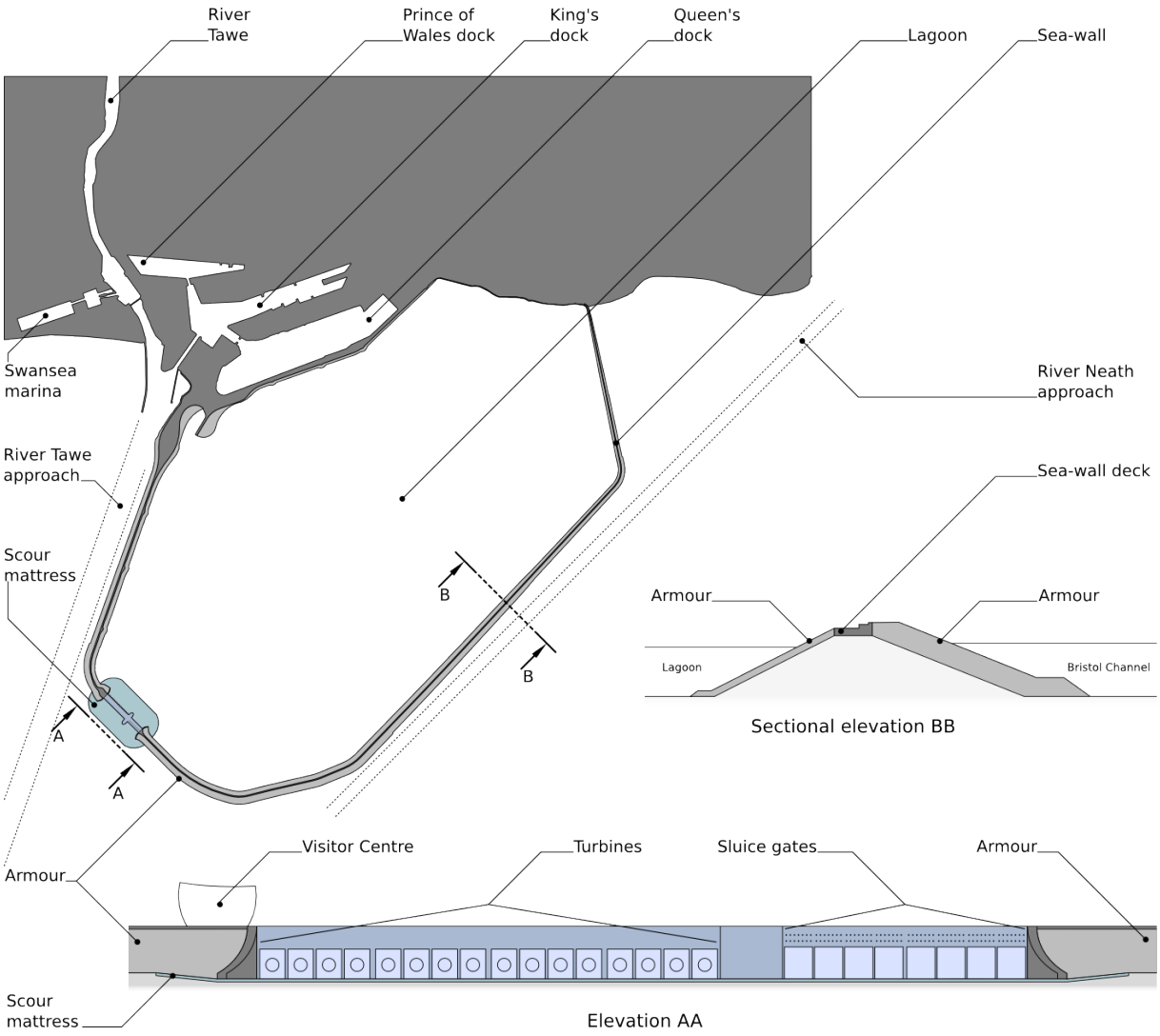


Figure 2: Schematic diagrams of the Swansea Bay tidal lagoon showing lagoon location, and broadly outlining lagoon construction, based on technical drawings of the proposed lagoon [15]. The plan view shows the location of the lagoon relative to the Swansea Docks, Swansea Marina, River Tawe and the approximate locations of the approach channels to the rivers Tawe and Neath. Elevation AA shows the housing of the 16 turbines and 8 sluice gates, viewed from the Bristol Channel. Sectional elevation BB outlines the construction of the sea–wall, identifying the armour either side and sea–wall deck.

boundary was used in the inner part of the Bristol Channel and Severn Estuary. The zero–elevation contour extracted from the GEBCO 2014 30-minute arc gridded bathymetry [14], shown in figure 1(a), was used in the portion close to the open boundary. The open boundary was formed by linearly blending two loxodromes (lines of constant bearing): The loxodrome starting from the Welsh coast at 5° East, 51.65° North, at a bearing 170° West, up to a latitude 51.0° North and the loxodrome starting from the Cornish Peninsula at 5° East, 50.5° North, at a bearing 30° West, up to a latitude 51.0° North.

The proposed Swansea Bay tidal lagoon would extend dock infrastructure into the Swansea Bay, as outlined in figure 2. As suggested by the bathymetry in figure 1(c,d) and figure 2 a $9.5km$ sea wall will follow the direction set out by existing dredged approach channels leading to rivers Neath and Tawe, and impound an area South of the Swansea Docks. As shown

in figure 2, the structure is planned to include sixteen turbines and eight sluice gates at a purpose built housing unit, located at the South–West side of the lagoon. The boundaries for the lagoon structure used here were obtained from technical drawings in [15]. Thus, as discussed in sections 1 and 3 the specification of the tidal lagoon structures was only available in a form originating from a CAD/CAE framework. Furthermore, only drawings of the structure were available, which were digitised as a first step, extracting the boundaries of the sea–wall deck, identified in the sectional elevation of figure 2. The digitisation was carried out within QGIS using geo–referencing utilities, so as to accurately locate the lagoon structures in relation to the Ordnance Survey boundaries [27]. The armour either side of the wall, identified in figure 2, is ignored in this case and the domain extends over the armour, up to the seawall crest. The qCAD open–source package was used to accurately edit the digitised boundaries and define the domain boundaries over the turbines and sluice gates.

A two–dimensional, shallow–water approximation, discussed in section 3.2, was chosen for simulating the flow in the domain. Turbines and sluice gates cannot be represented by the shallow–water framework, due to their complex three-dimensional geometry. Instead, only their effect on the larger scales can be captured. The energy extraction from the turbines is usually parameterised as a sub–grid–scale process, applied only to mesh elements in the vicinity of the device [28]. Therefore, the turbines are here represented by adjacent rectangles (see figure 1(e)), that will be covered by mesh elements, while the mesh is conforming to polygon boundaries (see figure 4(d)). Appropriate parameterisations can be applied in elements inside the turbine–representing regions. The definition of adjacent, meshed, polygons creates a number of *internal boundaries*. The flow through the turbines is also controlled by turbine gates. The effect of the turbine gates can be captured by appropriate conditions on the flux through the edges lying on the internal boundaries between the lagoon and turbine regions. The sluice gates are represented as internal boundaries between the lake and the Bristol Channel.

A small element edge length is required for an accurate representation of the detail present in the shorelines obtained from the Ordnance Survey, whereas a lower resolution is desired close to boundaries described by the $0m$ contour from the GEBCO 2014 bathymetry. The distribution of element size is shown in figure 3 and is based on the proximity to the shorelines. The mesh at the boundaries obtained from the Ordnance Survey has an element edge length of $10m$. This resolution is maintained up to a distance of $150m$ from the boundary. The edge length then increases to an edge length of $10km$ over a distance of $25km$. An identical element size distribution was chosen towards the boundaries of the tidal lagoon structure. A $1km$ element edge size is prescribed at the boundaries expressed as a $0m$ contour of the GEBCO 2014 bathymetry [14]. This resolution is maintained up to a distance of $2km$ from the boundary and then increases to $10km$ over a distance of $15km$. Panel (a) in figure 4 shows the mesh *without* the tidal lagoon structure and panels (b,c,d) show the mesh *with* the tidal lagoon structure. The mesh in the lagoon and the turbine polygons are coloured differently highlighting the individual regions and internal boundaries.

3.2 Simulation Results

In order to show the utility of the meshes shown in figure 4, the Telemac package was used to simulate the tidal flow using the mesh shown in figure 4(a). The "depth–velocity" formulation

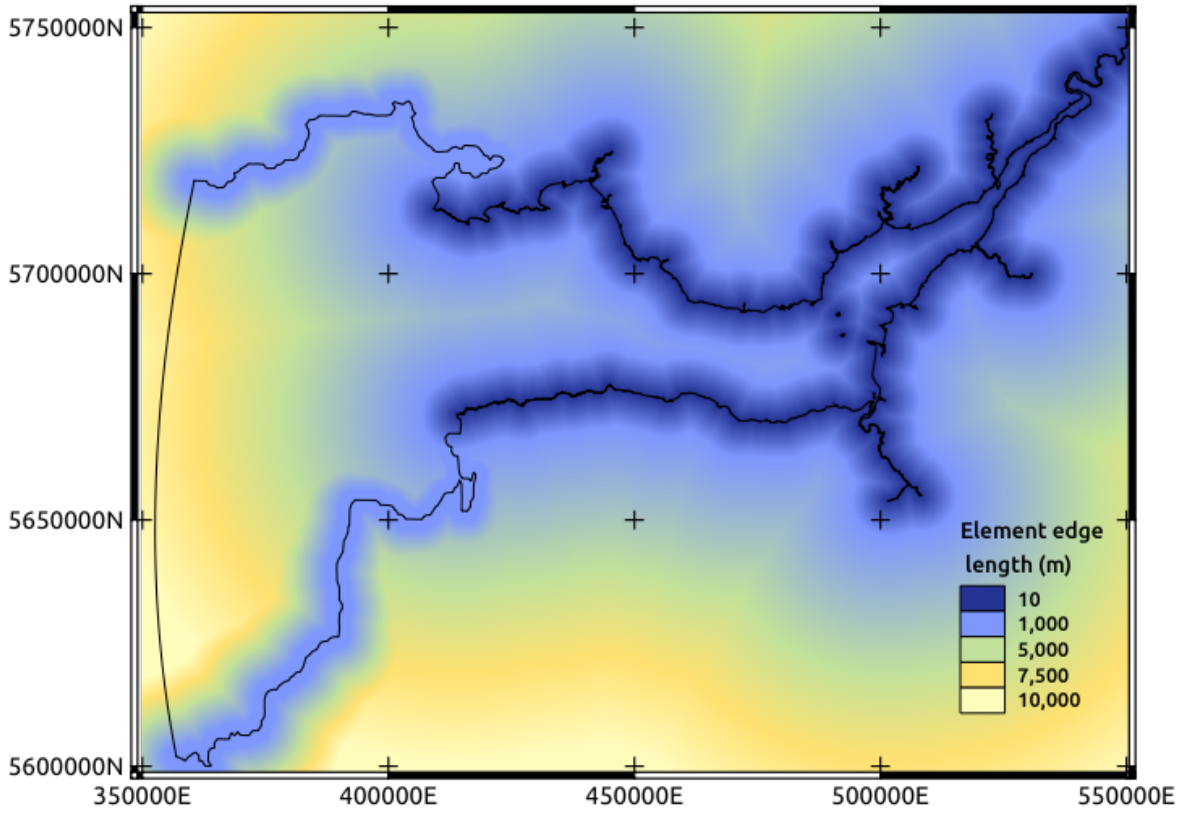


Figure 3: Map of user–specified mesh size metric (element edge length) in the UTM30 coordinate reference system, for the simulation domain without the tidal lagoon.

of the shallow water (Saint-Venant) equations [29] was used:

$$\frac{\partial h}{\partial t} + u_i \frac{\partial h}{\partial x_i} + h \frac{\partial u_i}{\partial x_i} = 0 \,, \tag{1}$$

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = - g \frac{\partial Z_s}{\partial x_i} + \frac{1}{h} \frac{\partial}{\partial x_j} \left(h \nu_e \frac{\partial u_i}{\partial x_j} \right) + C_i + S_f u_i \,, \tag{2}$$

where repeated indices imply summation, ν_e is the effective viscosity and h is the total water depth measured from the free-surface to the ocean floor. Z_s is the free surface elevation measured from the UTM30N vertical datum and Z_f denotes bathymetry, such that $h = Z_s - Z_f$. C_i are the Coriolis force terms, defined as:

$$\begin{aligned} C_1 &= f u_2 \,, \\ C_2 &= -f v_1 \,, \end{aligned} \tag{3}$$

where $f = 1.1 \times 10^{-4} s^{-1}$, valid for a latitude of approximately $50°$. S_f in equation 2 is the floor friction term, expressed by the Chézy law:

$$S_f = - \frac{g}{\cos(a) h C^2} \sqrt{u_1^2 + u_2^2} \,, \tag{4}$$

where $C = 62.6 m^{\frac{1}{2}} s$.

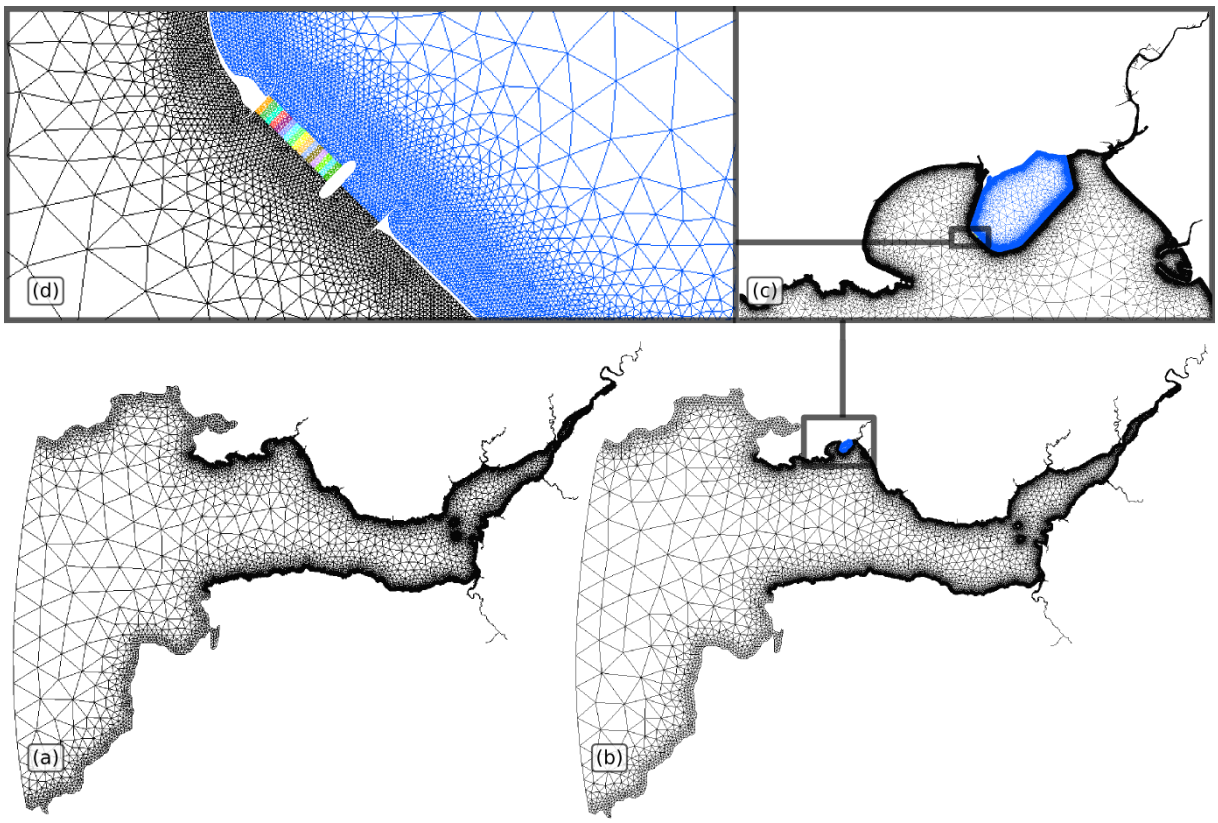


Figure 4: Multi–scale unstructured meshes over the Bristol Channel and Severn Estuary simulation domain, in UTM30: (a) Mesh of the domain without the tidal lagoon, containing $1,166,851$ vertices, $2,246,905$ triangles; (b) mesh of the domain with the tidal lagoon, containing $1,205,631$ vertices, $2,323,049$ triangles; (c) mesh detail over the Swansea Bay and tidal lagoon; (d) detail over the Swansea Bay Tidal Lagoon, showing the mesh around the sea–wall and the mesh over the turbines and sluice gates.

Equations 1 and 2 are discretised using a mixed finite element pair, composed of semi–quadratic (*quasi–bubble* in [29]) elements for velocity and linear elements for depth. Time stepping is affected using a fractional step method, where the advective terms are used to calculate intermediate values during a first stage using the method of characteristics detailed in [29]:

$$\begin{aligned}\frac{h^* - h^n}{\Delta t} + u_i \frac{\partial h}{\partial x_i} &= 0\,,\\ \frac{u_i^* - u_i^n}{\Delta t} + u_j \frac{\partial u_i}{\partial x_j} &= 0\,,\end{aligned}$$

where superscripts denote time–levels. The primitive variables at the end of the time–step are calculated in a second stage, using a semi–implicit discretisation of the terms with a θ parameter controlling the implicitness of each term:

$$\begin{aligned}&\frac{h^{n+1} - h^*}{\Delta t} + \left[h \frac{\partial u_i}{\partial x_i}\right]^{n+\theta_{hp}} = 0\,,\\ &\frac{u_i^{n+1} - u_i^*}{\Delta t} + \left[h \frac{\partial u_i}{\partial x_i}\right]^{n+\theta_{vp}} = \left[-g \frac{\partial Z_s}{\partial x_i}\right]^{n} + \left[\frac{1}{h}\frac{\partial}{\partial x_j}\left(h \nu_e \frac{\partial u_i}{\partial x_j}\right)\right]^{n+\theta_{vd}} + C_i^n + S_f^n u_i^{n+1}\,.\end{aligned}$$

In the simulations presented here $\theta_{hp} = \theta_{vp} = 0.55$ and $\theta_{vd} = 1$.

The OTPS tidal dataset over the North–Western European shelf [30] is used to prescribe the depth at the open boundary. The 1 arc–second *Digimap Marine, HydroSpatial One* [16] gridded bathymetry data, shown in panels (b,c,d,e) of figure 1, was used to define the bathymetry in the simulation domain. The Severn Estuary is known to have large *tidal flats*, which are submerged during high tides. The cyclical exposure of tidal flats requires special *wetting-and–drying* algorithms in the solution procedure, and such algorithms are available in the Telemac modelling framework. However, given the focus of this study is mesh generation in the context of coastal engineering, wetting–and–drying algorithms have not been incorporated into the solution procedure. Instead, the bathymetry was modified to be $-10m$ in any areas that exceeded that value. While this may affect the predictive accuracy of the simulation results presented here, the mesh generation procedure has not been altered substantially in ongoing studies incorporating wetting–and–drying.

Five days are simulated, starting at $00 : 00$ 2 July 2011, using a time step of $0.2s$. Figure 5 shows the instantaneous velocity magnitude contours at the end of the simulation. Details over the Severn Estuary and the Swansea Bay are shown. The velocity vectors are also shown in the Swansea Bay area. At the time of the shown instant the flow is in ebb tide and strong velocities are seen in the Bristol Channel, close to the shorelines. In the Severn Estuary high velocities are indicated in the shallow areas. However, the bathymetry modification and lack of wetting–and–drying compromise the predictive accuracy in shallow regions. This also applies to the Swansea Bay results, where the velocity vectors indicate an along–shore flow, with a recirculation region in the lee of the piers around the river Tawe. Nonetheless, the preliminary results presented here show the suitability of the mesh for coastal engineering fluid dynamics simulations.

4 CONCLUSIONS

The use of geophysical fluid dynamics simulations in the context of environmental impact assessment [2, 3, 4, 5, 17, 26] and coastal infrastructure optimisation [31] is becoming common place. Robust mesh generation is therefore required, enabling an adequate representation of coastal infrastructure, within inherently complex geophysical domains. GIS and CAD have both been used as interfaces to mesh generation, but in the context of coastal engineering the complementary capabilities of both frameworks are useful: while CAD can be used to accurately specify the geometry of coastal structures, the impact of coastal structures can be seen on a much larger scale where GIS is the appropriate framework for geometry description and data analysis. We have here shown how the combination of GIS and CAD can form an efficient framework for mesh generation.

The Swansea Bay Tidal Lagoon is a typical example of coastal infrastructure where environmental impact, construction cost and financial return must be quantified and balanced. We have shown how the proposed mesh generation framework can be used to produce meshes with very accurate shoreline representations, including the proposed lagoon structure. On–going work is aimed at understanding the effect of the tidal lagoon on the Severn Estuary, Bristol Channel and Irish Sea, as past studies [3, 2] have shown that a tidal barrage will affect the tidal range in a much wider area. To that end further simulations are needed, incorporating wetting–and–drying, a sediment model [32], realistic operation and parameterisation of sluice gates and turbines, with high mesh resolution over large parts of the Severn Estuary and Bristol Channel.

Future work will be aimed towards mesh generation for simulations requiring detailed description of structures within very large domains. For example, the parameterisation of geo-

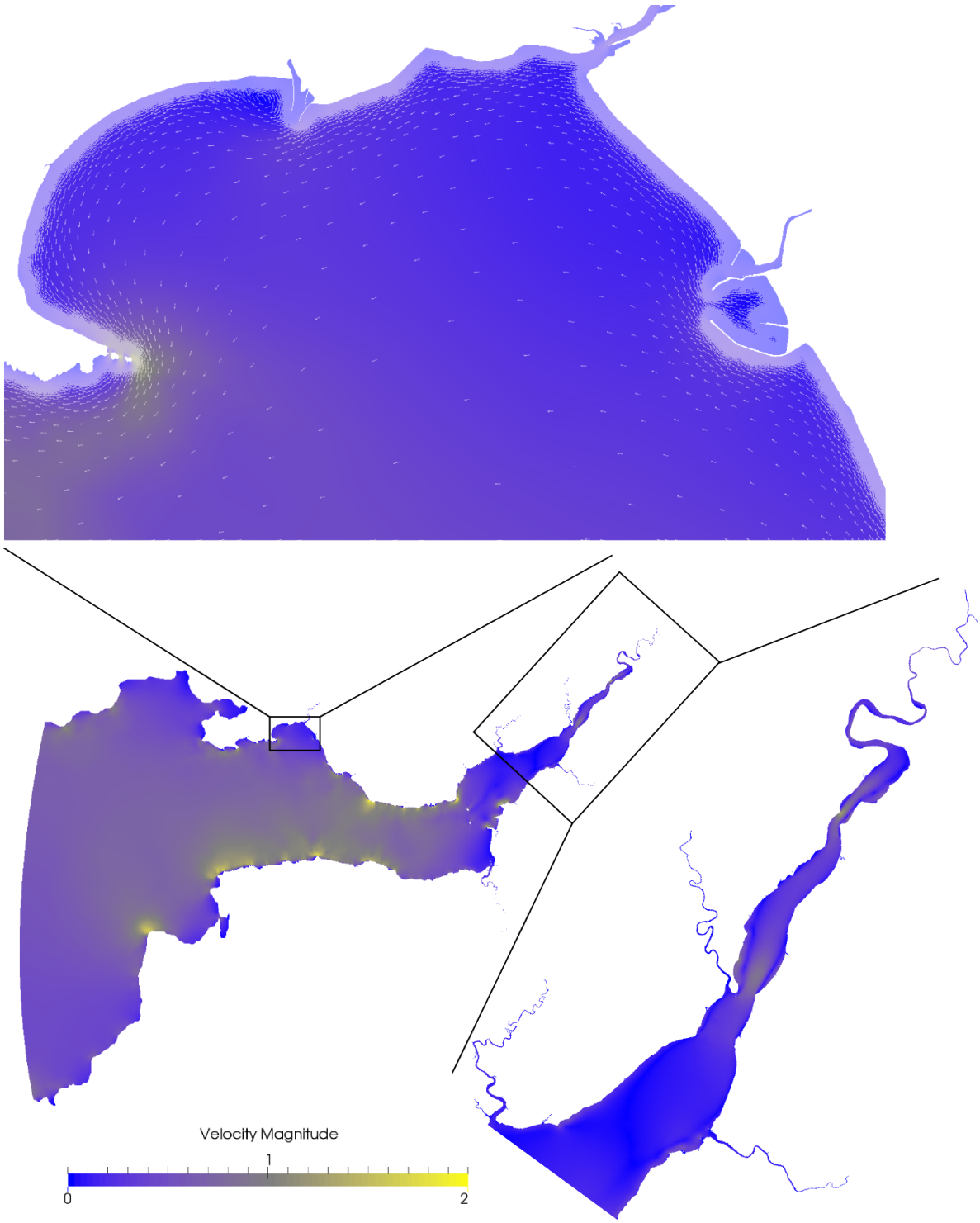


Figure 5: Simulation results, after five days of simulated time, at 00 : 00, 7 July 2011: Bottom left panel shows instantaneous velocity magnitude contours in the simulation domain; bottom right panel shows the Severn Estuary; top panel shows the Swansea Bay, also with velocity vectors.

metrically complex turbines has been common practice in studying the impact of tidal power generation arrays on sedimentation patterns [5]. However, recent studies have an increasingly detailed geometric description of the tidal turbines [33], as the focus shifts on tidal device per-

formance, loading, survivability, interaction between devices, and tidal array optimisation in realistic domains. In addition, the numerical framework is envisaged to combine a shallow water representation over the far region, with an accurate, three–dimensional description of the flow around the device.

REFERENCES

[1] Benoit B Mandelbrot and John A Wheeler. The fractal geometry of nature. *American Journal of Physics*, 51(3):286–287, 1 March 1983.

[2] Samuel Bray, Reza Ahmadian, and Roger A Falconer. Impact of representation of hydraulic structures in modelling a severn barrage. *Computers and Geosciences*, 2016.

[3] Juntao Zhou, Shunqi Pan, and Roger A Falconer. Effects of open boundary location on the far-field hydrodynamics of a severn barrage. *Ocean Modelling*, 73:19–29, 2014.

[4] Donald C Gordon, Jr. and Alan R Longhurst. The environmental aspects of a tidal power project in the upper reaches of the bay of fundy. *Marine Pollution Bulletin*, 10(2):38–45, February 1979.

[5] R Martin-Short, J Hill, S C Kramer, A Avdis, P A Allison, and M D Piggott. Tidal resource extraction in the pentland firth, UK: Potential impacts on flow regime and sediment transport in the inner sound of stroma. *Renewable Energy*, 76:596–607, April 2015.

[6] Jon Hill, Gareth S Collins, Alexandros Avdis, Stephan C Kramer, and Matthew D Piggott. How does multiscale modelling and inclusion of realistic palaeobathymetry affect numerical simulation of the Storegga Slide tsunami? *Ocean Modelling*, 83:11–25, November 2014.

[7] Lin Luo, Jia Wang, David J Schwab, Henry Vanderploeg, George Leshkevich, Xuezhi Bai, Haoguo Hu, and Dongxiao Wang. Simulating the 1998 spring bloom in lake michigan using a coupled physical–biological model. *Journal of Geophysical Research*, 117(C10):C10011, 1 October 2012.

[8] Alexandros Avdis, Christian T Jacobs, Jon Hill, Matthew D Piggott, and Gerard J Gorman. Shoreline and bathymetry approximation in mesh generation for tidal renewable simulations. In *Proceeding of 11th EWTEC*, 2015.

[9] Alexandros Avdis et al. Efficient unstructured mesh generation for renewable tidal energy. *In preparation*.

[10] Christian T Jacobs, Alexandros Avdis, Simon L Mouradian, and Matthew D Piggott. Integrating research data management into geographical information systems. In *Proceedings of the 5th International Workshop on Semantic Digital Archives*, pages 7–17, 18 September 2015.

[11] P Berry, S Bonduá, V Bortolotti, C Cormio, and E M Vasini. A GIS-based open source preprocessor for georesources numerical modeling. *Environmental Modelling & Software*, 62:52–64, 2014.

[12] QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation, 2016.

[13] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3–D finite element mesh generator with built-in pre- and post-processing facilities. *Int. J. Numer. Methods Eng.*, 79(11):1309–1331, 2009.

[14] British Oceanographic Data Centre, NERC. The GEBCO 2014 grid. Technical report, 2014.

[15] Tidal Lagoon (Swansea Bay) plc. Application for Development Consent Order, Volume 4 (Plans), Folder 1 (Planning Drawings). Technical report, February 2014.

[16] Digimap Marine© Crown Copyright/SeaZone Solutions. All Rights Reserved. Licence No 052006.001 31st July 2011. Not to be Used for Navigation.

[17] Junqiang Xia, Roger A Falconer, and Binliang Lin. Impact of different tidal renewable energy projects on the hydrodynamic processes in the Severn Estuary, UK. *Ocean Modelling*, 32(1–2):86–104, 2010.

[18] S J Taylor. The Severn Barrage - Definition study for a new appraisal of the project. Technical report, The Severn Tidal Power Group, 2002.

[19] The Energy and Climate Change Committee. A Severn Barrage? Technical report, House of Commons, 21 May 2013.

[20] *Autumn Statement 2014*. HM Treasury, 2014.

[21] *Budget 2015*. HM Treasury, 18 March 2015.

[22] Friends of the Earth Cymru. A Severn barrage or tidal lagoons? A comparison. Technical report, January 2004.

[23] Food Certification International Ltd, Prepared for Defra. Severn Estuary; Inshore Fisheries and Conservation Status. Technical report, May 2011.

[24] P A Henderson and D J Bird. Fish and macro-crustacean communities and their dynamics in the Severn Estuary. *Marine Pollution Bulletin*, 61(1-3):100–114, 13 January 2010.

[25] Department of Energy & Climate Change. Review of Tidal Lagoons, 10 February 2016.

[26] Junqiang Xia, Roger A Falconer, and Binliang Lin. Hydrodynamic impact of a tidal barrage in the severn estuary, UK. *Renewable Energy*, 35(7):1455–1468, 2010.

[27] Ordnance Survey data© Crown copyright and database right, 2012.

[28] Stephan C Kramer and Matthew D Piggott. A correction to the enhanced bottom drag parameterisation of tidal turbines. *Renewable Energy*, 92:385–396, July 2016.

[29] Jean Michel Hervouet. *Hydrodynamics of Free Surface Flows: Modelling with the finite element method*. Wiley, 11 May 2007.

[30] Gary D Egbert, Svetlana Y Erofeeva, and Richard D Ray. Assimilation of altimetry data for nonlinear shallow-water tides: Quarter-diurnal tides of the northwest european shelf. *Continental Shelf Research*, 30(6):668–679, 1 April 2010.

[31] Simon W Funke, Stephan C Kramer, and Matthew D Piggott. Design optimisation and resource assessment for tidal-stream renewable energy farms using a new continuous turbine approach. *Submitted to Renewable Energy*, 21 July 2015.

[32] W S Merritt, R A Letcher, and A J Jakeman. A review of erosion and sediment transport models. *Environmental Modelling and Software*, 18(8–9):761–799, October 2003.

[33] Karina Soto and Cristian Escauriaza. Numerical simulations of marine hydrokinetic devices in complex bathymetries. In *Proceeding of 11th EWTEC*, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

HIGH-ORDER TRACKING METHOD FOR THE SIMULATION OF BURNING FRONTS IN 3D

Denis Gueyffier and Bastien Andrieu

ONERA – The French Aerospace Lab
Palaiseau, France
e-mail: {denis.gueyffier, bastien.andrieu}@onera.fr

Keywords: front-tracking method, Eikonal flow, spectral method.

Abstract. *We present a novel set of techniques designed for very accurately tracking a propagating front of arbitrarily complex 3*D *geometry. The method relies on a segmentation into smooth patches, allowing for the use of spectral methods while avoiding Gibbs phenomenon. High-order accuracy is thus obtained using few degrees of freedom. We show how this method robustly and accurately tackles difficult geometrical problems arising when a surface evolves under an Eikonal flow, e.g. formation of singularities and large variation in time of the surface area. In particular, interactions between patches are handled by using a topological representation of the front and taking appropriate actions in order to preserve a physically valid front.*

1 INTRODUCTION

Simulation of fluid flow in presence of moving interfaces or boundaries is a difficult task, yet of critical importance in domains such as solid rocket propulsion. A solid rocket motor is basically composed of a casing in which a solid propellant – the grain – is inserted. Once ignited, the solid propellant converts to high-temperature, high-pressure gases that are expelled through a nozzle, producing thrust. As the grain burns radially outward, the volume of the chamber increases in time. On the other hand, the propellant's burning velocity is linked to local properties of the internal flow inside the combustion chamber. Accurately predicting the time evolution of the burning grain surface is thus of utmost importance, as the propellant's geometry is the only means to control rocket thrust.

This paper presents recent advances in the development of a high-order method for tracking this burning surface.

2 BURNING FRONT PROBLEM

When burning, the propellant regresses following the local normal direction to the grain surface. The burning velocity is given by empirical laws based upon the propellant composition and local variables of the internal flow such as static pressure or velocity.

Several approaches exist when dealing with the representation of a dynamic interface, and can be divided into two main categories. The so-called front-*capturing* methods rely on an implicit description of the front (e.g. level-set methods [4]). While they are well suited to address interfaces of complex topology, very fine grids are generally needed to obtain decent accuracy. On the other hand, front-*tracking* methods are based on an explicit representation and thus need to deform a mesh. These explicit methods offer a potentially more accurate description of the front yet are in general more complex to implement.

The method presented in this paper falls into the second category. The time-dependent grain surface is assumed to be represented parametrically as follows

$$\begin{array}{c} \gamma : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^3 \\ (u, v, t) \mapsto \mathbf{x} = (x, y, z) \end{array} \tag{1}$$

The position vector $\mathbf{x}$ is then solution to the following partial differential equation (PDE)

$$\frac{\partial \mathbf{x}}{\partial t} = V(\mathbf{x})\mathbf{n}(\mathbf{x}) \,, \tag{2}$$

V being the grain burning velocity and $\mathbf{n}$ the unit outward normal vector, obtained using the cross-product of two tangent vectors.

$$\mathbf{n} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} \,, \tag{3}$$

where subscript u (resp. v) denotes differentiation with respect to the u parameter (resp. v).

The front can be represented implicitly by a level-set function T such that surfaces $\mathbf{x} = \gamma(t)$ are solution of equation

$$T(\mathbf{x}) = t \,. \tag{4}$$

Following [4], equation 4 can be restated as a form of the well-known Eikonal equation, which in our case reads

$$\|\nabla T\|^2 = \frac{1}{V^2} \,. \tag{5}$$

The Eikonal equation is relevant to many front-propagation problems. This equation is a hyperbolic PDE whose solutions often exhibit singularities, also called *caustics*. Dealing with Eikonal flows is particularly difficult as the propagating front undergoes dramatic distortions.

In [1], a novel method has been presented for the simulation of the internal flow inside the chamber of a solid rocket booster, coupled with the computation of the burning grain regression in a 2D configuration. This approach makes use of differentiation in Fourier space to track the burning front very accurately, while handling the singularities (caustics) which develop naturally in time. A dynamic surface mesh of the front is then constructed. A volume mesh of the chamber is deformed to follow the propagation of the front. The method is also designed to preserve the connectivity of both the surface and volume meshes, as most CFD codes do not handle either addition or deletion of mesh elements during the calculation. It is finally demonstrated that the method yields excellent accuracy even with very few degrees of freedom.

In the following sections, we present an extended version of this novel approach, designed to address more complex, fully three-dimensional geometries. This paper focuses essentially on the tracking of the burning front. The improvement of volume mesh deformation techniques will be addressed in future work.

3 HIGH-ORDER FRONT-TRACKING METHOD

3.1 Motivation

Most grain geometries used in solid rocket boosters are not globally smooth, therefore the direct extension of the previous method yields poor results, as truncated Fourier series of non-smooth or non-periodic functions are polluted by the well-known Gibbs phenomenon (fig. 1 left). A solution to this problem has been proposed by Bruno et al. [2]. Their idea is to decompose a piecewise smooth surface into patches. A smooth, periodic extension of each patch is then computed, yielding a high-order local parameterization for each one of them. This strategy allows to circumvent the Gibbs phenomenon and therefore to recover the good properties intrinsic to spectral methods such as exponential convergence with the number of degrees of freedom.

An outline of our method is given in the following sections, including the dynamic implementation of Bruno's technique in our front-propagation algorithm.

3.2 Segmentation into patches

The first stage consists in extracting smooth, quadrilateral patches from the piecewise smooth initial grain surface. Such decomposition can easily be derived from a CAD file, as solids are already defined as unions of smooth faces (Bézier surfaces, NURBS, ...) in most standard CAD file formats. Smooth patches can also be extracted from a surface mesh, given an appropriate feature-detection algorithm [5].

3.3 Front-propagation algorithm

On each patch, computation of normal vectors and other high-order geometric quantities is performed via differentiation in Fourier space. Unlike in [1], Fourier coefficients are not obtained with a Fast Fourier Transform (FFT) algorithm but result from the "Fourier continuation" method presented in [2] and briefly described in the paragraph below.

Fourier continuation

The principle of this technique is the following: let f be a smooth, non-periodic function $f : [0, 1] \to \mathbb{R}$, sampled at n data points. We seek a trigonometric polynomial $\hat{f}$, truncated to $M \leq n$ modes and with period $T > 1$ that matches f on $[0, 1]$ in the least square sense.

$$\hat{f}(x_j) = \sum_{k=1}^{M} a_m e^{\frac{i2\pi}{T} m_k x_j} \approx f(x_j) \qquad (1 \leq j \leq n)\,, \tag{6}$$

with $m_k = k - (\lfloor M/2 \rfloor + 1)$, $(1 \leq k \leq M)$. This is equivalent to solving a least-square minimization problem.

$$\min_{\mathbf{a}} \|A\mathbf{a} - \mathbf{f}\|^2\,, \tag{7}$$

with $\mathbf{a} = (a_1, \ldots, a_M)^T$, $\mathbf{f} = (f(x_1), \ldots, f(x_n))^T$ and $A_{j,k} = e^{\frac{i2\pi}{T} m_k x_j}$.
Solving problem 7 yields the M first coefficients of a trigonometric polynomial which fits f on $[0, 1]$ up to machine precision (see figure 1) and is a periodic continuation of f on a larger domain $[0, T]$. This polynomial can be used to interpolate f between data points or even to approximate its derivatives. The p-th derivative of $\hat{f}$ reads

$$\hat{f}^{(p)}(x) = \sum_{k=1}^{M} \left(\frac{i2\pi}{T} m_k\right)^p a_m e^{\frac{i2\pi}{T} m_k x} \approx f^{(p)}(x)\,. \tag{8}$$

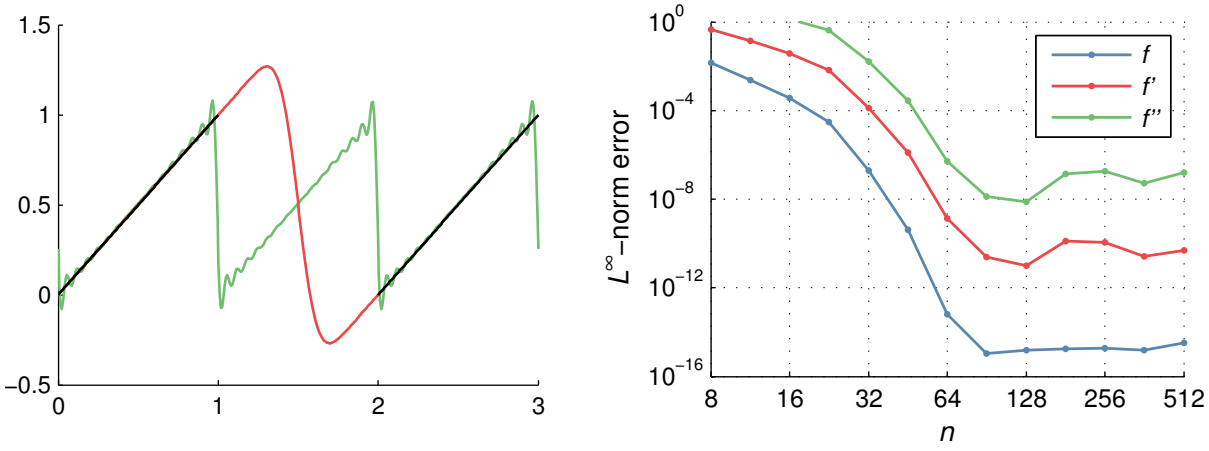


Figure 1: Left: the truncated Fourier series (green) of a non-periodic function (e.g. $f(x) = x$, in black) is polluted by oscillations. However, the periodic extension (red) computed by the Fourier continuation method matches f on its definition domain. Right: maximal approximation error on f and its first two derivatives for various degrees of freedom n.

This technique can easily be extended to higher dimensions, and is used to construct a high-order parameterization of each patch $\Phi : (u, v) \mapsto \mathbf{x} = (x, y, z)$, where Φ is a trigonometric polynomial.

Propagation of the burning front

Time integration of equation 2 is performed using simple forward Euler time stepping, the normal direction being given by the normalized cross products of the two tangent vectors $\mathbf{x}_u$ and $\mathbf{x}_v$.

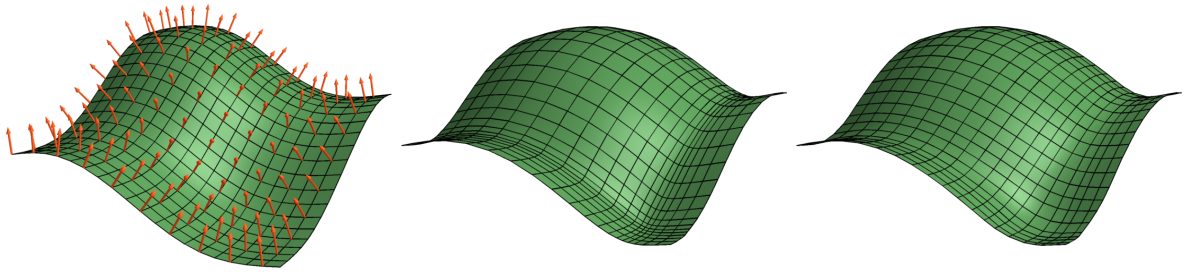


Figure 2: Left: initial surface patch. Center: same patch after an offset along its local normal directions, the surface mesh is distorted. Right: same patch after a few steps of regularization.

Surface mesh regularization

As the surface is transported along its local normal direction, vertices are drawn apart in convex regions, whereas they are clustered together in concave areas, leading to large distortions of the surface mesh. These distortions can be the source of aliasing errors and numerical instability. To prevent this, we add a tangential motion to the vertices. The shape of the surface is then preserved, provided that the boundaries are properly constrained. This tangential motion is governed by an ellipitic operator analogous to a manifold-projected Laplacian that prevents clustering of vertices and leads to a better distribution of vertices along the surface (fig. 2).

In practice, several small steps of tangential regularization are performed for each step of propagation along the normal direction.

Topological representation

Each patch is treated separately by the front propagation algorithm. However, when reconstructing a single mesh for the entire surface, one must handle interactions between patches (stitching, removal of invalid intersections, ...). Toward this end, the patches can be connected into a graph-like structure (*layout*) which stores only topological information. This *layout* is constituted of two types of elements, *nodes* and *segments*:

- a *node* corresponds either to a vertex shared by several patches or to a patch corner, and is linked to the local parametric coordinates (u, v) of the vertex in each of its indicent patches ;
- a *segment* is a pair of connected *nodes*, and can be seen as a portion of a boundary between two patches.

These elements are quite similar to those used for topological representation in several CAD file formats. An example of such a layout is given in fig. 3 for a simple geometry.

3.4 Singularities

The initial geometry might contain slope discontinuities such as sharp ridges or creases. These geometric singularities are necessarily located at patch boundaries. When patches on both sides of such a singularity are propagated individually, they either leave a gap or intersect. In both situations, the surface mesh will become invalid if no special action is taken. In the following two paragraphs, we describe how we handle singularities, thanks to the layout introduced previously.

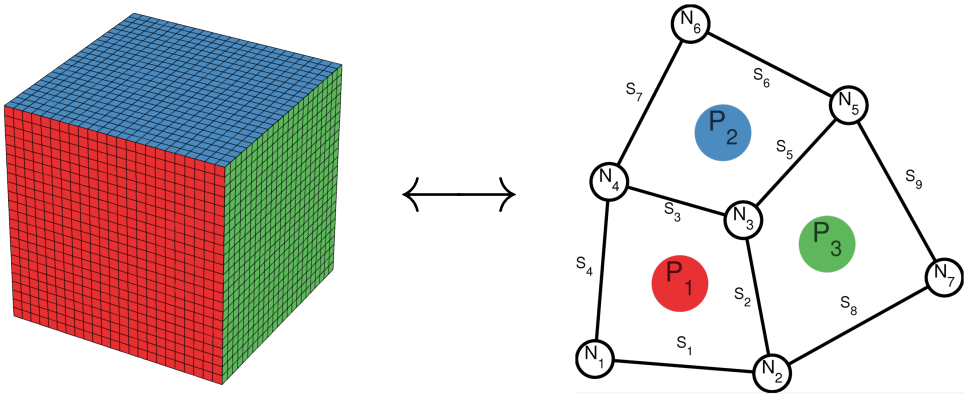


Figure 3: Exemple of a piecewise smooth surface (three faces of a cube, left) and its corresponding layout (right). Each patch is represented in a different color.

Ridges

Patches incident to a ridge separate when propagating as vertices along the ridge follow multiple directions (fig. 4). The non-physical gap thus formed must be filled. Such regions are easily identified thanks to the topological layout. New patches can then be generated from propagating ridges and inserted between the patches that are no longer adjacent. The geometrical construction of these new patches follows the entropy-satisfying Huygens' principle, which is also at the base of the level-set methods [4]. Once constructed, the new patches are treated like the usual ones using the front-propagation algorithm.

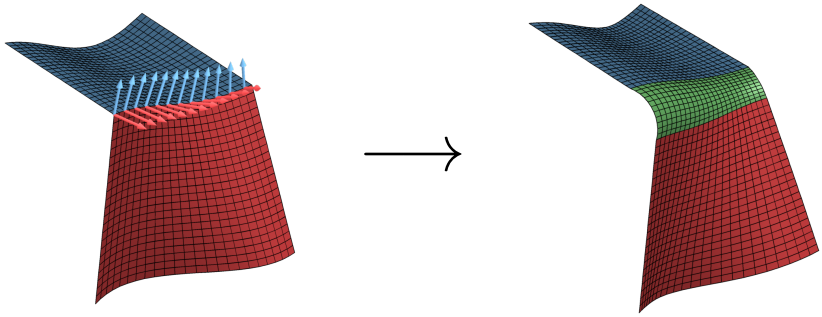


Figure 4: The vertices along a ridge have different velocities in the two incident patches, forming a gap when they propagate. A new patch (green) is created to fill the gap, respecting Huygens' principle.

Creases

Two patches incident to the same crease intersect when being propagated, forming caustics. Only the exterior portion of each patch remains valid. Intersection curves are therefore computed in order to delineate invalid parts in each patch, which are then trimmed off.

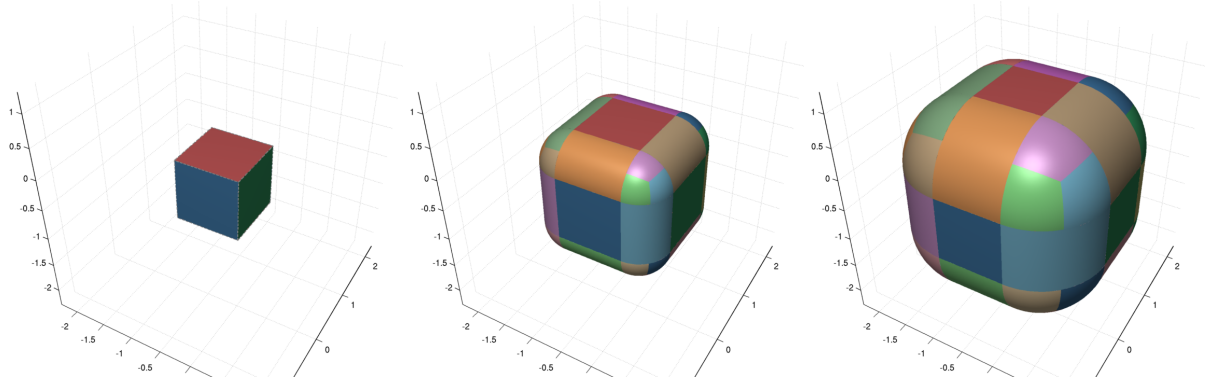
Despite the effort in regularizing the surface mesh, additional caustics are bound to develop in finite time in smooth concave regions due to the very nature of the Eikonal flow. Patches in which this occurs are no longer smooth and thus become prone to the Gibbs phenomenon. If no action is taken, oscillations develop at the vicinity of the caustic and quickly propagate to the whole patch, leading to numerical instability. However, formation of such caustics can be anticipated, as curvature grows indefinitely where new creases are being formed. Thus, we can handle this issue by splitting patches on the verge of becoming singular, yielding two new

smooth patches that are incident to a new crease. This splitting procedure is triggered when local maximal curvature exceeds a threshold value (some fraction of the normal displacement in one time step).

3.5 Preservation of the mesh connectivity

The method described in the previous section is thus able to propagate a piecewise smooth front by treating smooth patches separately and handling interactions between them in order to preserve the topology of the whole surface. However, a single mesh of the front is not yet available. On top of that, the connectivity of the union of all the patches is not preserved, mostly due to addition of new patches. However, this can be solved by constructing a global parameterization of the front based upon the layout described at the end of section 3.3. This structure can indeed be used to gather each patch's local parameterization into a set of charts, forming a global mapping from parameter space to the surface.

Once a global parameterization is available, a unique connectivity can be mapped to the front at each time step. The result is a dynamic surface mesh that keeps the same connectivity over time.



offset distance $= 0$ *offset distance* $= 0.5$ *offset distance* $= 1$

Figure 5: Three stages of the expansion of a cube computed by our method. Each patch is represented in a different color.

4 NUMERICAL RESULTS

Experiments have been run on simple geometries for which analytic results are known. The test exposed in the present section considers the expansion of a unit cube under uniform speed. The convergence of our method is studied using different levels of refinement of the surface mesh: a coarse one and a finer one (with twice as many vertices).

As the cube expands, its surface can be seen as the level-set of a distance function from the initial cube. Mean error on distance is computed in both configuration for an offset distance ranging from 0 to 1 (fig. 6 left).

The area of each patch can also be computed by integrating the area element $d\mathcal{A}$ given by the first fundamental form.

$$d\mathcal{A} = \sqrt{EG - F^2} du dv \,, \tag{9}$$

where $E = \mathbf{x}_u \cdot \mathbf{x}_u$, $F = \mathbf{x}_u \cdot \mathbf{x}_v$, and $G = \mathbf{x}_v \cdot \mathbf{x}_v$ are computed by spectral differentiation, as described in section 3.3. Total area of the expanding cube is then compared to an analytic solution.

$$\mathcal{A}(d) = 8 + 6\pi d + 4\pi d^2 \,, \tag{10}$$

where d denotes the offset distance from the initial unit cube. (Each edge of the cube propagates as a quarter of a cylinder, and each vertex as an eighth of a sphere, yielding formula 10).

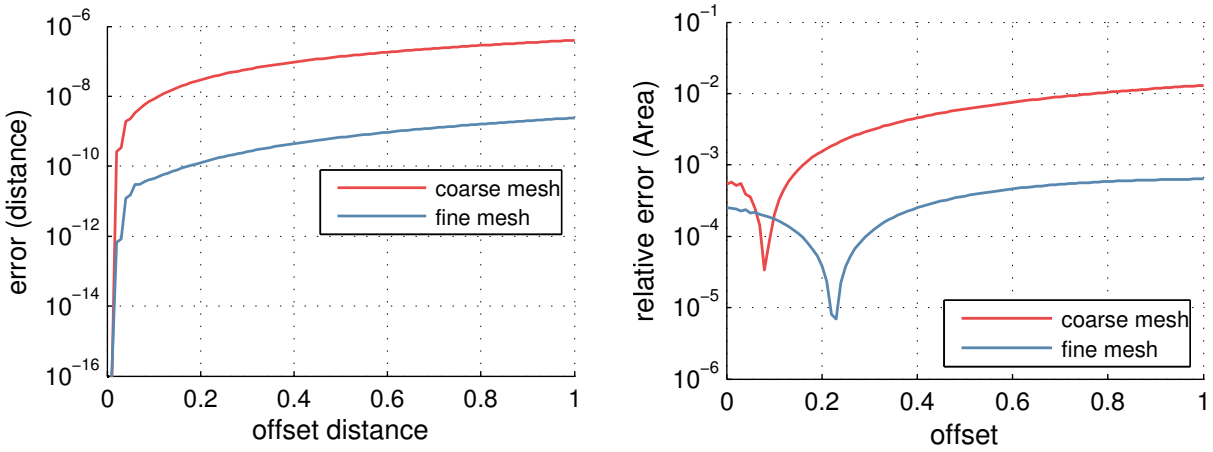


Figure 6: Left: error on distance from the initial cube. Right: relative error on total area.

5 CONCLUSION

We have presented a novel spectrally accurate tracking method for burning fronts. Numerical experiments have demonstrated the validity of the method on simple test cases and further tests of increasing complexity and realism will be performed in future studies. Eventually, this algorithm will be integrated into deformable volume mesh methodologies and coupled to a Navier-Stokes solver for the simulation of the internal fluid flow in a solid rocket motor with burning propellant grain.

REFERENCES

[1] D. Gueyffier, F. X. Roux, Y. Fabignon, G. Chaineray, N. Lupoglazoff, F. Vuillot, J. Hijlkema, and F. Alauzet, Accurate Computation of Grain Burning Coupled with Flow Simulation in Rocket Chamber, *Journal of Propulsion and Power*, **31:6**, 1761–1776, 2015.

[2] O. P. Bruno, Y. Han, and M. M. Pohlman, Accurate, high-order representation of complex three-dimensional surfaces via Fourier continuation analysis, *Journal of Computational Physics*, **227**, 1094–1125, 2007.

[3] S. O. Unverdi, and G. Tryggvason, A Front Tracking Method for Viscous Incompressible Flows, *Journal of Computational Physics*, **100**, 25–37, 1992.

[4] J. A. Sethian, Level Set Methods and Fast Marching Methods, *Cambridge University Press*, 1999.

[5] A. Myles, N. Pietroni, D. Zorin, Robust Field-aligned Global Parametrization, *ACM Transactions on Graphics*, **33:4**, 135:1–135:14, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SINGULAR MESH GENERATION FROM MULTIPLE OVERSET MESHES: A TOOL FOR INDUSTRIAL APPLICATIONS

Gennaro Abbruzzese[1], Marta Cordero-Gracia[2], Mariola Gómez[3] and Nuno Vinha[4]

Departamento Matemática Aplicada y Estadistica ETSI Aeronauticos,
Universidad Politécnica de Madrid, Spain
[1]gennaro.abbruzzese@upm.es
[2]marta.cordero@upm.es
[3]mariola.gomez@upm.es
[4]n.vinha@upm.es

Keywords: overset, mesh, generation, remeshing, moving, chimera, finite volume.

Abstract. *In this work we present a strategy for single mesh generation from multiple overset grids for an unstructured finite volume vertex-centered Navier-Stokes equations solver. The final aim is to slim down the mesh generation process for complex industrial cases, especially in case of mesh displacement, and boundary layer overlapping where Chimera interpolation schemes are often applied. At the same time the tool developed has to take place in between of the mesh generation and solver, in such a way that it is possible to work on the final mesh with no modification of the grid generator code. The validity of the approach will be verified on the basis of the results obtained by running DLR-TAU solver, directly comparing them with the outcome of Chimera interpolation routines.*

1 INTRODUCTION

Mesh generation can be considered the real bottleneck in the CFD analysis process in industry since meshing time consumption widely exceed the solver running time. Furthermore each solver needs a particular mesh, not only for quality requirement but also for topological requirements. The impossibility of the solvers to process any kind of mesh or damaged mesh can cause a huge time loss from the mesh generation side. In academia the problems coming from a bad mesh are generally overcome by employing refined and up to date routines but in an industrial environment often is not possible to change the tools or put the hands in the source code of the software employed. This is the key to understand the need of a tool which in the CFD investigation process is placed between the mesh generator and the solver.

The intent of this work is to develop a method to build/optimize meshes employed by vertex-centered DLR-TAU solver in such a way that it could be used to process any kind of overset, hybrid, non-matching boundaries, damaged or negative volume grid. The scheme adopted by DLR-TAU comes from the family of the Godunov type methods [1], wide spread for compressible flow prediction, which smooths the fluxes distribution solving the Riemann problem between each cell.

Common practices suggest to use classical geometrical quality criteria to evaluate computational grids. Especially for schemes as vertex-centered finite volume, which manipulate the initial mesh, this evaluation is not sufficient if not misleading. A good looking mesh does not bring always an accurate solution. However the influence of mesh features on the CFD predictions are far from being clear [2] [3] [4] [5] and we lack of actual guidelines to build a high quality mesh.

For viscous flow calculations special feature for the cells are desirable, one being a structured aligned hexahedral mesh in boundary layers. This brings up to several issues and obstacles with regards to automatic of the mesh generation, for example:

- Mesh generation tools often have problems dealing with complex geometries;
- In moving meshes or in the case of relocating components, the mesh must be regenerated;
- The presence of a structured mesh creates a huge amount of cells which are far from boundary layers, for example in the wake and in the free stream regions around aerofoils.

In order to overcome lack of accuracy or resolution in a mesh, generally techniques are employed as adaptation to increase resolution, a-posteriori evaluation of the mesh, accurate schemes on the solver side or interpolation of solution on overlapping grids (Chimera approach). In particular, this last approach is gaining popularity in the industry and interest from the academic community and DLR-TAU has a robust implementation of this scheme. It consists in a set of high quality meshes generated independently and communicating by mean of interpolation procedures in overlapping zones and presents some desirable properties as, on the other hand, some disadvantages [7][8]:

- Provide good quality mesh,
- It is versatile even for highly complex geometries,
- Relatively well tested,
- Allows to skip moving parts re-meshing.

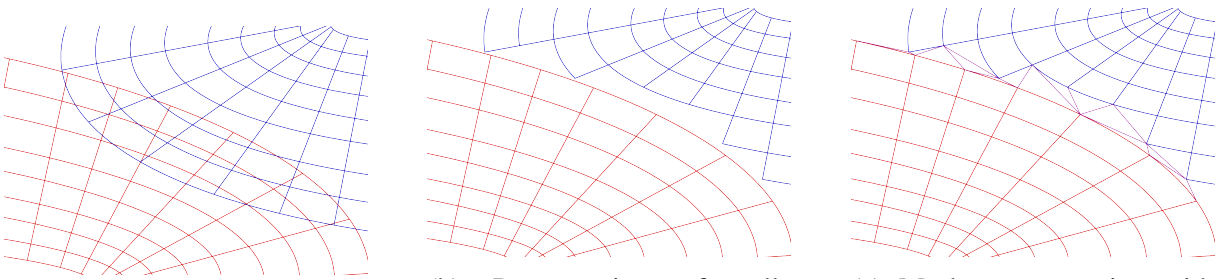
But:

- Higher computational effort,
- Poor solver performance,
- Increased problem setup time,
- Flux non conservative in presence of interpolation and non-matching boundaries.

An attempt to bypass interpolation problems among overset grids and to generate a single dual mesh has been done by Sørensen [9]. In his paper the author builds an algorithm in ten steps to merge two different grids in a single one. The intention of the author is to potentially fully automate the meshing process, to employ the same solver of a single component grid, to avoid interpolation and to have control on the whole domain mesh. Gremmo *et al.*[10] have presented research notes about the implementation of a tool for partial remeshing of background mesh when a solid body and relative mesh are subject to large displacements. In the following paragraph a sequence of actions are shown, to build incrementally a multi-component mesh for a three component *GARTEUR* A310 aerofoil [11][12].

2 THE ALGORITHM

The idea at the basis of the algorithm we are looking for is similar to the one coming from Sørensen [9] with a focus on primary mesh, which makes the implementation simpler and robust. The idea is still to start from two or more grids and supply a single one for the reasons explained above. The outcome of the process has to present the desired features minimising the eventual distortion. The problem is in principle simple to solve, but in practice presents a multitude of issues. First step is, of course identify the zones covered by the different grids. Once identified the new boundaries of each mesh, it is necessary to fill the holes corresponding to the degenerated cells in proximity of the interfaces (fig.1).



(a) Overlapping meshes.

(b) Degeneration of cells around interface.

(c) Mesh reconstruction with constrained triangulation.

Figure 1: Degeneration and reconstruction process in the interface proximity

2.1 Interface definition and cells degeneration

Modifications of the topology of the two meshes will be applied in the overlapping zones. Once identified, these areas must be transformed in single meshes, always taking into account the constraints coming from the non-overlapping meshes. The reason is that it is preferable to maintain the original meshes untouched where possible, since we can assume that these are

generated in order to fulfill some criteria imposed by the user for the particular case. This point raises another problem when the meshes have equal priority. If we try to maintain the topology of the original meshes and at the same time we try to modify the smallest possible number of cells in the mesh we will have to choose only one among the meshes in each zone in the entire overlapping region. The objective is then to split the overlapping area by mean of area interfaces based on mesh priority criteria. The employment of segments in two dimensions or surfaces in the three dimensional case as interface allows us to degenerate the minimum amount of cells in the entire domain and to substitute the cells of each mesh with higher quality cells coming from other mesh. There are many possible formulations for obtaining the interface. A way to define such interface is by mean of a criterion based on the distance from the walls. Simply, the elements lying in the overlapping zones and having points closer to a wall belonging to another mesh will lose priority and replaced by cells from the other mesh or, if needed, by new cells specifically created.

With this approach:

- It allows to deal with the problem of cells with equal priority (and it should be employed for this cases only);
- User defined cells like boundary layer structured mesh for viscous flow simulations will remain untouched as much as possible;
- Assuming that the grids are all built for the same purpose, it is reasonable to think that they will have same y^+ and same growth ratio and, then, that their cells will have similar dimension in proximity of the interface.

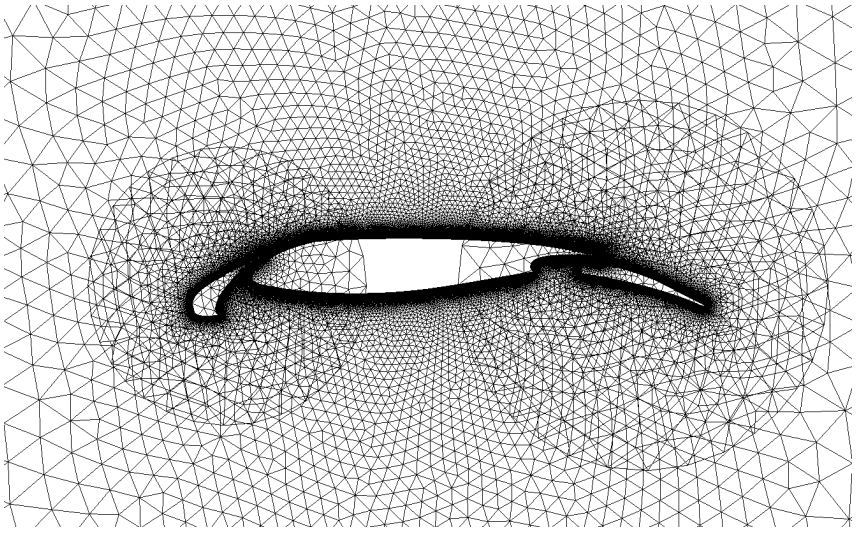


Figure 2: Overlapping meshes.

The outcome of such degeneration is shown in figure 3. It is possible to see that the slat lies in the boundary layer of the wing. This is a typical case where unstructured grid generation software find difficult to handle the complexity of the geometry. In the picture 4 the median line between the two walls is shown which is implicitly defined by the criterion used. All the cells crossing the median line are deleted. By the way the cells to be eliminated have been marked with a simple distance check and with the AABB tree routine in *CGAL* [13] to speed up the calculation.

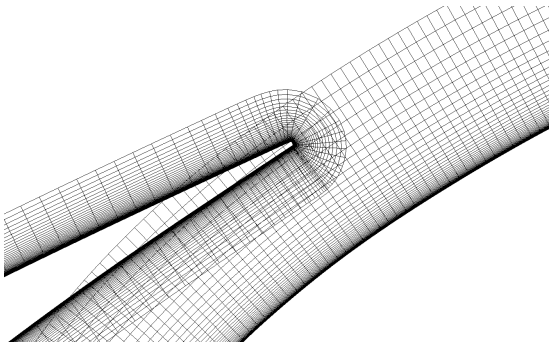


Figure 3: Overlapping boundary layers of wing and slat.

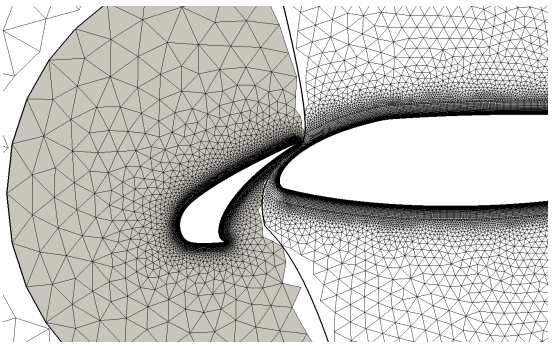


Figure 4: Degeneration of cells external to area defined by distance from teh walls.

2.2 Single mesh reconstruction

The simplest and most efficient way to reconstruct the mesh in arbitrary shaped hull (an example in fig.5) is the triangulation. Of course we can expect that some bad quality elements will be created in this way, but also it will be possible to apply smoothing tools once the single mesh will be created. For this reason we resorted to the *CGAL* [13] library. The use of this tool will guarantee a high quality grid as output and a high level of automation of this step. After the degeneration the two meshes are not overlapping anymore and they are separated by an empty hull. In order to refill this hull *CGAL* constrained Delaunay 2D triangulation routine has been employed. This triangulation still creates a convex polygon as a simple Delaunay triangulation but forces the creation of the edges on some constraints. In this case the borders of the void hull are imposed as constraints in order to obtain a triangulation which matches original meshes elements. The polygon cells outside the void hull borders are ignored and the remaining are added to the full mesh to patch the two degenerated meshes. The full single mesh is in this way generated.

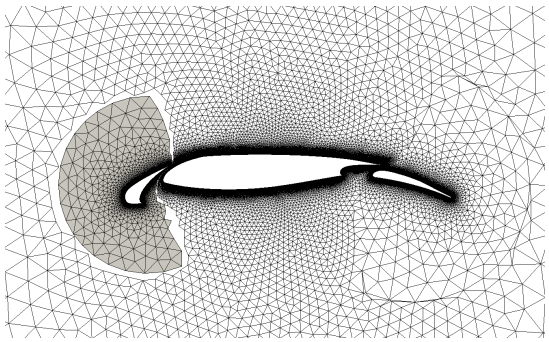


Figure 5: Slat mesh added to the wing background mesh right before joining.

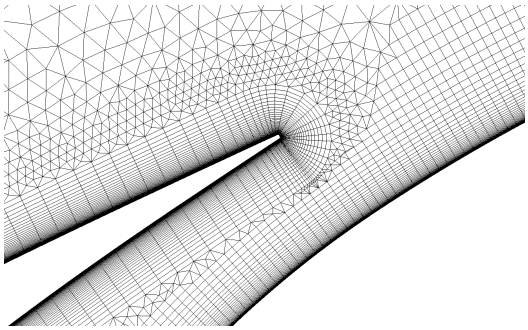


Figure 6: Boundary layers of wing and slat.

In the picture 5 it is possible to see the mesh where the flap mesh has been already degenerated and mended and the same process has to be applied to the slat. In the figure 6, the mesh between the slat and the wing has been magnified. As predicted the triangulation in this zone has a favorable quality since it has been preformed in a space defined by elements of comparable dimension. In the figure 7 we can see the full single mesh after reconstruction and observe that the same said for near wall zones is not valid walking away from the walls. Furthermore, the cells have been intentionally left untreated to check the influence of such small amount of

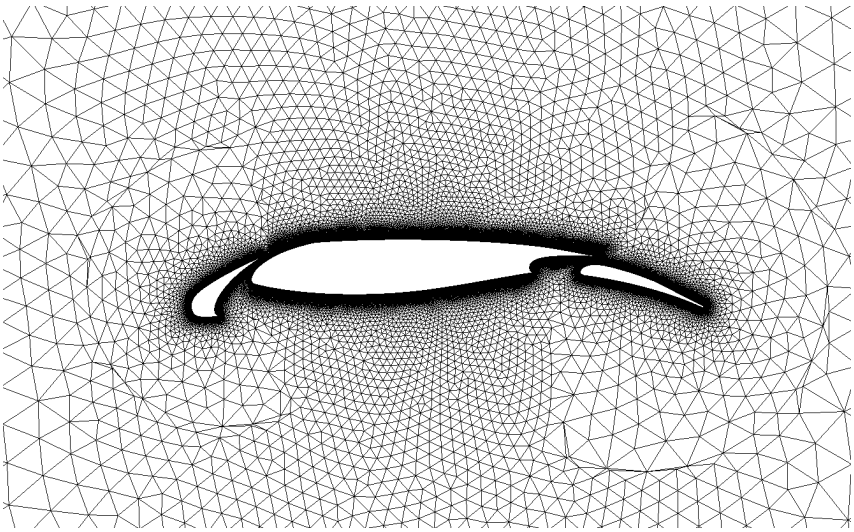


Figure 7: Final incrementally generated mesh.

bad cells. In this case for example, the cells created in the process, by the addition of the slat, are 292 out of a total number of 191025 cells in the final full mesh.

3 A BRIEF TEST

3.1 Cases setup

In this paragraph we show a small test with *DLR-TAU* finite volume Navier-Stokes equations solver. The simulations have been run in order to get an idea of the possible loss of accuracy of the prediction in a manipulated mesh comparing it with a mesh fully generated simultaneously for the three components by the automated mesh generator available in *CENTAUR*[14]. The mesh is shown in figure 8. We can notice that, although it has a number of points comparable to the one of the incrementally generated mesh (tab.1), it has a different distribution of points. From the zoom of the boundary layers between slat and wing in fig.9 we can observe that the complexity of the geometry is an obstacle for the mesh generator since the growth of the boundary layers is limited by the space between slat and wing.

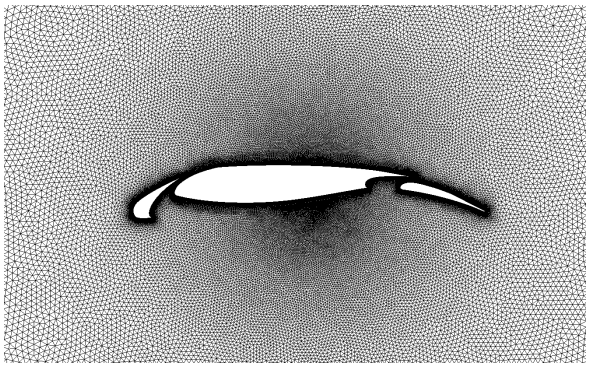


Figure 8: *CENTAUR* mesh simultaneously generated for the three components.

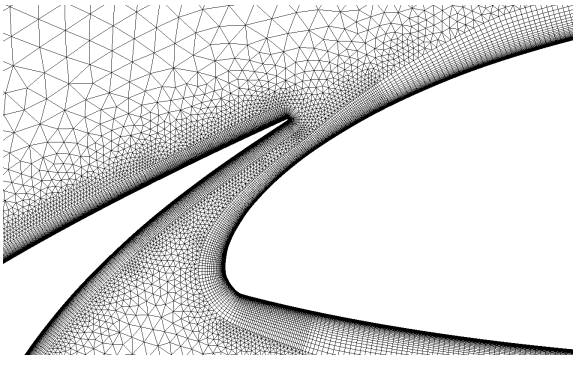


Figure 9: Zoom of the boundary layers between slat and wing in *CENTAUR* mesh simultaneously generated for the three components.

Furthermore, in order to evaluate the advantages coming from the presented method, the same meshes of the components individually generated (still with *CENTAUR*) have been used to run a Chimera simulation without applying any topology modification (fig. 2). The table 1 shows the number of points of the meshes.

Mesh	Number of points
Incremental	***345740***
3-Components	***330784***
Chimera (slat)	*29892*
Chimera (wing)	*165212*
Chimera (flap)	*156722*
Chimera (total)	***351826***

Table 1: Meshes dimension.

The cases are a quasi 2D simulations with surface mesh with $y = 0$ extruded to $y = 1$. Boundary conditions imposed in TAU are *farfield* boundary conditions for farfield faces, *viscous wall* for the solid bodies faces and *symmetry plane* for the plane in constant $y = 0$ and $y = 1$ [15]. All the meshes have been run with Spalart-Allmaras one-equation turbulence model under the conditions summarized in table.2.

operating conditions	
Re	0.605×10^7
Mach	0.224
AoA	12.19°
wing configuration	*take-off*

Table 2: Operating conditions for the three cases.

3.2 Results

The aim of this test is to reveal any possible malfunctioning of the mesh incrementally generated with the method explained above. It is certainly not detailed enough to fully validate the newly assembled mesh but as we will see, it becomes useful to confirm some valuable features of the algorithm. All the simulations have been run until convergence of the lift and drag coefficients. It is important to point out that the Chimera scheme could not reach a calculation speed over 35% of the speed of the scheme employed in the other two simulations. All of them have shown a rapid convergence of these parameters while the residuals of the density seem to have encountered a resistance around 10E-5.

Figure 10 shows the plot of lift coefficient history for the three simulations. We can observe that all of them are converging rapidly with a small percentage of error to the final value. The *CENTAUR* mesh has at the beginning high fluctuations of the value while the other two have a smoother behavior. With this data is still not possible to determine if the smooth curves are similar because of the similarities in mesh topology or to a higher amount of cells in the boundary layers. The converged values of the lift coefficient are summarized in the table 3 with the error relative to the experimental results [11].

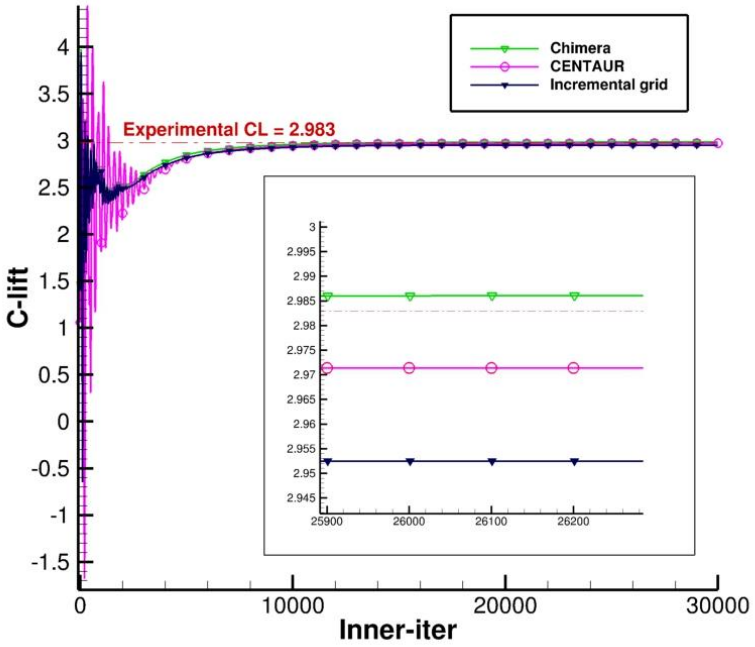
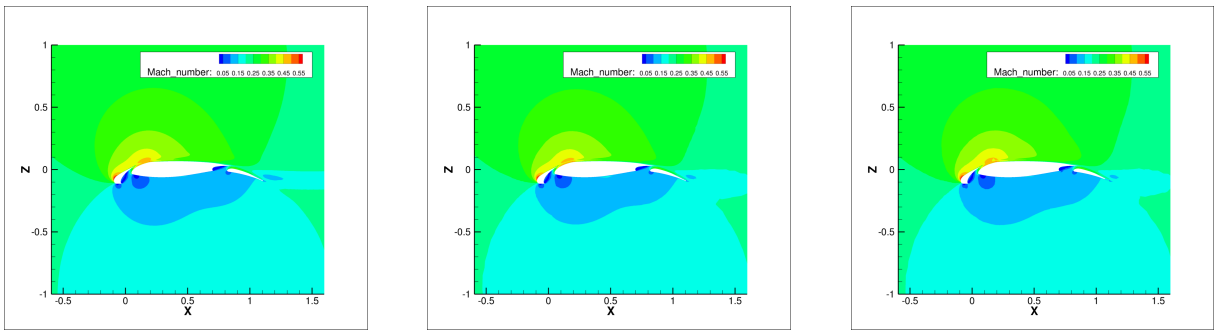


Figure 10: Lift coefficient history for the three simulations.

lift coefficients (exp = 2.983)		
	CL	*err(%)*
Incremental	2.953	1.006
CENTAUR	2.971	0.4022
CHIMERA	2.987	0.4

Table 3: Lift coefficient result and error relative to experimetal results [11].

The mesh run with Chimera scheme seems to have the best behavior among the three. Still the data is not sufficent for a verdict. With only this test is not possible to determine if the error is sistematic. What is possible to see, nevertheless, is that the assembled mesh has an error on the prediction which does not overcome the 1%. Furthermore, although no mesh smoothing has been applied, the convergence is not affected by the bad cells generated with constrained triangulation, as it could be expected. We continue the observation of the results with the plot of the contours of the Mach number (fig.11). This time the Chimera seems to have an inferior performance among the three, showing a ”pleating” of the contours in proximity of the wing downwind to the slat, whereas the assembled mesh seems to be the one which better captures the gradient between the flows coming from the upper part of the slat and the channel composed by slat and wing.



(a) *CENTAUR* mesh. (b) Incremental mesh. (c) Chimera mesh.

Figure 11: Contours of the Mach number for the three simulations

4 CONCLUSIONS

A method for mesh assebling by incremental components addition has been implemented and tested. With the small amount of data available is not possible to have a verdict on the performances once the mesh is used for aerofoil flow simulation. Nevertheless, it has been shown that the method has some valuable features. The algorithm is a simple way to give flexibility to the meshing problems when neither the solver nor the mesh generator codes are available. It is easy to automate and presents some of the advantages of the Chimera approach but with no need to introduce interpolation with a remarkable time saving and with a better accuracy for complex cases. Hence, it would be interesting to develop the algorithm and further studies can be carried out with a more comprehensive set of test-cases but also introducing a smoothing of the newly created cells and, of course, expanding it to real non symmetrical 3D cases.

REFERENCES

[1] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics : a practical introduction*. Springer, Berlin, New York, 1997.

[2] F. J. Dannenhoffer, *Correlation of grid quality metrics and solution accuracy for supersonic flows*. AIAA-2012-0610, 2012.

[3] B. Diskin and J. L. Thomas, *Accuracy of Gradient Reconstruction on Grids with High Aspect Ratio*. NIA Report No. 2008-12, 2008.

[4] B. Diskin and J. L. Thomas, *Mesh effects on accuracy of finite-volume discretization schemes*. AIAA-2012-0609, 2012.

[5] B. Diskin and J. L. Thomas, *Notes on accuracy of finite-volume discretization schemes on irregular grids*. Applied Numerical Mathematics, 2009.

[6] S. Crippa, *Application of novel hybrid mesh generation methodologies for improved unstructured cfd simulations*. 28th AIAA Applied Aerodynamics Conference - CFD Drag Prediction Workshop Results, 28 June - 1 July 2010, 2010.

[7] S. Crippa, *Application of chimera with hexahedral blocks in solar meshes*. Third Symposium Simulation of Wing and Nacelle Stall, 21st - 22nd June 2012, 2012.

[8] S. Crippa, *Application of Chimera with Hexahedral Blocks in Solar Grid*, PerPre-Deliverable 09.2011, $C^2A^2S^2E$, Institute of Aerodynamics and Flow Technology, DLR.

[9] K. A. Sørensen, *Multi-component modeling with the dual mesh connect approach*. Computational Flight Testing: Results of the Closing Symposium of the German Research Initiative ComFliTe, Braunschweig, Germany, June 11th-12th, 2012.

[10] S.Gremmo, I.Antonik, R.Nigro, C.Hirsch, and G.Coussement, *Partial remeshing strategy for cfd simulations when large displacements are taken into account*. 23rd International Meshing Roundtable (IMR23), Research Notes, 2014.

[11] F. Manie, O. Piccin, and J.P. Ray, *Test Report of the 2D model M1 in the Onera F1 Wind Tunnel*, GARTEUR High Lift Action Group AD (AG-08), TP041, October 1989.

[12] M.C.Iorio, L. M. Gonzalez and A. Martinez Cava, *Global Stability Analysis of a Compressible Turbulent Flow around a High-Lift Configuration*. 45th AIAA Fluid Dynamics Conference, 22-26 June 2015, Dallas, Texas, USA. pp. 1-25.

[13] *CGAL*, Computational Geometry Algorithms Library.

[14] *CENTAUR*, www.centaursoft.com.

[15] TAU-Code User Guide Release 2015.1.0

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

GENERATING TRIANGULAR LATTICES FOR SURFACES WITH IRREGULAR BOUNDARY

Tatiana Sá Marques[1], and Vitor Dias da Silva[2]

[1]Department of Civil Engineering, University of Coimbra
Rua Luís Reis Santos - Pólo II, 3030-788 COIMBRA, Portugal
e-mail: t.marques@cobagroup.com

[2] INESCC, Department of Civil Engineering, University of Coimbra
Rua Luís Reis Santos - Pólo II, 3030-788 COIMBRA, Portugal
e-mail: vdsilva@dec.uc.pt

Keywords: Lattice, Delauney Triangulation, Bézier Curves, Ringwise Triangulation.

Abstract. *In this contribution algorithms for the generation of triangular lattice meshes for surfaces with an irregular boundary are presented. A central objective of this study was to get meshes with the maximum possible number of bars of a given length. Three different approaches are described. In all three the meshing process starts at the boundary and advances to the central part of the domain, by defining successive rings of bars. The differences between the three algorithms lie on the different ways as the problem of node elimination and smoothing is solved, as the rings become smaller. Algorithm A relies on the Delauney criterion for mesh smoothing. Algorithm B uses Bézier curves to approximate the cloud of initially generated points in a new ring, when smoothing is necessary. In algorithm C first an automatic elimination of points is performed, when the points in the new ring ate too close, which is followed by a relocation of interactively chosen points. A representative example is presented, allowing a comparison of the performance of the three approaches.*

1 INTRODUCTION AND LITERATURE REVIEW

The purpose of this work was to develop a pre-processor with a specific automatic mesh generator that can be used to study grid shells. The constraints imposed a priori were: to use a triangulated system to prevent the excessive deformability and resulting instability of non-triangulated systems; to develop a 2D generator based on a sound theoretical footing that could be extended to grid shells; to create a triangulated system of continuous rings that should enable the analysis of different stages of a self-supporting construction, since there is a belief in this potential capacity in these type of structures (double curvature shells); to use as many bars of equal length as possible, thereby offering advantages in terms of both aesthetics and the construction process.

Three separate algorithms were developed that correspond to different approaches of the used method and solve the smoothing of the mesh, which allows a geometric solution without the elements overlapping. Algorithm C corresponds to the semi-automating of a manual pre-processor with automatic generation of the triangulation with bars of equal length, followed by slight automatic smoothing (getting rid of very short bars), and manual smoothing (manual choice of points to be displaced). The heavy reliance on manual smoothing makes it difficult to generate two identical meshes. Algorithm B adjusts the Bézier curves to a series of points. The Bézier curves are widely used in computer graphics, but no report of their use in mesh generation was found in the literature. Bézier curve adjustment is used for smoothing between the rings of the mesh. Coupling the triangulation programmes in algorithm C, the Bézier curves and creating the connection mesh gave algorithm B. Algorithm A is the only wholly automatic generator, based on geometric criteria and it fulfils the objectives set initially.

The efficacy of the computer programs is demonstrated in the numerical applications. Lattices generated on a plane surface with a irregular curved boundary, and obtained with the three algorithms, are presented. The results are compared with respect to the total number of bars, the number of same-size bars and the regularity of the meshes.

References in the literature on the automatic generation of meshes based on geometric criteria are scarce in comparison with references on automatic generation based on adaptive finite elements techniques. No models were found in the literature that could be directly adapted to solve the specific problem of double curvature grid shells.

Of the three types of mesh - structured, unstructured and hybrid - this work uses unstructured meshes because there may be a variable number of adjacent elements in the internal nodes, without explicit connective relations and they are more adaptable to arbitrary and/or irregular geometry domains. Unstructured meshes also allow local and adaptive refinement, impossible to achieve in structured meshes.

With Sibson, 1979 [1] and Watson, 1981 [2], we find triangulation algorithms of domains based on the Delaunay triangulation technique. S.H.Lo, 1985 [3] presents a discretization algorithm of arbitrary planar domains in triangles that generates finite element meshes based on the advancing front and Delaunay triangulation techniques. Chew, 1989 [4] and S.H.Lo, 1989 [5] present restricted Delaunay triangulation versions applied to planar domains of irregular, non-convex boundaries. Zienkiewicz and Wu, 1994 [6] and El-Hamalawi, 2004 [7] introduced the possibility of controlling the direction of mesh elongation.

2 DOMAIN BOUNDARY

The domain boundary is defined by means of a polygon and interpolated by means of Bézier curves, as shown in Figure 1. Each of the vertices in Figure 1 is considered in turn, as vertex

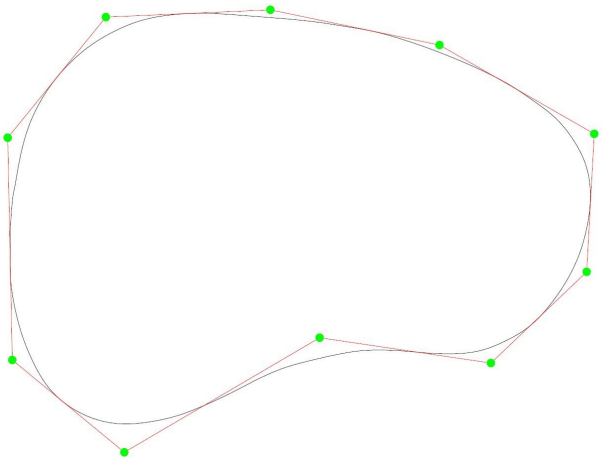


Figure 1: Base polygon and boundary (defined by a set of Bézier curves).

V_i, defined by the adjacent segments designated generically as V_{i-1} and V_{i+1}. In each segment, $V_{i-1}V_i$ and V_iV_{i+1}, the midpoints are designated respectively by 0 and 3. The midpoints will be the endpoints of the Bézier curve, as shown in Figure 2-a. The curve developed between points 0 and 3 and is controlled by means of two more points, 1 and 2, as shown in Figure 2-b, with the coordinates given by means of a parameter α

$$x_1 = (1-\alpha)x_{V_i} + \alpha x_0 \quad (1)$$
$$y_1 = (1-\alpha)y_{V_i} + \alpha y_0 \quad (2)$$
$$x_2 = (1-\alpha)x_{V_i} + \alpha x_3 \quad (3)$$
$$y_2 = (1-\alpha)y_{V_i} + \alpha y_3 \quad (4)$$

In the set of Bézier curves depicted in Figure 1 $\alpha = 0.35$ has been used.

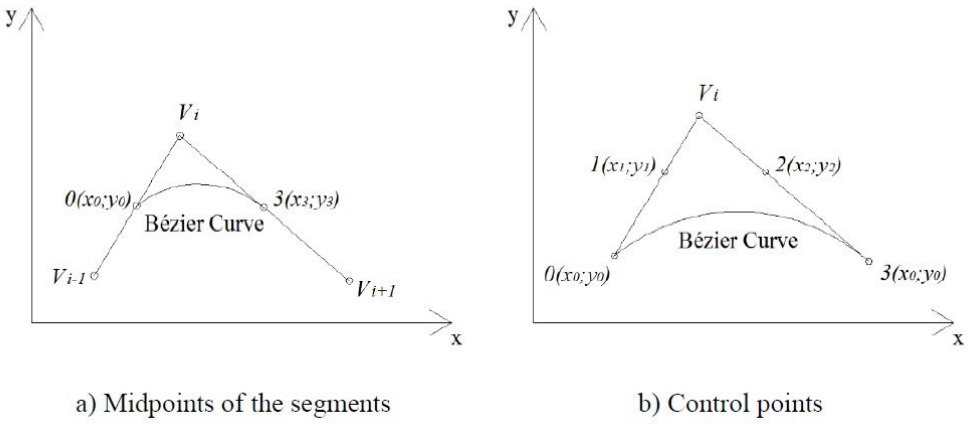


Figure 2: Construction of the Bézier polygon and its curve for each vertex V_i.

3 LATTICE GENERATION

3.1 Common features to the three approaches

In every of the three considered approaches the first step consists of defining the nodal points that lie on the boundary. The distance between two neighbouring points is as close as possible to the targeted length of the bars C_b, under the condition that all these distances have the same value.

Once these points are defined, a new interior ring of points is defined, which lie on the left, as the progression on the boundary points takes place in an anti-clockwise direction. Considering two consecutive points, A and B, on the boundary, the coordinates of a point C of the new ring are obtained by the expressions (Figure 3)

$$x_c = x_a + C_b \cos\gamma \tag{5}$$

$$y_c = y_a + C_b \sin\gamma \tag{6}$$

with $\gamma = \beta + \theta$ and $\theta = \arccos\left(\frac{\sqrt{(x_b-x_a)^2+(y_b-y_a)^2}}{2}\frac{1}{C_b}\right)$.

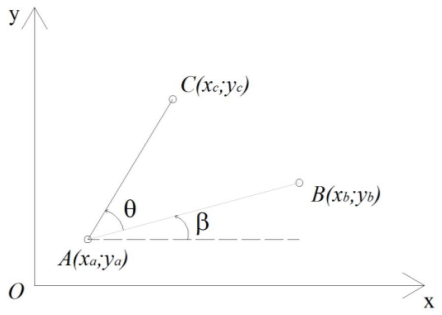


Figure 3: Determination of point C.

After the needed rings have been generated, using one of the three algorithms described in the next sub-sections, the lattice is closed by linking every point on the last ring with a mean point, as shown in Figure 4.

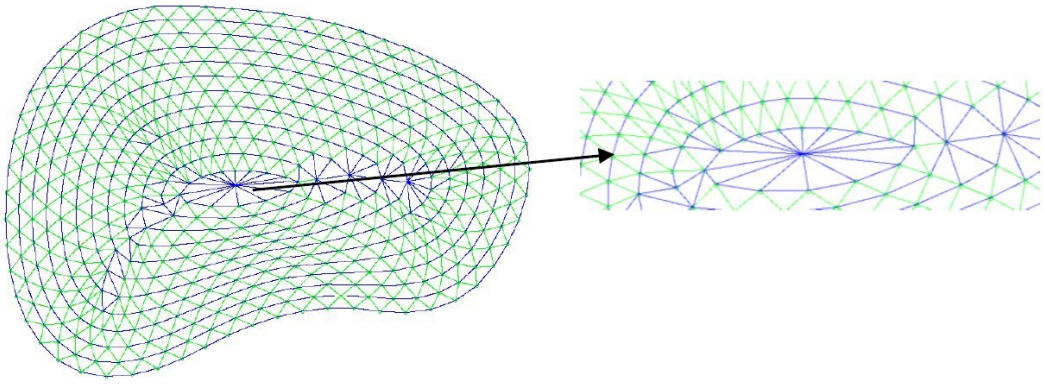


Figure 4: A complete mesh with detail of mesh closure.

3.2 Algorithm A

After the formation of ring we have to update the generation front that serves as the base for the construction of the next ring, i_a. The ring ia will be formed from i points of the ring i_{a-1} and will give k points that will be the generation front of the ring i_{a+1}. The number of k points is equal to i under these conditions:

- the distance between points k and $k+1$ is greater than the value accepted, according to the smoothing criterion established for the distance;
- the triangulation obtained with i and k points are Delaunay triangles.

In the absence of any these conditions the number of k points is smaller than i, and we need to determine which points will constitute the new generation front.

The main property of the Delaunay triangulation or Delaunay criterion is that no point of a triangular element can belong to the interior of the circle defined by the three vertices of another triangle. Figure 5 shows an example in which the criterion is violated by the presence of point P inside the circle that circumscribes triangle ABC.

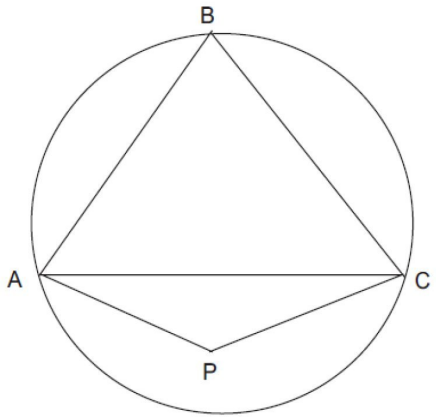


Figure 5: Violation of the Delaunay criterion.

The points that do not obey the Delaunay criterion are detected as follows:

1. For each set of i and $i+1$ points a k point is determined, and the circumference that passes through i, $i+1$ and k. The circumference is defined from the intersection of the perpendicular bisectors of the edges of the triangles.
2. Once the circle that passes through i, $i+1$ and k has been determined, those k_j points formed by all the sets of 2 successive points of the front located inside the circle are verified. The existence of an arbitrary point P inside the circle is verified by comparing the distance of P from the centre of the circle O, d_{PO}, with the radius of the circle: if d_{PO} is the same as or greater than the radius, then P does not violate the Delaunay criterion and thus can be regarded as a valid point. Thus the lack of overlapping elements in the triangulation is ensured.
3. All the points of the generation front that give rise to triangulations that do not meet the Delaunay criterion are included in the i line of the matrix corresponding to the circle formed by i, $i+1$ and k.

4. At the end of the verifications, a matrix $[mp]_{nx\times(n+1)}$ is obtained, in which n is the number of generation front points.

5. After constructing the matrix $[mp]$ it is necessary to define those k points that should be eliminated from future generation front and triangulation to be carried out with the i points that will no longer have two-way correspondence with k. This procedure is called matrix condensation.

Condensation of the detection matrix $[mp]$

The matrix $[mp]$ contains information about overlaps that will occur in the formation of the ring if the triangulation continues with the same number n of front points as were used in the formation of the previous ring. With this information we must decide which points to keep on the generation front, which points to remove, what new points to introduce and how the triangulation should be done when new points are eliminated or created. The method follows a logic similar to Gauss-Jordan elimination. The decisions taken, with regard to the above situations, are as follows:

- If the circle C_i contains point $n + j$, then circle C_j is deemed to contain point $n + i$, and the occurrence is referenced on the matrix $[mp]$. This procedure requires going through each of the columns $j = npc_a + 2, \ldots, npc_a + n + 1$, of $[mp]$ and putting in the first row where the value is not null all values of the points that give rise to the non-violation of the Delaunay criterion.

- Once the sets of points to be removed from the generation front are established, we must check if some of the sets intersect. If they do, the matrix $[mp]$ is transformed so that the points to remove are between the smallest and largest points in the sets, taking into consideration the special case of the set of points that encloses the end and the beginning of the ring.

After the matrix [mp] condensation we know which points to remove, so that there is no overlap of triangulation when processing the ring.

Elimination/creation of new points

Whenever it is necessary to remove generation front points this results in the distortion of the mesh, which is greater the higher the number of points to be eliminated. Criteria must be introduced to lead to the least distortion possible. Several options can be taken. The following strategies were adopted in the algorithm developed:

- when an i point is removed from the generation front, the end of the triangulation segments that would connect to the i point pass to the $i - 1$ point;

- when two or more points are removed from the generation front, a point is created, a midpoint, whose coordinates are the average of the coordinates of the points removed and the two adjacent points (before and after) and which remain on the front. The end of all triangulation segments which would connect to the eliminated points replaced then has its end at the midpoint.

These criteria proved to be effective in the triangulations carried out. Other criteria could be introduced, in particular: the successive use recursive Delaunay criterion; consideration of

the criterion of smoothing angle (angle limitation); introduction more than one midpoint when distances are greater than $2 \times C_b$. The first two options did not show good results in the examples studied.

Distance-based smoothing

Different smoothing methods were developed in this work. One eliminates the bars whose length is less than a percentage value of the length prescribed for the bars (C_b) and this was applied to algorithm A. Whenever the distance between consecutive points is below the present value, the points are removed from the generation front such that the bars converging at those points will converge at the point immediately preceding it. These considerations are valid for all the bars, apart from those converging on the last k point of the ring, which are converging at the first point of the ring. Figure 6 illustrates the application of this method in a mesh and Figure 7 shows local enlargement of the smoothing.

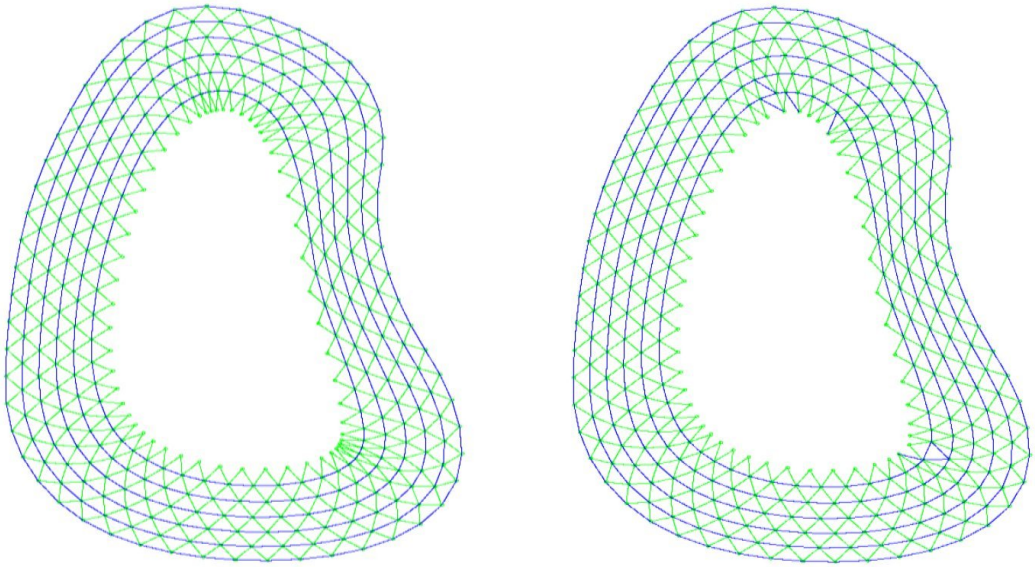
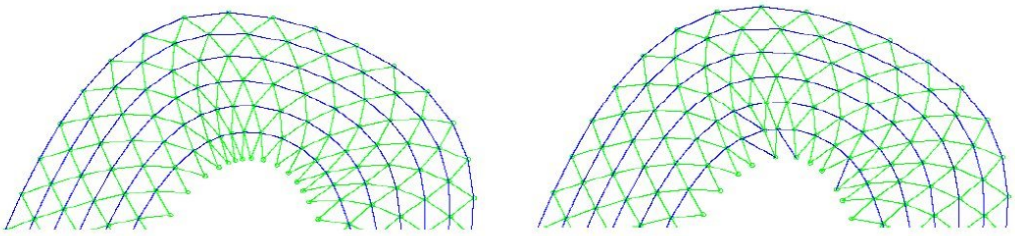


Figure 6: Overall effect of distance-based smoothing process.



a) Before smoothing b) After smoothing

Figure 7: Local enlargement of effect of distance-based smoothing.

3.3 Algorithm B

In each ring we have to update the generation front that serves as the base for the construction of the next ia ring. The ia ring will be formed from npr (number of points of the ring) of the $ia-1$ ring and will give k points that will be the generation front of the ring $ia+1$. In this algorithm, the number of k points is equal to npr, except when the smoothing based on the Bézier curves is applied. In that case, the number of k points is generally different from npr, depending on the number of points given by dividing the curve into segments of equal length, C_b.

As mentioned above, after the formation of a ring, smoothing corresponds to making the initially generated npr points approximate a curve. The goal is to use the Bézier curves to find the curve that best approximates the npr points of the ring. This closed curve can be related to a base polygon formed by a number of previously established points, as was analysed in the definition of the domain boundary. The number of points of the polygon, nv should be chosen/tested to adjust/fit the description of the desired geometry with sufficient accuracy. In the examples presented in this paper five-node polygons ($nv = 5$) have been used. With the number of base points set, we have to find the coordinates of the points corresponding to a given fixed value of the parameter α, which best allows a fit between the curve and the points. The curve is obtained iteratively, using the procedure set out in the following. Once the curve that best fits the points of the generation front, we can recalculate the value of the α parameter which leads to a total length of the curve that is a multiple of the length of the pre-defined bar, C_b. The value of α is also determined iteratively. The iterative method considered in the two approximations was the Newton-Raphson method.

The main steps of this procedure are

1. Using the procedure described in sub-section 3.1 new rings are generated, until any of the interior bars become smaller than a given value, which, in the example presented in Figure 8, occurs in ring 7.

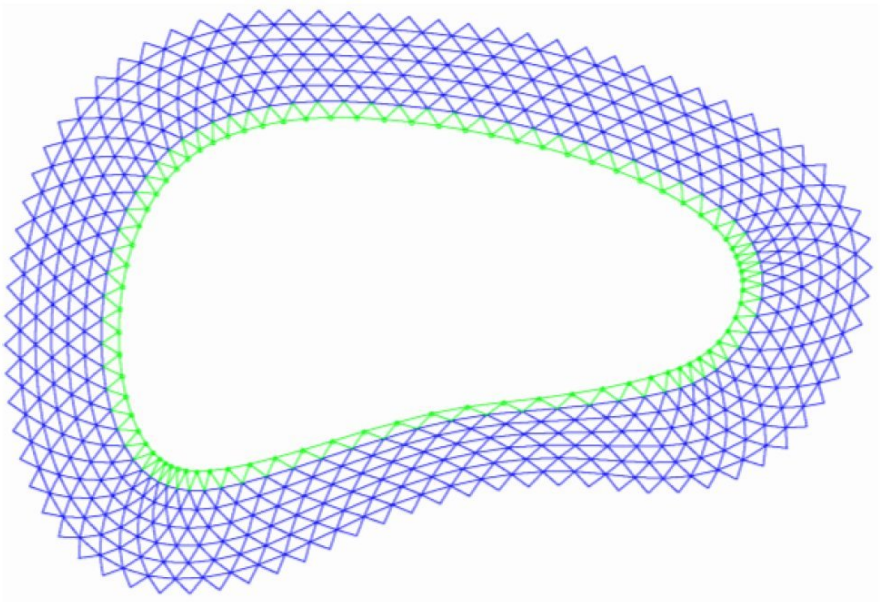


Figure 8: Initially generated nodes and bars for ring 7 of a mesh (green).

2. A starting set of Bézier curves (5, in this case) is defined, taking as vertices 5 of the initially generated npr points, as depicted in Figure 9

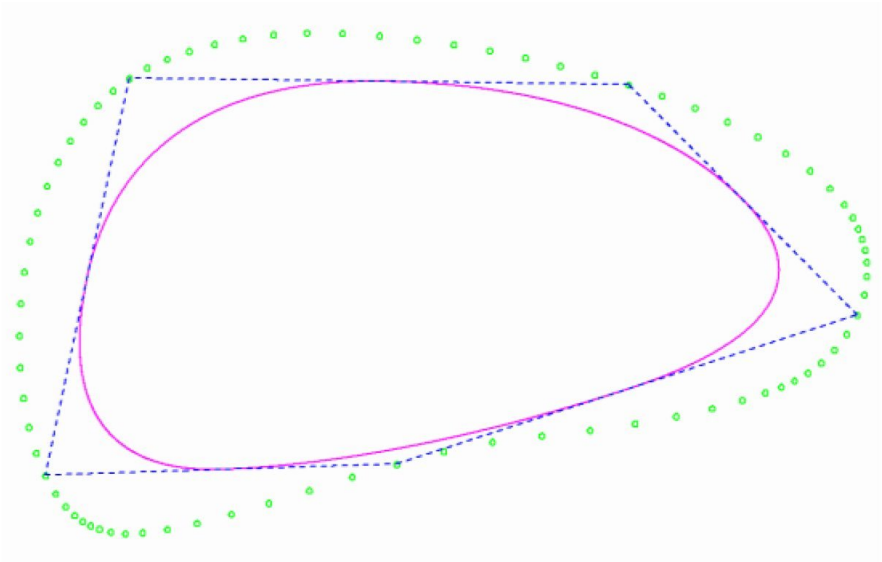


Figure 9: Starting configuration of a 5-vertex approximation set of Bézier curves.

3. An error function E is defined as the sum of the minimum distances of each of the npr points to the set of Bézier curves (Figure 10) [8]

$$E = \sum_{n=1}^{npr} d_{k_n}(x_1, x_2, \ldots, x_{nv}, y_{nv}) = \sum_{n=1}^{npr} d_{k_n}(c_1, c_2, \ldots, c_{nc}) \tag{7}$$

where $nc = 2nv$ is the number of unknowns to be iteratively approximated (10, in this case).

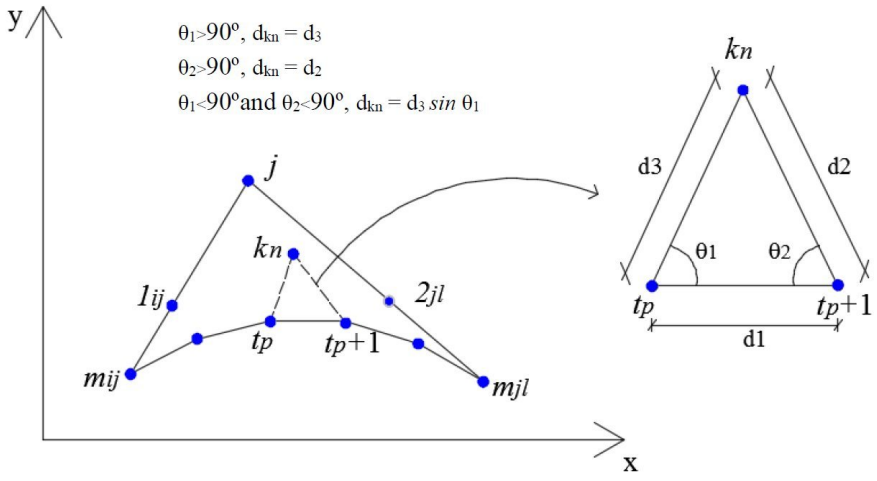


Figure 10: Minimum distance from point kn of the npr set, to the linearized Bézier curve.

4. The minimization of the error function E is performed by defining nc equations $f_j = 0$ ($j = 1, \ldots, nc$)

$$f_j = \frac{\partial E}{\partial c_j} = \sum_{n=1}^{npr} \frac{\partial d_{k_n}}{\partial c_j} = 0 \tag{8}$$

These derivatives are computed numerically. This system of equations is solved by using the Newton-Raphson algorithm (i - iteration counter)

$$\mathbf{C}_{i+1} = \mathbf{C}_i - \mathbf{J}_i^{-1}\mathbf{F}_i \quad \Longleftrightarrow \quad \mathbf{J}_i \Delta\mathbf{C} = -\mathbf{F}_i \tag{9}$$

with $\Delta\mathbf{C} = \mathbf{C}_{i+1} - \mathbf{C}_i$,

$$\mathbf{C} = \begin{Bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{nc} \end{Bmatrix} \quad , \qquad \mathbf{F} = \begin{Bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{nc} \end{Bmatrix} \tag{10}$$

and

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial c_1} = \frac{\partial^2 E}{\partial c_1^2} & \frac{\partial f_1}{\partial c_2} = \frac{\partial^2 E}{\partial c_1 c_2} & \cdots & \frac{\partial f_1}{\partial c_{nc}} = \frac{\partial^2 E}{\partial c_1 c_{nc}} \\ \frac{\partial f_2}{\partial c_1} = \frac{\partial^2 E}{\partial c_2 c_1} & \frac{\partial f_2}{\partial c_2} = \frac{\partial^2 E}{\partial c_2^2} & \cdots & \frac{\partial f_2}{\partial c_{nc}} = \frac{\partial^2 E}{\partial c_2 c_{nc}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{nc}}{\partial c_1} = \frac{\partial^2 E}{\partial c_{nc} c_1} & \frac{\partial f_{nc}}{\partial c_2} = \frac{\partial^2 E}{\partial c_{nc} c_2} & \cdots & \frac{\partial f_{nc}}{\partial c_{nc}} = \frac{\partial^2 E}{\partial c_{nc}^2} \end{bmatrix} \tag{11}$$

These derivatives are also computed numerically, using high precision arithmetic.

5. Determination of the α parameter of the set of Bézier curves - Once the Bézier curve has been defined it has to be divided into several segments of the pre-defined length C_b so as to restart the whole mesh generation process for the following rings. The last segment of the curve naturally results in a length different from C_b. The curve is thus very slightly readjusted to change its length so that the last segment also has the prescribed length, C_b. The used approach consists of an iterative computation of the α parameter of the Bézier curves in such a way that the division of the final length of the set of Bézier curves into segments of length C_b gives a whole number. The iterative process for determining the α parameter can be summarized as follows.

   - Initial estimation of α: $\alpha = 0.35$.
   - Evaluation of the length of the last bar, $llb(\alpha)$, and definition of function $g(\alpha) = llb(\alpha) - C_b$.
   - Solution of equation $g(\alpha) = 0$ via Newton-Raphson iteration: $\alpha_{i+1} = \alpha_i - g/\left(\frac{\mathrm{d}g}{\mathrm{d}\alpha}\right)_i$; the derivative is numerically computed.

   Figure 11 shows the set of Bézier curves after convergence of both iterative procedures.

6. Connection bars between the previous ring and the new ring - The nodal points are numbered from i_{initial} to i_{final} in the previous ring and from j_{initial} to j_{final} to in the new ring, where j_{initial} is the point in the new ring that is closest to point i_{initial} of the previous ring. A triangulated connection between the two rings is generated as follows (Figure 12)

   (a) First bar connects i_{initial} to j_{initial}.

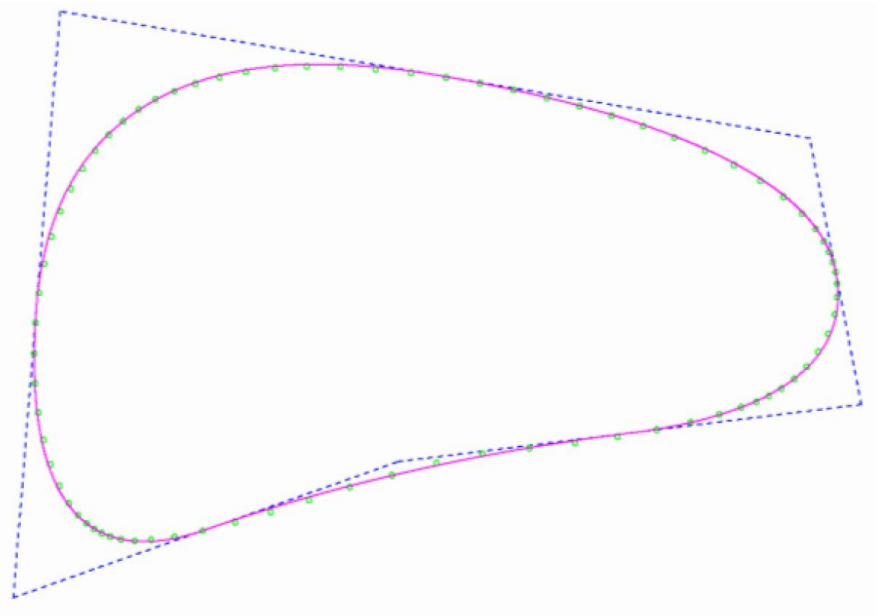


Figure 11: Set of Bézier curves after convergence of the iterative processes.

(b) For points i, from $i_{\text{initial}+1}$ to i_{final}:

- identify j_p as the closest nodal point to i in the new ring;
- connection bars between i and the points between the last connected point in the new ring, j_l and j_p; if $i = i_{\text{final}}$ and $j_p < j_{\text{final}}$, bars connecting i_{final} with j_{p+1} to j_{final} are added;
- the last bar connects j_{final} to i_{initial}

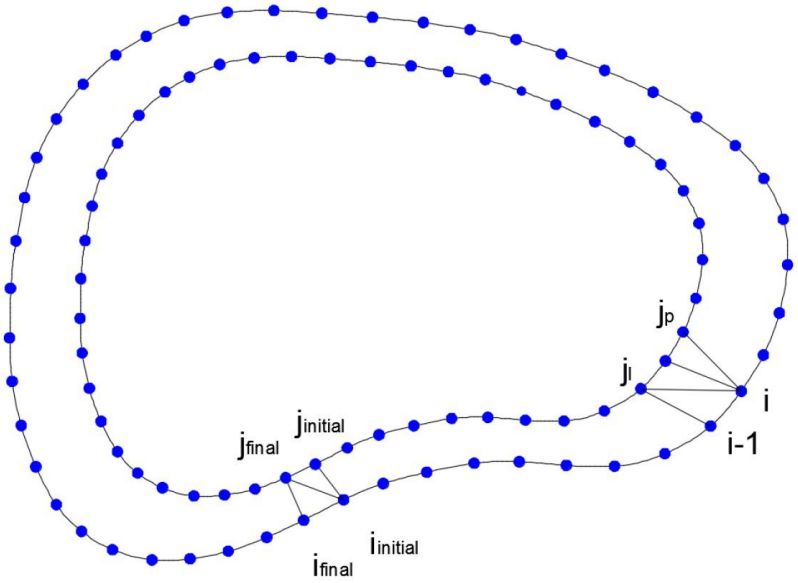


Figure 12: Connection bars between rings.

Figure 13 shows the new ring obtained using the above described procedure.

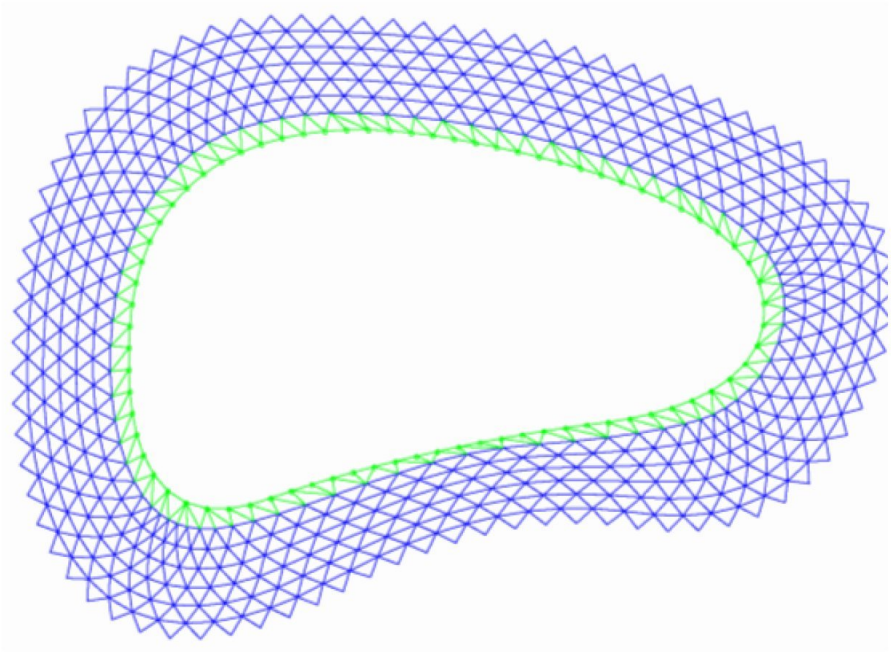


Figure 13: Final configuration of the nodes and bars in ring 7 (green).

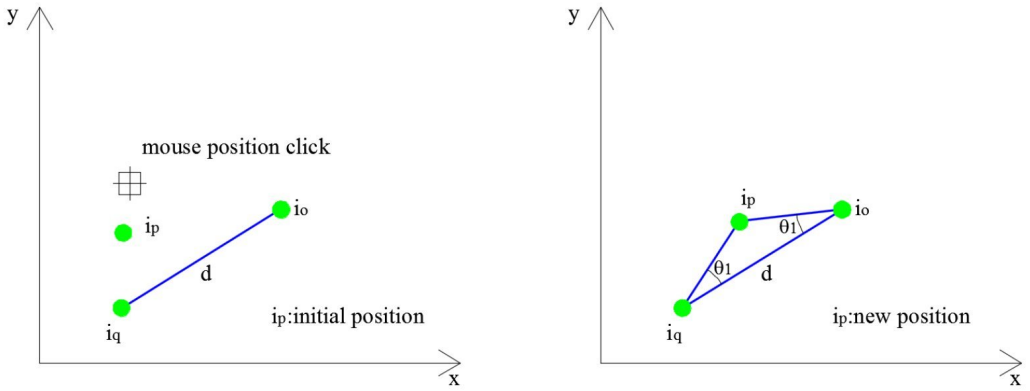
3.4 Algorithm C

As with algorithms A and B, this method, too, is based on the advancing front method, and it is therefore necessary to update the generation front after each ring is formed since this forms the basis for constructing the next i_a ring. The i_a ring is formed from i points of the i_{a-1} ring and gives k points that will be the generation front of the i_{a+1} ring. The number of k points is equal to i when the distance between two consecutive points in the new ring is greater than a given accepted value, according to the smoothing criterion established for the distance. In the absence of this condition, point C in Figure 3 is rejected and points A and B are connected to the last not-rejected point C. As a consequence, the number of k points is smaller than i. Besides this automatic smoothing process, smoothing is complemented by manually selecting the nodal points of the new ring that must be moved, using the algorithm described in Figure 14: the program automatically adjusts the point that is closest to the mouse pointer (Fig. 14-a) to a position that makes angles θ_1 with segment $i_q i_o$, which are the points before and after i_p in the new ring (Fig. 14-b).

The angle θ_1 in the examples was 5^o. Other values can be taken, depending on the type of adjustment intended. This is a simple smoothing method (no renumbering of the nodes or bars is performed) and it is quite effective at regularizing the lattice as the examples in the next chapter show.

4 EXAMPLE

This example represents an irregular boundary with distinct convexities and concavities. The maximum horizontal length is about $200\,m$ and the maximum vertical length is about $75\,m$. A value $C_b = 7\,m$ was considered. Table 1 gives the number of fixed and variable length bars



a) Manual choice of point to be moved i_p. b) Relocation of point i_p to new position.

Figure 14: Manual smoothing in Algorithm C.

obtained by the algorithms. Lattices generated with each algorithm are represented in Figures 15, 16 and 17.

	No. of fixed length bars	No. of variable length bars	Total no. of bars
Algorithm A	957 (62%)	590 (38%)	1547
Algorithm B	807 (51%)	774 (49%)	1581
Algorithm C	903 (61%)	589 (39%)	1492

Table 1: Bars of fixed and variable length for a mesh with an irregular boundary.

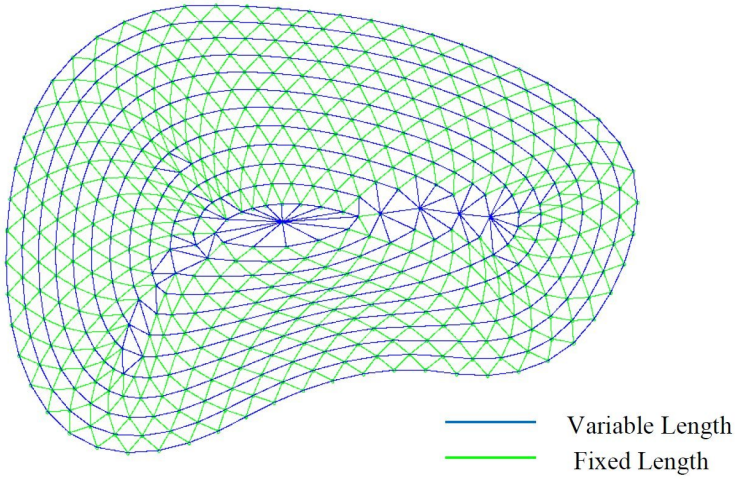


Figure 15: Lattice obtained with Algorithm A.

In this example, Algorithm B generated the mesh with the largest number of elements, the smallest number of bars of standard length and the most irregular boundary. Algorithms A and C give a larger number of bars of fixed length simultaneously with more regular meshes than Algorithm B does. Note the regularity of the Algorithm A mesh compared with that of C.

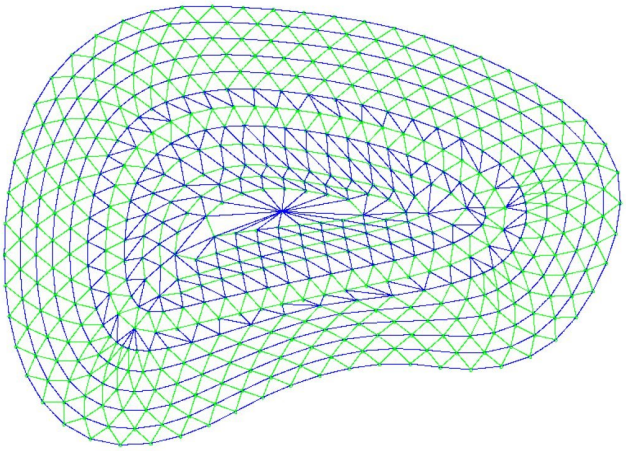


Figure 16: Lattice obtained with Algorithm B.

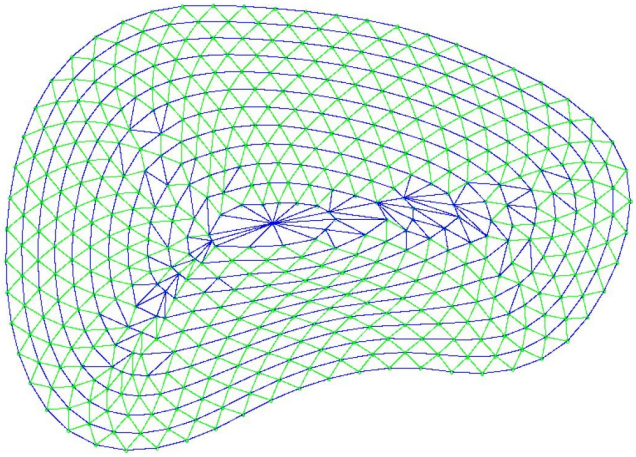


Figure 17: Lattice obtained with Algorithm c.

5 CONCLUSIONS

Three algorithms for generating triangulated meshes for curved boundaries were developed so that the meshes could be used to study the behaviour of double curvature shells, including analysis of construction processes, without shoring. The feasibility of simulating the construction stages of shells without shoring warranted the procedure used for generating triangulated meshes in successive rings. Another restriction imposed is the use of as many bars of the same length and equal to a predefined value as possible. The development of curved boundaries with successive rings of gradually smaller perimeters makes it impossible to use bars of just one length.

The three types of mesh generation can be classified according to their automation, ease of implementation and efficiency. Measures of efficiency are the number of elements, number and percentage of elements of the same length. The following conclusions can be drawn from applying the three algorithms:

Algorithm A - This is the only fully automatic algorithm. It always let to the highest percentage of bars of standard length and to the most regular meshes, that is, with triangles that were

less distorted compared with the other algorithms;

Algorithm B - Using Bézier curves for the smoothing process made this a semi-automatic algorithm: smoothing is introduced into the predefined rings by the user and the vertices of the support polygon can be used to improve the fit of the curve. However, the connection of the ring smoothed by the Bézier curve with the rings formed previously results in a less regular mesh and fewer bars of standard length. This algorithm led to the generation of meshes that were less regular and less efficient;

Algorithm C - This is a semi-automatic algorithm in which the user smoothes the mesh manually. It is easy to implement, with less supporting theory and provides efficient solutions with regular meshes. It was noted that the percentage of bars of the same length is similar to Algorithm A and that it generated the smallest number of elements. The main drawback is that relies wholly on the user and the solution cannot be replicated.

REFERENCES

[1] Sibson, R., Locally equiangular triangulations. *Computer Journal*, Vol.2, 243-245, 1979.

[2] Watson, D.F., Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes. *Computer Journal*, Vol.24, 167-172, 1981.

[3] S.H.Lo, A new mesh generation scheme for arbitrary planar domains. *Inter. J. N. Meth. Eng.*, Vol. 21, 1403-1426, 1985.

[4] Chew, L.Paul, Constrained Delaunay triangulations, *Algorithmica*, Vol.4, 97-108, 1989.

[5] S.H.Lo, Delaunay triangulation of non-convex planar domains, *Inter. J. N. Meth. Eng.*, Vol. 28, 2695-2707, 1989.

[6] Zienkieviewicz, O. Wu, J., Automatic directional refinement in adaptative analysis of compressible flows. *Inter. J. N. Meth. Eng.*, Vol. 37, 2189-2210, 1994.

[7] El-Hamalawi, A., A 2D combined advancing front-Delaunay mesh generation scheme, *Finite Elements in Analysis and Design.*, Vol. 4, 967-989, 2004.

[8] J. Argyris; I. St. Doltsinis; V. D. da Silva. Constitutive modelling and computation of non-linear viscoelastic solids. Part II: Application to orthotropic PVC-coated fabrics, *Comput. Meth. Appl. Engrg.*, Vol. 98, 159-226, 1992.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

WAVE TRANSMISSION THROUGH NONLINEAR IMPACTING METAMATERIAL UNIT

Arnab Banerjee[1], Raj Das[2], and Emilio P. Calius[3]

[1] PhD student, Mechanical Engineering Dept., University of Auckland, New Zealand
e-mail: aban991@aucklanduni.ac.nz

[2] Senior Lecturer, Mechanical Engineering Dept., University of Auckland, New Zealand
email: r.das@auckland.ac.nz

[3] Senior Scientist, Callaghan Innovation, Auckland, New Zealand
email: Emilio.Calius@callaghaninnovation.govt.nz

Keywords: Nonlinear metamaterial, Mechanical metamaterial, Acoustic metamaterial, Nonlinear oscillation, attenuation bandwidth, wideband metamaterial, mass-in-mass lattice

Abstract. *Metamaterials are an increasingly well-known type of generalized composites that can exhibit unconventional behaviors and responses due to its exciting frequency dependent properties which are not commonly encountered in natural materials. Elastodynamic metamaterials based on mechanical structures have a range of potential applications of importance in sound, vibration and seismic engineering. However, the effectiveness of metamaterials is limited to a relatively narrow frequency band as they are generally based on linear resonance mechanisms. These linear metamaterials do not perform well when under the broadband excitation spectra that are common in real life applications.*

Mathematically, the 1-D metamaterial is represented by a series of periodic resonating spring and mass lattice structures. In this paper, the bandwidth of a metamaterial is examined by introducing nonlinearity in the resonating frequency. The performance of nonlinear metamaterial is well investigated in the field of electromagnetic wave propagation although the existing literature on nonlinear mechanical metamaterial is very limited and investigated only bistable and monostable type resonating metamaterial. Till date, any research on the impacting resonating metamaterial has not been reported, in spite of its excellent vibration insulation and resonating bandwidth enhancement properties. Therefore, as a first step towards exploring the impacting metamaterial, we have analytically estimated the bandwidth enhancement of a one dimensional impacting mechanical metamaterial unit over a linear metamaterial. First, the steady state response of an impacting system is computed analytically and then the transmission loss in the frequency domain is estimated using MATLAB. From the analysis, it is found that the resonating bandwidth of an impacting system is wider than that of an equivalent linear system. Therefore, it can be effectively used as a wideband mechanical filter, acoustic insulator or shield.

1 INTRODUCTION

Natural materials have their own excitation independent material properties, such as mass density, Poisson's ratio and Young's modulus. Therefore, most of the natural materials react in phase with the excitation under external simulation; whereas, metamaterial behaves in out of phase [1]. Metamaterials are a special kind of artificial composite, designed to have exotic behaviours, such as negative mass [2-16], negative Poisson's ratio [17-22] and negative Young's modulus [4], in certain range of frequencies due to the out of phase response of multiple resonating units which are present in a metamaterial. Negative mass is achieved due to the out of phase motion of resonating unit during resonance [4, 6, 10, 11], for an example, the periodic arrangement of steel balls in a silicon rubber matrix can act as a negative effective mass system [23]. In this discussion, the definition of dynamic effective mass is according to the Newton's 2nd law of motion, which is the ratio of force and acceleration, instead of the amount of matter present in a body. An acoustic metamaterial is used as a waveguide and sound insulator. The elastic metamaterial has an extensive application in the field of vibration insulation [13]. Elastic wave propagation through a medium depends on the impedance of the medium (Z_0), which is the product of the mass density (ρ) and the P-wave (v_p) and S-wave (v_s) velocity of the medium. On the other hand, the P-wave velocity depends on the Young's modulus (E) and Poisson's ratio (υ) of the medium. So, the frequency dependent negative material property changes the impedance of the medium which reflects the incident waves back. Due to the out phase response of resonating units, effective mass which is the ratio of momentum and velocity becomes negative [6, 10, 13] which refracts the propagating waves opposite to the conventional direction.

The metamaterial is used widely in the various field of application associated with the wave propagation. However, due to the dependency on the linear resonance, the attenuation bandwidth of the linear metamaterial is limited to a very narrow bandwidth. On the other hand, nonlinearity has a potential to widen the resonating bandwidth of an oscillator. Therefore the nonlinear metamaterial attracts the attention of the current researchers. The recent development in field of nonlinear electromagnetic metamaterial is condensed in this review article [24]. To widen the attenuation band of a mechanical metamaterial, Huang and Sun [5] proposed multi-resonance unit inside of the single block of the metamaterial. In contrary, an analytical study on nonlinear elastic dispersion for the elastic metamaterial was carried out by Khajehtuorian and Hossain [25]. This research concludes that the location and the length of the attenuation bandwidth of nonlinear elastic metamaterial depends on the amplitude of the wave. Nadkarni et al [26] analyzed a metamaterial containing bistable nonlinear resonating units for low and high amplitude wave motion and concluded that nonlinearity is proportional with the amplitude of wave and the nonlinearity enhanced the performance of metamaterial as a filter. Klatt and Haberman [27] proposed a multi-scale model for nonlinear metamaterial. Attenuation bandwidth of cubic [28, 29] and sinusoidal [30-32] nonlinear metamaterial is much wider than the linear metamaterial. Beside that, stress wave propagation through granular material or impacting chain of monoatomic lattice is also studied and found that it can attenuate much wider frequency band[33, 34]; thus, can be applicable in designing of helmet or protecting devices. Moreover, the resonating bandwidth of an energy harvester can be extended upto twice compared to an equivalent linear system by introducing impacting non-linearity [35, 36] although, the impacting resonating metamaterial has not been reported in any publication till now.

Impact or contact takes place when the two bodies come to very close vicinity. Starting from the time of Newton, plenty of literatures are available on the modelling of the impacting phenomena which are condensed and critically evaluated in this review article [37]. To model

the normal directional impact from a compliance based approach, nonlinear Hertzian element or Kelvin element can be used, where a small penetration is assumed for a finite duration. In the contrary, stereo-mechanical approach models the impact as an instantaneous event; thus, the velocity response becomes discontinuous. Modelling of this discontinuous response for a steady state vibration adds extra difficulty. A straight forward solution with standard numerical integration solver can solve the transient state response very easily but cannot converge to a steady state solution. Several techniques on the computation of the steady state response of an impacting device are proposed by various researchers[38]. In this paper, a solver designed by Nigm and Shabana [39] is used with stereo-mechanical unilateral impact model to solve the steady state response of the impacting metamaterial unit.

2 COMPUTATION OF STEADY STATE RESPONSE

The simplest resonating metamaterial can be mathematically conceptualized as a chain of mass-in-mass units. A single unit of the resonating metamaterial is depicted in Figure 1a) where the internal mass can vibrate without any obstruction; whereas, in Figure 1b) the motion of the internal mass is bounded by two stoppers. The impact with the stopper and the inner mass induced a nonlinear response to the system.

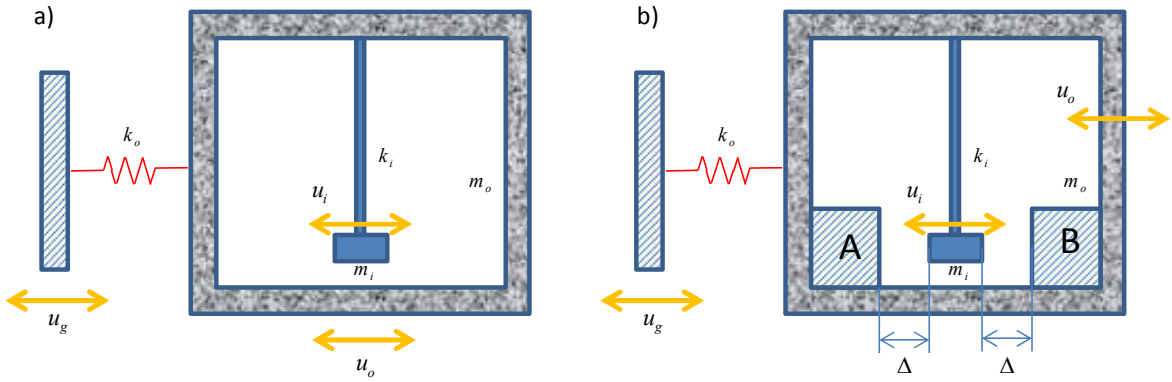


Figure 1: A single unit of metamaterial a) linear b) piece wise linear or impacting

The main aim of this paper is to compare the attenuation bandwidth for the linear and impacting metamaterial, subjected to a monochromatic displacement excitation from the base, as shown in Figure 1.

2.1 Linear system

The equation of motion of the linear system which is shown in Figure 1a) can be written as:

$$\underbrace{\begin{bmatrix} m_o & 0 \\ 0 & m_i \end{bmatrix}}_{m} \underbrace{\begin{Bmatrix} \ddot{u}_o \\ \ddot{u}_i \end{Bmatrix}}_{\ddot{x}} + \underbrace{\begin{bmatrix} k_o + k_i & -k_i \\ -k_i & k_i \end{bmatrix}}_{k} \underbrace{\begin{Bmatrix} u_o \\ u_i \end{Bmatrix}}_{x} = \underbrace{\begin{Bmatrix} k_o u_g \cos \bar{\omega} t \\ 0 \end{Bmatrix}}_{f} \tag{1}$$

where, $[m],[k]$ are respectively mass and stiffness matrix of the system. Therefore, the natural frequencies of the two degree of freedom system are:

$$\omega_{1,2} = \frac{k_i}{2m_i} + \left(\frac{k_o + k_i}{2m_o}\right) \pm \sqrt{-\frac{k_i}{m_i}\frac{k_o}{m_o} + \frac{1}{4}\left(\frac{k_o + k_i}{m_0} + \frac{k_i}{m_i}\right)^2} \tag{2}$$

and the modal matrix, which contains the mode shapes for each natural frequency can be expressed as:

$$\tilde{\Phi} = \begin{bmatrix} a_1 & a_2 \\ 1 & 1 \end{bmatrix}; \quad a_n = -\frac{m_i}{k_i}\omega_n^2 + 1 \tag{3}$$

Now, the Eq. (1) is a coupled second order differential equation. To uncouple this Eq.(1), the system $\{x\}$ is transferred to the principal coordinate $\{q\}$, by coordinate transformation.

$$\underbrace{\begin{Bmatrix} u_o \\ u_i \end{Bmatrix}}_{x} = \underbrace{\begin{bmatrix} \dfrac{a_1}{m_i a_1^2 + m_o} & \dfrac{a_2}{m_i a_2^2 + m_o} \\ \dfrac{1}{m_i a_1^2 + m_o} & \dfrac{1}{m_i a_2^2 + m_o} \end{bmatrix}}_{\Phi} \underbrace{\begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix}}_{q} \tag{4}$$

where, Φ is the mass normalized modal matrix. After shifting the coordinate system to principal coordinate, the mass, stiffness and force matrix changes to:

$$M = \Phi^T m \Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; K = \Phi^T k \Phi = \begin{bmatrix} \omega_1^2 & 0 \\ 0 & \omega_2^2 \end{bmatrix}; P = \Phi^T f \tag{5}$$

Therefore, Eq.(1) can be solved in different modes as a single degree of freedom system and the equation of motion of the metamaterial unit in the principal coordinate system is:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{Bmatrix} + \begin{bmatrix} \omega_1^2 & 0 \\ 0 & \omega_2^2 \end{bmatrix} \begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} = \begin{Bmatrix} P_1 \\ P_2 \end{Bmatrix} \cos \bar{\omega} t$$
$$\rightarrow \ddot{q}_i + \omega_i^2 q_i = P_i \cos \bar{\omega} t \quad i = 1,2 \tag{6}$$

The solution of the i[th] mode Eq.(6) is:

$$q_i = \sin(\omega_i t) A + \cos(\omega_i t) B + Z_i \cos \bar{\omega} t \tag{7}$$

where, A and B are two unknowns based on the initial condition and $Z_i = P_i / (\omega_i^2 - \bar{\omega}^2)$.

2.2 An example of linear system

Now, as an example, a linear system, having the inner mass is 1kg, outer mass is 2 kg, inner stiffness 3N/m and outer stiffness is 6N/m, subjected to a base excitation, having amplitude of 1m displacement and frequency 10 Hz = 0.628 rad/s is solved. The natural frequencies are $(\omega_1^2, \omega_2^2) = (1.5, 6)$ and the reduced force vector is $(P_1, P_2) = (2.4495, 3.4641)$. Now, to estimate the values of the unknowns, A and B, initial condition need to be substituted. At $t = 0$, the displacement and velocity of inner and outer mass are $(x_1, x_2) = (0,0)$ and $(\dot{x}_1, \dot{x}_2) = (0,0)$, which implies that $(q_1, q_2, \dot{q}_1, \dot{q}_2) = (0,0,0,0)$; thus solution of Eq.(7) is $q_i = Z_i \cos \bar{\omega} t$. The displacement of inner and outer mass with its both side, assuming sides are of 0.5 m apart, is plotted in Figure 2.

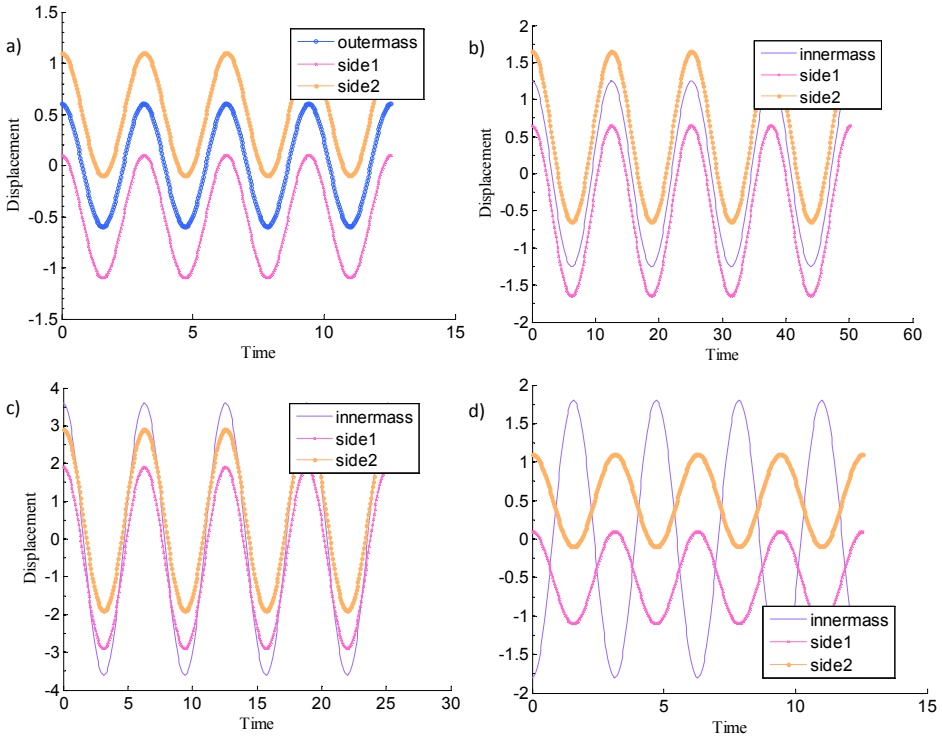


Figure 2: Steady-state displacement response of inner and outer mass showing the side1 and side2 for various frequency b) 0.5 rad/s, c) 1.0 rad/s, d) 2.0 rad/s

Figure 2 depicts that at low frequency, for example $\bar{\omega} = 0.5$ rad/s, inner mass vibrates in-phase with the outer block and it does not touch the either side of the outer block. On the other hand, inner mass touches the side1 and side2 of the outer mass although both the responses are in same phase. A further increment of excitation frequency makes the vibration of inner mass out of phase with the outer mass and therefore it touches both the side of outer mass more frequently. In both the cases, when the displacement response of inner mass crosses the boundary of the outer block then impact occurs and it needs a special analysis.

2.3 Impacting system

From the discussion of the linear system, it can be easily found that the inner mass can comes in contact several times with the two side walls of the outer mass not only during the out-of-phase response but also for the in phase response. Therefore, impacting system demands a thorough investigation on the responses of the system. In this paper, the frequency response of an impacting system is computed by mostly extending the method described by Nigm and Shabana [39].

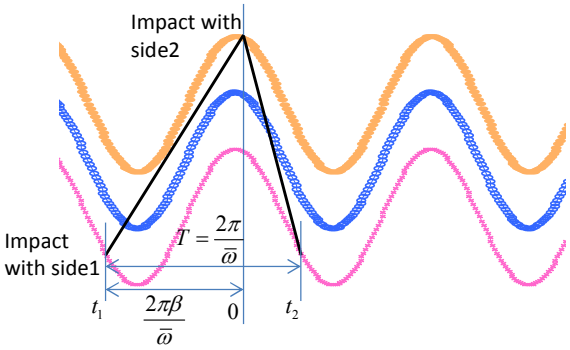


Figure 3: conceptualization of the impacting events

Now, as the vibration of the blocks is steady state in nature, after a constant interval, time period $T = 2\pi/\bar{\omega}$, inner mass hits the side1 of the outer block. It can be assumed that inner mass hits the side 1 of outer block at $t=t_1$ and $t=t_2$. Assuming that only double impact takes place in each period, the solution of the Eq. (6) can be written as:

$$\begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} = \underbrace{\begin{bmatrix} \sin\omega_1 t & 0 \\ 0 & \sin\omega_2 t \end{bmatrix}}_{s} \underbrace{\begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix}}_{A} + \underbrace{\begin{bmatrix} \cos\omega_1 t & 0 \\ 0 & \cos\omega_2 t \end{bmatrix}}_{c} \underbrace{\begin{Bmatrix} B_1 \\ B_2 \end{Bmatrix}}_{B} + \begin{Bmatrix} Z_1 \cos(\bar{\omega}t+\alpha) \\ Z_2 \cos(\bar{\omega}t+\alpha) \end{Bmatrix}, \quad t_1 \le t \le 0^-$$

$$\begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} = \underbrace{\begin{bmatrix} \sin\omega_1 t & 0 \\ 0 & \sin\omega_2 t \end{bmatrix}}_{s} \underbrace{\begin{Bmatrix} C_1 \\ C_2 \end{Bmatrix}}_{C} + \underbrace{\begin{bmatrix} \cos\omega_1 t & 0 \\ 0 & \cos\omega_2 t \end{bmatrix}}_{c} \underbrace{\begin{Bmatrix} D_1 \\ D_2 \end{Bmatrix}}_{D} + \begin{Bmatrix} Z_1 \cos(\bar{\omega}t+\alpha) \\ Z_2 \cos(\bar{\omega}t+\alpha) \end{Bmatrix}, \quad 0^+ \le t \le t_2 \tag{8}$$

where, 0^+ and 0^- are represent the condition just after and just before t=0. The velocities are:

$$\begin{Bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{Bmatrix} = \underbrace{\begin{bmatrix} \omega_1\cos\omega_1 t & 0 \\ 0 & \omega_2\cos\omega_2 t \end{bmatrix}}_{\omega c} \underbrace{\begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix}}_{A} - \underbrace{\begin{bmatrix} \omega_1\sin\omega_1 t & 0 \\ 0 & \omega_2\sin\omega_2 t \end{bmatrix}}_{\omega s} \underbrace{\begin{Bmatrix} B_1 \\ B_2 \end{Bmatrix}}_{B} - \begin{Bmatrix} Z_1\bar{\omega}\sin(\bar{\omega}t+\alpha) \\ Z_2\bar{\omega}\sin(\bar{\omega}t+\alpha) \end{Bmatrix}, t_1 \le t \le 0^-$$

$$\begin{Bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{Bmatrix} = \underbrace{\begin{bmatrix} \omega_1\cos\omega_1 t & 0 \\ 0 & \omega_2\cos\omega_2 t \end{bmatrix}}_{\omega c} \underbrace{\begin{Bmatrix} C_1 \\ C_2 \end{Bmatrix}}_{C} - \underbrace{\begin{bmatrix} \omega_1\sin\omega_1 t & 0 \\ 0 & \omega_2\sin\omega_2 t \end{bmatrix}}_{\omega s} \underbrace{\begin{Bmatrix} D_1 \\ D_2 \end{Bmatrix}}_{D} - \begin{Bmatrix} Z_1\bar{\omega}\sin(\bar{\omega}t+\alpha) \\ Z_2\bar{\omega}\sin(\bar{\omega}t+\alpha) \end{Bmatrix}, 0^+ \le t \le t_2 \tag{9}$$

where, A, B, A_1, B_1 are unknown initial condition and $Z_n = \dfrac{P_n}{\left(\omega_n^2 - \bar{\omega}^2\right)}$.

Due to the periodicity,

$$x\left(0^-\right) = x\left(0^+\right); \quad x\left(t_2\right) = x\left(t_1\right) \tag{10}$$

After converting Eq.(10)a) to the principal coordinate, we can get:

$$\begin{aligned} x\left(0^-\right) &= \Phi q = \Phi\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{Bmatrix} B_1 \\ B_2 \end{Bmatrix} + \begin{Bmatrix} Z_1\cos\alpha \\ Z_2\cos\alpha \end{Bmatrix}\right) \\ x\left(0^+\right) &= \Phi q = \Phi\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{Bmatrix} D_1 \\ D_2 \end{Bmatrix} + \begin{Bmatrix} Z_1\cos\alpha \\ Z_2\cos\alpha \end{Bmatrix}\right) \\ \rightarrow B_1 &= D_1; \quad B_2 = D_2 \end{aligned} \tag{11}$$

The converted equation of the Eq.(10) b) in principal coordinate is:

$$
\begin{aligned}
x(t_2) &= \Phi q = \Phi\left(s_2 C + c_2 B + \begin{Bmatrix} Z_1 \cos\bar{\omega}t_2 \\ Z_2 \cos\bar{\omega}t_2 \end{Bmatrix} \cos\alpha - \begin{Bmatrix} Z_1 \sin\bar{\omega}t_2 \\ Z_2 \sin\bar{\omega}t_2 \end{Bmatrix} \sin\alpha \right) \\
x(t_1) &= \Phi q = \Phi\left(s_1 A + c_1 B + \begin{Bmatrix} Z_1 \cos\bar{\omega}t_1 \\ Z_2 \cos\bar{\omega}t_1 \end{Bmatrix} \cos\alpha - \begin{Bmatrix} Z_1 \sin\bar{\omega}t_1 \\ Z_2 \sin\bar{\omega}t_1 \end{Bmatrix} \sin\alpha \right) \\
\rightarrow & s_2 C - s_1 A + (c_2 - c_1) B + \underbrace{\begin{Bmatrix} Z_1(\cos\bar{\omega}t_2 - \cos\bar{\omega}t_1) \\ Z_2(\cos\bar{\omega}t_2 - \cos\bar{\omega}t_1) \end{Bmatrix}}_{Z\bar{c}} \cos\alpha - \underbrace{\begin{Bmatrix} Z_1(\sin\bar{\omega}t_2 - \sin\bar{\omega}t_1) \\ Z_2(\sin\bar{\omega}t_2 - \sin\bar{\omega}t_1) \end{Bmatrix}}_{Z\bar{s}} \sin\alpha = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}
\end{aligned} \tag{12}
$$

Assuming the impact as an instantaneous event, the change of velocity can be estimated by:

$$
\left\{\dot{x}(0^+)\right\} = G\left\{\dot{x}(0^-)\right\}; \left\{\dot{x}(t_2)\right\} = G\left\{\dot{x}(t_1)\right\}; G = \frac{1}{m_i + m_o}\begin{bmatrix} m_o - m_i\varepsilon & m_i(1+\varepsilon) \\ m_o(1+\varepsilon) & m_i - m_o\varepsilon \end{bmatrix} \tag{13}
$$

where, ε is the coefficient of restitution between the inner and the outer mass. Eq.(13) a) can be converted in principal coordinate system as:

$$
\begin{aligned}
\dot{x}(0^+) &= \Phi\dot{q}(0^+) = \Phi\left(\underbrace{\begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}}_{\omega} \underbrace{\begin{Bmatrix} C_1 \\ C_2 \end{Bmatrix}}_{C} - \underbrace{\begin{Bmatrix} Z_1\bar{\omega} \\ Z_2\bar{\omega} \end{Bmatrix}}_{Z\bar{\omega}} \sin\alpha \right) \\
\dot{x}(0^-) &= \Phi\dot{q}(0^-) = \Phi\left(\underbrace{\begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}}_{\omega} \underbrace{\begin{Bmatrix} A_1 \\ A_2 \end{Bmatrix}}_{A} - \begin{Bmatrix} Z_1\bar{\omega} \\ Z_2\bar{\omega} \end{Bmatrix} \sin\alpha \right) \\
\rightarrow & -G\Phi\omega A + \Phi\omega C - (\Phi Z\bar{\omega} - G\Phi Z\bar{\omega})\sin\alpha = \{0\}_{2\times 1}
\end{aligned} \tag{14}
$$

Similarly, Eq.(13) b) can be converted in principal coordinate as:

$$
\begin{aligned}
& \dot{x}(t_2) = G\dot{x}(t_1) \\
& \Phi\left(\omega c_2 C - \omega s_2 B - \begin{Bmatrix} Z_1\bar{\omega}\sin\bar{\omega}t_2 \\ Z_2\bar{\omega}\sin\bar{\omega}t_2 \end{Bmatrix} \cos\alpha - \begin{Bmatrix} Z_1\bar{\omega}\cos\bar{\omega}t_2 \\ Z_2\bar{\omega}\cos\bar{\omega}t_2 \end{Bmatrix} \sin\alpha \right) \\
& = G\Phi\left(\omega c_1 A - \omega s_1 B - \begin{Bmatrix} Z_1\bar{\omega}\sin\bar{\omega}t_1 \\ Z_2\bar{\omega}\sin\bar{\omega}t_1 \end{Bmatrix} \cos\alpha - \begin{Bmatrix} Z_1\bar{\omega}\cos\bar{\omega}t_1 \\ Z_2\bar{\omega}\cos\bar{\omega}t_1 \end{Bmatrix} \sin\alpha \right) \\
& \rightarrow -G\Phi\omega c_1 A + (G\Phi\omega s_1 - \Phi\omega s_2) B + \Phi\omega c_2 C + \left(G\Phi \begin{Bmatrix} Z_1\bar{\omega}\sin\bar{\omega}t_1 \\ Z_2\bar{\omega}\sin\bar{\omega}t_1 \end{Bmatrix} - \Phi \begin{Bmatrix} Z_1\bar{\omega}\sin\bar{\omega}t_2 \\ Z_2\bar{\omega}\sin\bar{\omega}t_2 \end{Bmatrix} \right) \cos\alpha \\
& + \left(G\Phi \begin{Bmatrix} Z_1\bar{\omega}\cos\bar{\omega}t_1 \\ Z_2\bar{\omega}\cos\bar{\omega}t_1 \end{Bmatrix} - \Phi \begin{Bmatrix} Z_1\bar{\omega}\cos\bar{\omega}t_2 \\ Z_2\bar{\omega}\cos\bar{\omega}t_2 \end{Bmatrix} \right) \sin\alpha = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}
\end{aligned} \tag{15}
$$

The impact occurs at the side 1 at t=0 and at side 2 at t=t_1 and t_2; therefore,

$$
u_i(0^-) - u_o(0^-) = \Delta; \quad u_i(t_1) - u_o(t_1) = -\Delta \tag{16}
$$

The transformation from Eq.(16)a) to principal coordinate system can be expressed as:

$$\Phi_2\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{Bmatrix}B_1\\ B_2\end{Bmatrix}+\begin{Bmatrix}Z_1\\ Z_2\end{Bmatrix}\cos\alpha\right)-\Phi_1\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{Bmatrix}B_1\\ B_2\end{Bmatrix}+\begin{Bmatrix}Z_1\\ Z_2\end{Bmatrix}\cos\alpha\right)=\Delta \tag{17}$$

The converted equation of Eq.(16)b) in principal coordinate is:

$$\begin{aligned}&\Phi_2\left(s_1A+c_1B+\begin{Bmatrix}Z_1\cos\bar{\omega}t_1\\ Z_2\cos\bar{\omega}t_1\end{Bmatrix}\cos\alpha-\begin{Bmatrix}Z_1\sin\bar{\omega}t_1\\ Z_2\sin\bar{\omega}t_1\end{Bmatrix}\sin\alpha\right)\\ &-\Phi_1\left(s_1A+c_1B+\begin{Bmatrix}Z_1\cos\bar{\omega}t_1\\ Z_2\cos\bar{\omega}t_1\end{Bmatrix}\cos\alpha-\begin{Bmatrix}Z_1\sin\bar{\omega}t_1\\ Z_2\sin\bar{\omega}t_1\end{Bmatrix}\sin\alpha\right)=-\Delta\end{aligned} \tag{18}$$

where, Φ_i is the i[th] row of the modal matrix Φ.

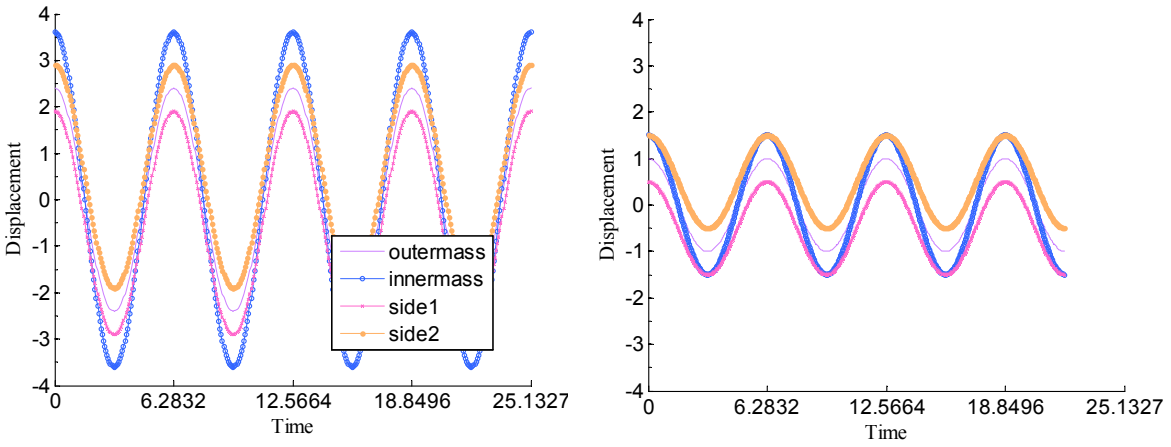
The unknown values of initial conditions, namely A, B, A_1, B_1, and phase angle α can be estimated by writing Eq.(12),(14),(15),(17) and (18) in a matrix form:

$$\begin{bmatrix}-s_1 & c_2-c_1 & s_2 & -Zs & Zc\\ -G\Phi\omega & 0 & \Phi\omega & T_{24} & 0\\ -G\Phi\omega c_1 & T_{32} & \Phi\omega c_2 & T_{34} & T_{35}\\ 0 & \Phi_2 I-\Phi_1 I & 0 & 0 & \Phi_2 Z-\Phi_1 Z\\ \Phi_2 s_1-\Phi_1 s_1 & \Phi_2 c_1-\Phi_1 c_1 & 0 & T_{54} & T_{55}\end{bmatrix}\begin{Bmatrix}A_{2\times1}\\ B_{2\times1}\\ C_{2\times1}\\ \sin\alpha\\ \cos\alpha\end{Bmatrix}=\begin{Bmatrix}\{0\}_{2\times1}\\ \{0\}_{2\times1}\\ \{0\}_{2\times1}\\ \Delta\\ -\Delta\end{Bmatrix} \tag{19}$$

where, $T_{24}=G\Phi Z\bar{\omega}-\Phi Z\bar{\omega}$, $T_{32}=G\Phi\omega s_1-\Phi\omega s_2$, $T_{35}=G\Phi Z\bar{\omega}\sin\bar{\omega}t_1-\Phi Z\bar{\omega}\sin\bar{\omega}t_2$, $T_{34}=G\Phi Z\bar{\omega}\cos\bar{\omega}t_1-\Phi Z\bar{\omega}\cos\bar{\omega}t_2$, $T_{54}=\left(\Phi_1Z-\Phi_2Z\right)\sin\bar{\omega}t_1$, $T_{55}=\left(\Phi_1Z-\Phi_2Z\right)\cos\bar{\omega}t_1$

2.4 Extension of the previous example with impact modification

Nigm and Shabana [39] shows that solution becomes non-viable for $0\le\beta\le0.18$. In this case a symmetric impact can be considered because the gaps on both the sides are same; therefore, $\beta=0.5$ is assumed to conduct the analysis. In case of the previous example, solved in section 2.2, shows that at excitation frequency 1 rad/s and 2 rad/s, response of inner mass crosses the boundary of outer mass, which is impractical. This issue is resolved and shown in . after consideration of impact.



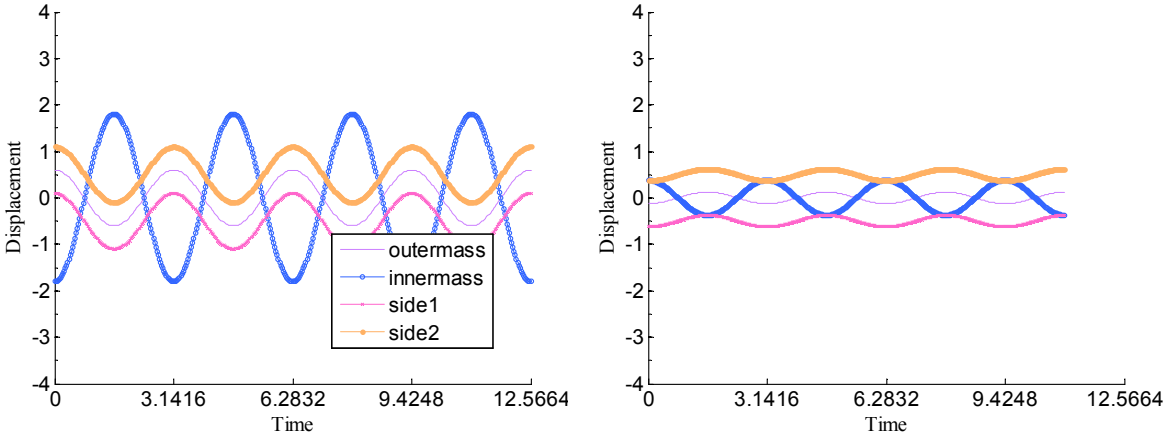


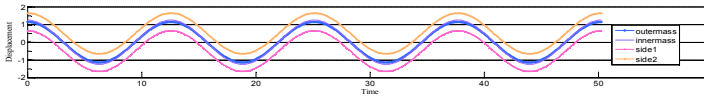
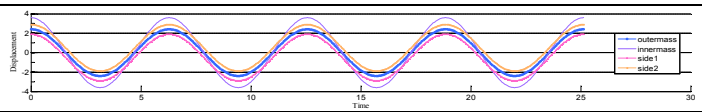
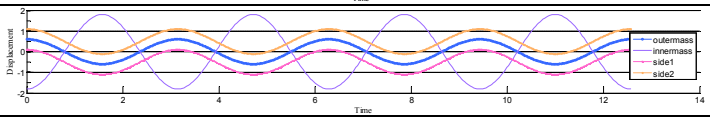
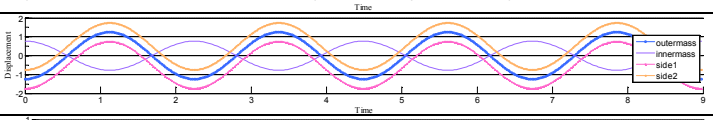
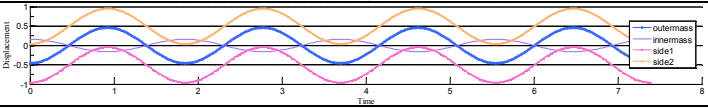
Figure 4: Time domain steady-state response of the linear system without considering impact (in left column) and then modified impacting solution (right column) for excitation frequency 1 rad/s and 2 rad/s

Figure 4 depicts the linear response of the two degree of freedom systems where the vibration of inner mass exceeds the boundary of the outer mass. In modified steady state solution the vibration of the inner mass is limited between the two side wall of the outer mass. The interaction between the side wall and the inner mass damped the vibration of the outer mass, even for the fully elastic collision; therefore, vibro-impacting system is potentially very useful for vibration attenuation.

3 FREQUENCY-RESPONSE ANALYSIS

Frequency-response curve of impacting system is computed to determine the attenuation bandwidth of a single cell of an impacting metamaterial due to the impact between the walls of the outer mass and the inner resonating unit. The total frequency dependent steady-state response can be classified into two main parts: in-phase motion, where the inner and outer masses are in same phase and out-of-phase motion, where both of them are in opposite phase. Different state of vibration for various excitation frequencies is summarized in Table 1.

Table 1: Different state of response for the system given in section 2.2

Phase	Condition	Plot (linear system)	Remarks
In-phase	$X_i < X_o + \Delta$		No impact
	$X_i > X_o + \Delta$		Impact
Out of phase	$X_i > X_o + \Delta$		Impact
	$X_i < X_o - \Delta$		Impact
	$X_i > \Delta - X_o$		Impact

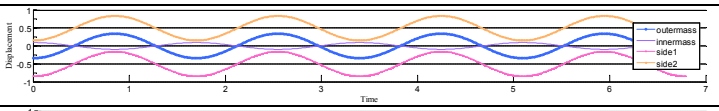
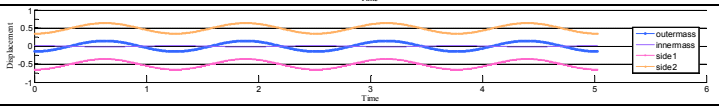
	$X_i < \lvert \Delta - X_o \rvert$		No impact
	$X_i < X_o$		No impact

Table 1 classifies the conditions for in-phase and out-of-phase impacts. These conditions are employed in computing the frequency response curve. Simple linear solution is adopted and maximum of the displacement is considered as amplitude in those frequency ranges, where no impact takes place, otherwise modified solution procedure discussed in section 2.3 is adopted. In Table 1, X_i and X_o are stands for the amplitude of the inner mass and outer mass. The transmittance of vibration, which is described as $T(dB) = 20\log(u_0/u_g)$, is plotted in the frequency domain in Figure 5.

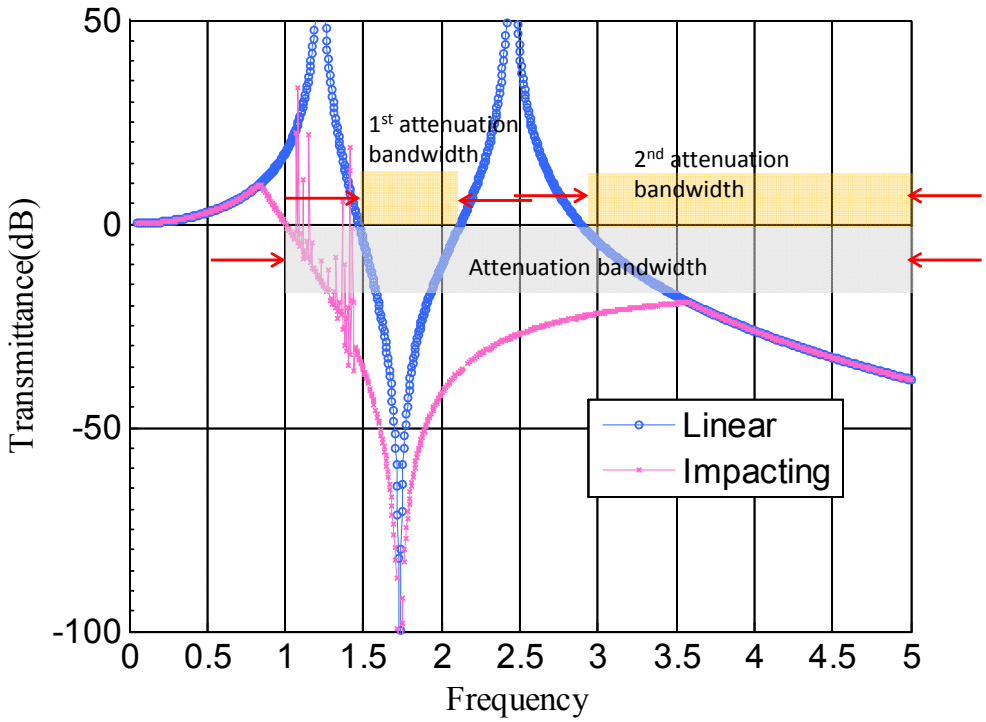


Figure 5: Attenuation bandwidth for linear and impacting metamaterial unit

Figure 5 depicts that the transmission of excitation is very high at the two resonating peaks, which are at 1.25 and 2.45 rad/s. In between these two peaks, an anti-resonating valley exists at 1.72 rad/s, which symbolized that almost no vibration is transmitted to the outer block from the base, in this frequency range. This anti-resonating valley is generally termed as attenuation bandwidth because it does not allow waves to propagate within the lattice in this frequency range. After the 2nd resonating peak, another infinitely long attenuation band generates. In case of the linear metamaterial unit, the 1st attenuation bandwidth starts at 1.5 rad/s and extended upto 2.2 rad/s and 2nd attenuation bandwidth originates from 2.9 rad/s and extended upto infinity. On the other hand, the attenuation bandwidth of the impacting metamaterial is initiate from 1.0 rad/s and extended upto infinity without having any transmission band inside.

Therefore, as the low frequency wave insulator and wideband mechanical filter, this impacting metamaterial can be used very efficiently.

4 CONCLUSIONS

The attenuation bandwidth increment capability of a single unit of impacting metamaterial over a single unit of equivalent linear metamaterial is analytically evaluated in this paper. The amplitude of the steady state response of impacting metamaterial unit for a specific monochromatic excitation is computed based on the method suggested by Nigm and Shabana. From the transmittance plot it can be easily conclude that impacting metamaterial widen the attenuation bandwidth in lower and higher frequency side, which proves that the impacting metamaterial can potentially be used as wideband filter and low frequency sound insulator. Further parametric study on the effect of the coefficient of restitution and the gap distance on the attenuation bandwidth is need to be investigated towards the designing of a wideband metamaterial unit.

REFERENCES

[1] Banerjee, B., *An Introduction to Metamaterials and Waves in Composites*. 2011, Auckland: Taylor & Fransis.

[2] Pai, P.F. and H. Peng. *Acoustic metamaterial structures based on multi-frequency vibration absorbers*. in *Proceedings of SPIE - The International Society for Optical Engineering*. 2014.

[3] Sun, H., et al. *Acoustic metamaterial with negative parameter*. in *Proceedings of SPIE - The International Society for Optical Engineering*. 2014.

[4] Huang, H.H. and C.T. Sun, *Anomalous wave propagation in a one- dimensional acoustic metamaterial having simultaneously negative mass density and Young's modulus.* The Journal of the Acoustical Society of America, 2012. **132**: p. 2887.

[5] Huang, G.L. and C.T. Sun, *Band Gaps in a Multiresonator Acoustic Metamaterial.* Journal of Vibration and Acoustics, 2010. **132**(3): p. 031003-031003.

[6] Yao, S., X. Zhou, and G. Hu, *Experimental study on negative effective mass in a 1D mass–spring system.* New Journal of Physics, 2008. **10**(4): p. 043020.

[7] Sheng, P., et al., *Locally resonant sonic materials.* Physica B: Condensed Matter, 2003. **338**(1–4): p. 201-205.

[8] Sun, H., X. Du, and P. Frank Pai. *Metamaterial broadband vibration absorbers by local resonance*. in *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. 2011.

[9] Pope, S.A. and H. Laalej, *A multi-layer active elastic metamaterial with tuneable and simultaneously negative mass and stiffness.* Smart Materials and Structures, 2014. **23**(7).

[10] Calius, E.P., et al., *Negative mass sound shielding structures: Early results.* Physica Status Solidi B-Basic Solid State Physics, 2009. **246**(9): p. 2089-2097.

[11] Huang, H.H., C.T. Sun, and G.L. Huang, *On the negative effective mass density in acoustic metamaterials.* International Journal of Engineering Science, 2009. **47**(4): p. 610-617.

[12] Lu, M.-H., L. Feng, and Y.-F. Chen, *Phononic crystals and acoustic metamaterials.* Materials Today, 2009. **12**(12): p. 34-42.

[13] Lin, G.C., et al., *Simulation of a metamaterial beam for mechanical wave absorption.* Wuli Xuebao/Acta Physica Sinica, 2011. **60**(3).

[14] Sun, H., et al. *Theory and experiment research of metamaterial beams for broadband vibration absorption*. in *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. 2011.

[15] Sun, H., et al. *Theory and experiment research of metamaterial panel for mechanical waves absorption*. in *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. 2013.

[16] Huang, H.H. and C.T. Sun, *Wave attenuation mechanism in an acoustic metamaterial with negative effective mass density.* New Journal of Physics, 2009. **11**(1): p. 013003.

[17] Baughman, R.H., et al., *Negative Poisson's ratios as a common feature of cubic metals.* Nature, 1998. **392**(6674): p. 362-365.

[18] Friis, E., R. Lakes, and J. Park, *Negative Poisson's ratio polymeric and metallic foams.* Journal of Materials Science, 1988. **23**(12): p. 4406-4414.

[19] Kocer, C., D.R. McKenzie, and M.M. Bilek, *Elastic properties of a material composed of alternating layers of negative and positive Poisson's ratio.* Materials Science and Engineering: A, 2009. **505**(1–2): p. 111-115.

[20] Lakes, R., *Foam structures with a negative Poisson's ratio.* Science, 1987. **235**(4792): p. 1038-1040.

[21] Lakes, R., *Advances in negative Poisson's ratio materials.* Advanced Materials, 1993. **5**(4): p. 293-296.

[22] Larsen, U.D., O. Sigmund, and S. Bouwstra. *Design and fabrication of compliant micromechanisms and structures with negative Poisson's ratio*. in *Micro Electro Mechanical Systems, 1996, MEMS'96, Proceedings. An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems. IEEE, The Ninth Annual International Workshop on*. 1996. IEEE.

[23] Calius, E.P., et al., *Negative mass sound shielding structures: Early results.* physica status solidi (b), 2009. **246**(9): p. 2089-2097.

[24] Lapine, M., I.V. Shadrivov, and Y.S. Kivshar, *Colloquium: Nonlinear metamaterials.* Reviews of Modern Physics, 2014. **86**(3): p. 1093.

[25] Khajehtourian, R. and M.I. Hussein, *Dispersion characteristics of a nonlinear elastic metamaterial.* AIP Advances, 2014. **4**(12): p. 124308.

[26] Nadkarni, N., C. Daraio, and D.M. Kochmann, *Dynamics of periodic mechanical structures containing bistable elastic elements: From elastic to solitary wave propagation.* Physical Review E, 2014. **90**(2): p. 023204.

[27] Klatt, T. and M.R. Haberman, *A nonlinear negative stiffness metamaterial unit cell and small-on-large multiscale material model.* Journal of Applied Physics, 2013. **114**(3).

[28] Romeo, F. and G. Rega, *Wave propagation properties in oscillatory chains with cubic nonlinearities via nonlinear map approach.* Chaos, Solitons & Fractals, 2006. **27**(3): p. 606-617.

[29] Lazarov, B.S. and J.S. Jensen, *Low-frequency band gaps in chains with attached non-linear oscillators.* International Journal of Non-Linear Mechanics, 2007. **42**(10): p. 1186-1193.

[30] Georgiou, I., *On the Global Geometric Structure of the Dynamics of the Elastic Pendulum.* Nonlinear Dynamics, 1999. **18**(1): p. 51-68.

[31] Chen, W.Z., B.B. Hu, and H. Zhang, *Interactions between impurities and nonlinear waves in a driven nonlinear pendulum chain.* Physical Review B, 2002. **65**(13): p. 134302.
[32] Hodges, C.H. and J. Woodhouse, *Vibration isolation from irregularity in a nearly periodic structure: Theory and measurements.* The Journal of the Acoustical Society of America, 1983. **74**: p. 894.
[33] Bonanomi, L., G. Theocharis, and C. Daraio, *Wave propagation in granular chains with local resonances.* Physical Review E, 2015. **91**(3): p. 033208.
[34] Lydon, J., G. Theocharis, and C. Daraio, *Nonlinear resonances and energy transfer in finite granular chains.* Physical Review E, 2015. **91**(2): p. 023208.
[35] Vandewater, L. and S. Moss, *Probability-of-existence of vibro-impact regimes in a nonlinear vibration energy harvester.* Smart Materials and Structures, 2013. **22**(9): p. 094025.
[36] Soliman, M.S.M., et al., *A wideband vibration- based energy harvester.* Journal of Micromechanics and Microengineering, 2008. **18**(11).
[37] Banerjee, A., A. Chanda, and R. Das, *Historical Origin and Recent Development on Normal Directional Impact Models for Rigid Body Contact Simulation: A Critical Review.* Archives of Computational Methods in Engineering, 2016: DOI: 10.1007/s11831-016-9164-5
[38] Hsu, C.S., *Impulsive Parametric Excitation: Theory.* Journal of Applied Mechanics, 1972. **39**(2): p. 551-558.
[39] Nigm, M.M. and A.A. Shabana, *Effect of an impact damper on a multi-degree of freedom system.* Journal of Sound and Vibration, 1983. **89**(4): p. 541-557.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A PARTICLE FINITE ELEMENT METHOD FOR MACHINING SIMULATIONS

Rodríguez J.M.[1], **Jonsén P.** [2], **and Svoboda A.**[2]

[1] Division of Mechanics of Solid Materials
Department of Engineering Sciences and Mathematics. Lulea University of Technology
e-mail: rodjua@ltu.se

[2] Division of Mechanics of Solid Materials
Department of Engineering Sciences and Mathematics. Lulea University of Technology

e-mail: par.jonsen@ltu.se and ales.svoboda@ltu.se

Keywords: Updated Lagrangian formulation – Particle Finite element method – Metal cutting - Mixed Stabilized method.

Abstract. *Metal cutting process is a nonlinear dynamic problem that includes geometrical, material, and contact nonlinearities. In this work a Lagrangian finite element approach for simulation of metal cutting processes is presented, based on the so-called Particle Finite Element Method (PFEM). The governing equations for the deformable bodies are discretized with the FEM via a mixed formulation using simplicial elements with equal linear interpolation for displacements, pressure and temperature. The use of PFEM for modeling of metal cutting processes includes the use of a remeshing process, α -shape concepts for detecting domain boundaries, contact mechanics laws and material constitutive models. The merits of the formulation are demonstrated in the solution of 2D and 3D thermally-coupled metal cutting processes using the particle finite element method. The method shows good results and is a promising method for future simulations of thermally/coupled machining processes.*

1 INTRODUCTION

Numerical simulation of machining had a tremendous development in the last twenty years [1-5]. Thanks to the increasing computer power, it can provide excellent results with reasonable computational times. Thanks to its flexibility and robustness commercial FEM codes are standard tools in the industry.

However, the main limitation of the FEM in this kind of application is that the quality of the results depends on the mesh. In the problems characterized by large deformations, such as the machining processes, if a Lagrangian formulation is used, the mesh moves with material and the elements become so distorted that numerical results lose their validity, unless a remeshing-rezoning technique is used[1, 2]. This improves robustness but, at the same time, the computational time increases and supplementary computational errors are introduced, in the form of numerical, artificial, diffusion of the results.

Metal cutting has been simulated also with Eulerian formulations [6] and arbitrary Lagrangian-Eulerian (ALE) [7-9], but for all cases additional drawbacks appear.

A possible alternative is the use of meshless methods. Actually, different mesh-free techniques have been developed in the last years [5, 10, 11] but in many cases the improvement of the results quality, with the same number of degrees of freedom, is vanished by the very high computational time of shape functions.

In this paper an innovative approach named Particle Finite Element Method (PFEM) has been used. PFEM overcomes the main difficulties in the numerical simulation of metal cutting, specifically, large strains and angular distortions, multiple contacts and self-contact, generation of new boundaries and fracture with multiple cracks and defragmentation.

The outline of this work is as follows. In the following section, the solid-dynamics equations in a Lagrangian framework are introduced, a space-time finite element semi-discretization is presented and the stabilization of the incompressibility condition is described. In section 3 the basic ideas of the PFEM are illustrated. In Section 4 the constitutive model used is explained, and finally a variety of machining and cutting problems are solved in Section 5 to show the efficiency of the proposed method.

2 GOVERNING EQUATIONS

Let $\mathbf{x}$ denote the current position of a solid particle in the current configuration at time t, $\mathbf{u} = \mathbf{u}(\mathbf{x},t)$ its displacement, $\theta = \theta(\mathbf{x},t)$ its temperature, $p = p(\mathbf{x},t)$ its pressure and $\sigma = \sigma(\mathbf{x},t)$ the Cauchy stress tensor at time $t \in [0,T]$. For a solid, momentum, mass conservation and energy write as

$$\begin{aligned} &\nabla\bullet\sigma + \rho\mathbf{b} = 0 \quad in \quad \Omega \\ &\frac{1}{\kappa}p - \ln(J) = 0 \quad in \quad \Omega \\ &c\frac{D\theta}{Dt} - \nabla\bullet(k\nabla\theta) + Q = 0 \quad in \quad \Omega \end{aligned} \tag{1}$$

where Ω is the volume occupied by the solid in the current configuration, ρ is the density, c the specific heat, k the thermal conductivity, J is the determinant of the deformation gradient and $\mathbf{b}$ the external body forces. The initial conditions are prescribed assigning the initial displacement $\mathbf{u}_0(\mathbf{x})$ and initial velocity $\mathbf{v}_0(\mathbf{x})$ at $t = 0$ whereas both Dirichlet and Neumann type boundary conditions are imposed as:

$$
\begin{aligned}
&\mathbf{u}(\mathbf{x},t) = \bar{\mathbf{u}}(\mathbf{x},t) \quad \forall \mathbf{x} \in \Gamma_D \\
&\mathbf{v}(\mathbf{x},t) = \bar{\mathbf{v}}(\mathbf{x},t) \quad \forall \mathbf{x} \in \Gamma_D \\
&\theta(\mathbf{x},t) = \bar{\mathbf{T}}(\mathbf{x},t) \quad \forall \mathbf{x} \in \Gamma_{D\theta} \\
&\sigma(\mathbf{x},t)\bullet\mathbf{n} = \bar{\mathbf{h}}(\mathbf{x},t) \quad \forall \mathbf{x} \in \Gamma_N \\
&-k\nabla\theta\bullet\mathbf{n} = \bar{\mathbf{q}}(\mathbf{x},t) \quad \forall \mathbf{x} \in \Gamma_{N\theta}
\end{aligned}
\tag{2}
$$

where $\bar{\mathbf{u}}(\mathbf{x},t)$, $\bar{\mathbf{v}}(\mathbf{x},t)$, $\bar{\mathbf{T}}(\mathbf{x},t)$, $\bar{\mathbf{h}}(\mathbf{x},t)$ and $\bar{\mathbf{q}}(\mathbf{x},t)$ are assigned functions, $\mathbf{n}$ denotes the outward normal to the boundary and $\Gamma_D \cup \Gamma_N = \partial\Omega$ and $\Gamma_{DT} \cup \Gamma_{NT} = \partial\Omega$

2.1 Space discretization

Introducing a Galerkin isoparametric finite element discretization, the semi-discrete form of equations (1)

$$
\begin{aligned}
&\mathbf{F}_{\mathbf{u},\text{int}}(\mathbf{u},p) - \mathbf{F}_{ext} = 0 \\
&\mathbf{F}_{\mathbf{p},press}(p) - \mathbf{F}_{p,vol}(\mathbf{u}) + \mathbf{F}_{p,stab}(p) = 0 \\
&\mathbf{F}_{\theta,dyn}(\dot{\theta}) - \mathbf{F}_{\theta,\text{int}}(\theta) + \mathbf{F}_{\theta,ext}(\theta) = 0
\end{aligned}
\tag{3}
$$

Where

$$
\begin{aligned}
&\mathbf{F}_{\mathbf{u},\text{int}}(\mathbf{u},p) = \int_V \mathbf{B}_u^T \sigma dV \\
&\mathbf{F}_{\mathbf{u},ext} = \int_V \mathbf{N}^T \mathbf{b} dV - \int_{\Gamma_N} \mathbf{N}^T \bar{\mathbf{h}} d\Gamma \\
&\mathbf{F}_{p,press}(p) = \int_V \frac{1}{\kappa} \mathbf{N}\mathbf{N}^T p dV \\
&\mathbf{F}_{p,vol}(\mathbf{u}) = \int_V \mathbf{N}^T \ln(J) dV \\
&\mathbf{F}_{p,stab}(p) = \int_{V^{(e)}} \frac{\alpha}{\mu} p^{(e)} \quad \mathbf{N}^{(e)}\mathbf{N}^{(e)T} - \tilde{\mathbf{N}}^{(e)}\tilde{\mathbf{N}}^{(e)T} \quad dV \\
&\mathbf{F}_{\theta,\text{int}}(\theta) = \int_V \mathbf{B}_\theta^T \mathbf{q} dV - \int_V \mathbf{N}^T D_{\text{int}} dV \\
&\mathbf{F}_{\theta,ext} = \int_{\Gamma_{NT}} \mathbf{N}^T (\mathbf{q}.\mathbf{n}) d\Gamma \\
&\mathbf{F}_{\theta,dyn}(\dot{\theta}) = \int_V c\mathbf{N}\mathbf{N}^T \dot{\theta} dV
\end{aligned}
\tag{4}
$$

In the above expressions, $\mathbf{N}$ are the shape functions, $\tilde{\mathbf{N}}^{(e)}$ the constant shape function through the element, $\mathbf{B}_\mathbf{u}$ and $\mathbf{B}_\theta$ are the strain displacement matrix and the gradient-temperature matrix, respectively. Those matrices contain the derivatives of the shape functions used in the interpolation of the problem variables.

In finite element computations, the above force vectors are obtained as the assemblies of element vectors. Given a nodal point, each component of the global force associated with a particular global node is obtained as the sum of the corresponding contributions from the element force vectors of all elements that share the node. In this work, the element force vectors are evaluated using Gaussian quadrature.

2.2 Pressure stabilization

The selected discretization (linear elements for displacement and pressure) is known to lead to spurious oscillations in the pressure field and a stabilization procedure is required. For the present purpose the stabilized formulation called the Polynomial Pressure Projection (PPP) presented and applied to the Stokes equations in [12] is used. The PPP introduces a pressure term in the mass conservation equation $\mathbf{F}_{p,stab}$ shown in Eq. (3) and (4). Unlike other stabilization methods, the Polynomial pressure projection (PPP) does not require specification of a mesh dependent stabilization parameter or calculation of higher-order derivatives. The PPP uses a projection on a discontinuous space and as a consequence can be implemented in an elementary level surpassing the need of mesh dependent and problem dependent parameters.

2.3 Time integration

A staggered approach[13] is adopted for the purpose of coupling the mechanical and thermal equations. Mechanical and thermal computations are staggered assuming constant temperature during the mechanical step and constant heat generation during the thermal step. A mechanical step is taken first on the current distribution of temperatures and the heat generated is computed as

$$D_{\text{int}} = \chi \sigma_y \dot{\xi} \tag{5}$$

Where χ is the Taylor-Quinney coefficient, σ_y the yield stress and $\dot{\xi}$ the equivalent strain rate. A Newton-Raphson algorithm solves the momentum equation and mass conservation for nodal displacement and pressures. The heat computed in the mechanical problem is transferred to the thermal problem and the balance of energy is discretized using an implicit Backward-Euler algorithm. The solution of discretized energy equation is carried out using a Newton-Raphson algorithm.

3 DESCRIPTION OF THE METHOD

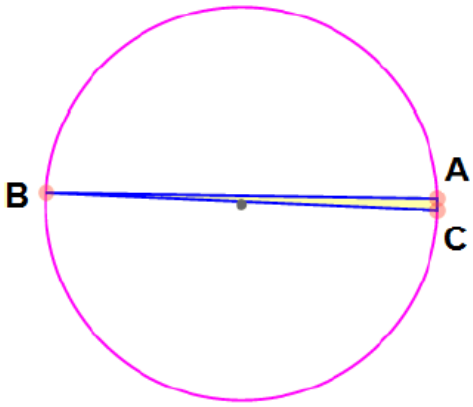
The PFEM is a set of numerical strategies combined for the solution of large deformation problems. The PFEM, in the form implemented in the present work, consists of the following conceptual steps

1. Definition of the domain(s) in the last converged configuration, $t = {}^{n}t$, keeping existing spatial discretization.
2. Transference of variables by a smoothing process – from Gauss points to nodes.
3. Discretization of the given domain(s) in a set of particles of infinitesimal size – elimination of existing connectivity.
4. Reconstruction of the mesh through a triangulation of the domain's convex-hull and the definition of the boundary applying the α-shape method [14], defining a new spatial discretization .
5. A contact method to recognize the multibody interaction.
6. Transference of information, interpolating nodal variables into the Gauss points.
7. Solution of the system of equations for ${}^{n+1}t = {}^{n}t + \Delta t$.
8. Go back to step 1 and repeat the solution process for the next time step.

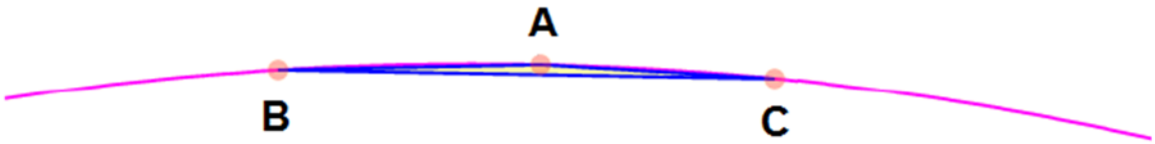
3.1 Mesh generation and identification of external boundaries

In a Lagrangian framework the external boundary and the reference volume are defined by the position of the material particles. Every time the mesh is regenerated, the particles belonging

to the boundary may change and the new boundary nodes (and therefore the particles) have to be identified. The Delaunay triangulation generates the convex hull of the set of particles. Moreover, the convex hull may not be conformal with the external boundaries. A possibility to overcome this problem is to correct the generated mesh using the so-called **α**-shape method[14] to remove the unnecessary triangles from the mesh using a criterion based on the mesh distortion. The **α**-shape method can also be used for the identification of the solid particles which separate from the rest of the domain.



(a) Sharp triangle



(b)Flat triangle

Figure 1. Triangles with the same mean size and their circumcircles.

In the numerical simulation of metal cutting process, despite the continuous Delaunay triangulation, elements arise with unacceptable aspect ratios. Neither the Alpha Shape method nor the Delaunay triangulation can prevent from the creation of highly distorted elements. Let consider, for example, the triangle depicted in Figure 1(a). The circumradius associated to this type of triangle is not excessively large. The Alpha Shape method and the Delaunay triangulation can only ensure the best physical mesh (so the best discretization that respects the physical boundaries of the domain) for a given cloud of points. However, during the motion, the nodes distribution may be such that it is impossible to avoid the presence of poor quality elements. In order to avoid this eventuality the remeshing with the Delaunay tessellation and the Alpha Shape method must be combined with some additional controls.

For example, for avoiding the presence of sharp elements in the triangulation, as the one

shown in Figure 1(b), one should monitor the edge lengths of each triangle. Whenever the distance between two nodes is smaller than a prearranged critical length, one of those nodes is removed.

Another type of critical elements are the slivers, so the simplices with null area or volume. These may form when there are three points contained in the same line, or four points laying in the same 3D plane as shown in Figure 2.

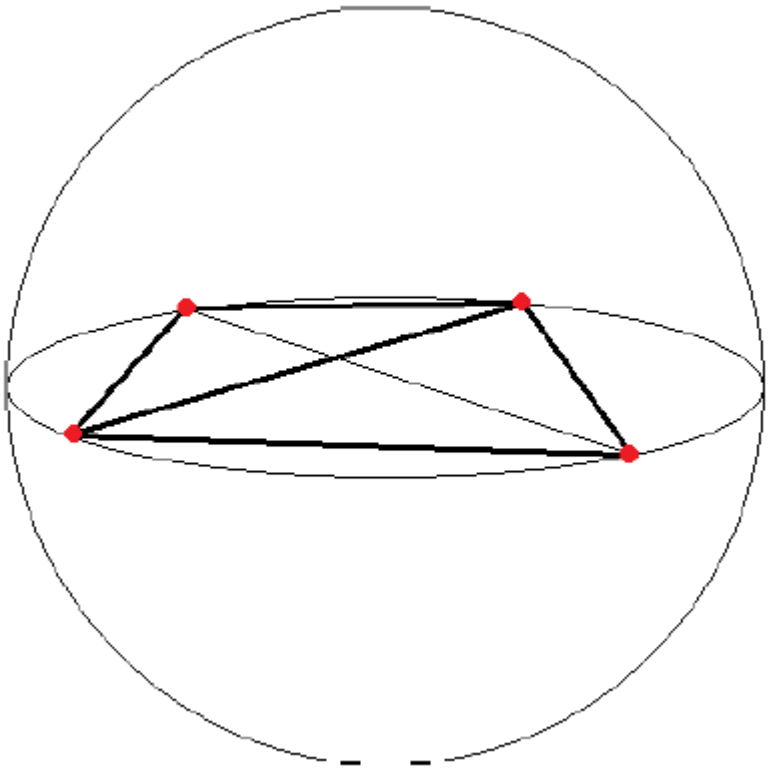


Figure 2. Sliver element in 3D.

In theory, the Alpha Shape method should not allow the formation of those elements because the associated circumradius is huge, see Figure 1(b). However, these elements are critical from the numerical point of view and they can escape to the Alpha Shape control. For all these reasons, the formation of sliver elements must be prevented. In this work, the formation of slivers is prevented by checking the element areas. Whenever the element surface or volume is less than a threshold value, computed for example with respect to the mean area of the mesh, one of the nodes of those elements is removed. Note that this control is also beneficial versus the formation of sharp elements.

After these operations over the nodes, the discretization needs to be updated. It is done in this work modifying locally the mesh and the Alpha Shape control. Once the updated mesh is created, the nodal variables of the new particles are assigned, via interpolation, with the values of the neighbor nodes.

3.2 Mesh refinement

The addition of particles is the principal tool, which we employ for sidestepping the difficulties, associated with the deformation-induced element distortion, and for resolving the different scales of the solution. The insertion of nodes is based on the equidistributiom of plastic

power, such that elements exceeding a prescribed tolerance TOL are targeted for refinement [13, 15].

4 THE CONSTITUTIVE MODEL

A summary of the J_2-plastic flow model with isotropic hardening is shown in Table 1. Table 2 shows the main steps involved in the radial return mapping algorithm and Table 3 shows the expression of the consistent tangent elastoplastic moduli considering non-linear isotropic hardening law known as Johnson-Cook.[16-19]

(i) Multiplicative decomposition of the deformation gradient

$$\mathbf{F} = \mathbf{F}^e\mathbf{F}^p$$

(ii) Free energy

$$\psi(J,\bar{\mathbf{b}}^e,\xi) \coloneqq U(J) + \bar{W}(\bar{\mathbf{b}}^e) + \mathrm{H}(\xi)$$

$$U(j) = \frac{1}{2}\kappa\log^2(J), \quad J = \det\left[\mathbf{F}\right] = \det\left[\mathbf{F}^e\right]$$

$$\bar{W}(\bar{\mathbf{b}}^e) = \frac{1}{2}\mu(tr\left[\bar{\mathbf{b}}^e\right] - 3), \bar{\mathbf{b}}^e = J^{-\frac{2}{3}}\mathbf{b}^e, \mathbf{b}^e = \mathbf{F}^e\mathbf{F}^{eT}$$

(iii) Mixed Hyperelastic and Hardening response

$$\tau(\mathbf{u},p) = p\mathbf{1} + \mathbf{s}(\mathbf{u}), p = JU'(J), \mathbf{s}(\mathbf{u}) = \mu dev(\bar{\mathbf{b}}^e)$$

$$\sigma_y = (A + B\xi^n)(1 + \log\left(\frac{\dot{\xi}}{\dot{\xi}_{ref}}\right))\left(1 - \left(\frac{T - T_o}{T_{melt} - T_o}\right)^m\right)$$

(iv) Von-Mises yield condition

$$\phi(\mathbf{s},\xi) \coloneqq \left\|\mathbf{s}\right\| - \sqrt{\frac{2}{3}}\sigma_y(\xi) \le 0$$

(v) Associative flow rule

$$L_v\bar{\mathbf{b}}^e = \gamma tr\left[\bar{\mathbf{b}}^e\right]\mathbf{n}, \dot{\xi} = \gamma\sqrt{\frac{2}{3}}, \mathbf{n} = \frac{\mathbf{s}}{\left\|\mathbf{s}\right\|}$$

(vi) Kuhn-Tucker loading/unloading conditions

$$\gamma \ge 0, \quad \phi \le 0, \quad \gamma\phi = 0$$

(vii) Consistency condition

$$\gamma\dot{\phi} = 0 \ \text{ if } \ \phi{=}0$$

Table 1. Hyperelastic J_2-flow model at finite strain

Step 1: Database and initial data $\mu, \mathbf{F}_{n+1}, \bar{\mathbf{b}}^e_n, \xi_n, p_{n+1}$

Step 2: Compute the elastic predictor

$$\bar{\mathbf{b}}^{e,trial}_{n+1} = \bar{\mathbf{F}}_{n+1}\bar{\mathbf{b}}^e_n\bar{\mathbf{F}}^T_{n+1}, \bar{\mathbf{F}}_{n+1} = \det(\bar{\mathbf{F}}_{n+1})^{-\frac{1}{3}}\bar{\mathbf{F}}_{n+1}$$

$$\xi^{trial}_n = \xi_n$$

$$\mathbf{s}^{trial}_{n+1} = \mu dev\left[\bar{\mathbf{b}}^{e,trial}_{n+1}\right], \sigma^{trial}_y = \sigma_y(\xi^{trial}_n)$$

$$f^{trial}_{n+1} = \left\|\mathbf{s}^{trial}_{n+1}\right\| - \sqrt{\frac{2}{3}}\sigma^{trial}_y$$

Step 3: Check for plastic loading

IF : $f_{n+1}^{trial} \leq 0$ THEN

Set $(.)_{n+1} = (.)_{+1}^{trial}$ and EXIT

END IF

Step 4: Radial return mapping

Set: $\bar{\mu}_{n+1} = \mu \bar{I}^{e}_{n+1}, \bar{I}^{e}_{n+1} = \frac{1}{3} tr\left[\bar{\mathbf{b}}^{e,trial}_{n+1} \right]$

Solve $\left\| \mathbf{s}^{trial}_{n+1} \right\| - 2\bar{\mu}_{n+1}\gamma_{n+1} - \sqrt{\frac{2}{3}}\sigma_y = 0$ for γ_{n+1}

$\mathbf{n}_{n+1} := \frac{\mathbf{s}^{trial}_{n+1}}{\left\| \mathbf{s}^{trial}_{n+1} \right\|}$

$\mathbf{s}_{n+1} = \mathbf{s}^{trial}_{n+1} - 2\bar{\mu}_{n+1}\gamma_{n+1}\mathbf{n}_{n+1}, \xi_{n+1} = \xi^{trial}_{n+1} + \sqrt{\frac{2}{3}}\gamma_{n+1}$

Step 5: Addition of the elastic pressure

$\tau_{n+1} = p_{n+1}\mathbf{1} + \mathbf{s}_{n+1}$

Table 2. Hyperelastic J_2-flow model at finite deformations: radial return mapping algorithm

$\mathbf{c}^{dev}_{n+1} = (1-\beta_1)\mathbf{c}^{dev,trial}_{n+1} - 2\bar{\mu}_{n+1}(\beta_3 \mathbf{n}_{n+1} \otimes \mathbf{n}_{n+1} + \beta_4(\mathbf{n}_{n+1} \otimes dev\left[\mathbf{n}^2_{n+1} \right]))$

$\mathbf{c}^{dev}_{n+1} = 2\bar{\mu}_{n+1}(\underline{\mathbf{1}} - \frac{1}{3}\mathbf{1} \otimes \mathbf{1}) - \frac{2}{3}\left\| \mathbf{s}^{trial}_{n+1} \right\| (\mathbf{n}_{n+1} \otimes \mathbf{1}_{n+1} + \mathbf{1}_{n+1} \otimes \mathbf{n}_{n+1})$

$\beta_0 = 1 + \frac{\partial\sigma_y / \partial\xi_{n+1}}{3\bar{\mu}}, \beta_1 = 2\bar{\mu}_{n+1}\frac{\gamma_{n+1}}{\left\| \mathbf{s}^{trial}_{n+1} \right\|}, \beta_2 = (1 - \frac{1}{\beta_0})\frac{2}{3}\frac{\left\| \mathbf{s}^{trial}_{n+1} \right\|}{\bar{\mu}}\gamma_{n+1}$

$\beta_3 = 1/\beta_0 - \beta_1 + \beta_2, \beta_4 = (1/\beta_0 - \beta_1)\frac{\left\| \mathbf{s}^{trial}_{n+1} \right\|}{\bar{\mu}}$

Table 3. Hyperelastic J_2-flow model at finite deformations: deviatoric consistent elastoplastic moduli

In Table 1 $\psi(J, \bar{\mathbf{b}}^e, \xi)$ represents the uncoupled volumetric/isochoric stored-energy function, $U(J)$ the volumetric part of the stored energy function, $\bar{W}(\bar{\mathbf{b}}^e)$ the isochoric part of the energy function defined in terms of the elastic left Cauchy-Green tensor $\bar{\mathbf{b}}^e$, $\tau(\mathbf{u}, p)$ the Kirchhoff stress tensor, p the Kirchhoff pressure, $\mathbf{s}(\mathbf{u})$ the deviatoric component of the Kirchhoff stress tensor. In Table 2 and Table 3 $(.)_n$ and $(.)_{n+1}$ denote the time discrete values of the variable $(.)$ at times t_n and t_{n+1}, respectively.

5 APPLICATIONS/NUMERICAL EXAMPLES

The objective of this section is to show the ability of the PFEM to simulate metal cutting problems in both 2D and 3D, where very large strains (around 10) are present.

5.1 Orthogonal cutting test: 2D approach

Our first case is an industrial application consisting of a rigid tool that is cutting a rectangular block of steel. The block has a length of 3.7 mm, a width of 1.8mm and a depth of 1 mm. The

cutting has an imposed velocity of 3333.3 mm/s, a cutting depth of 0.60 mm, a rake angle of 0º, a clearance angle of 0º and a tool radius of 0.4mm. The rigid tool is composed by to straight lines connected by a circular arch on the tool tip with the characteristics of the cutting parameters described, see Figure 3. The workpiece material behavior is given by a Johnson-Cook constitutive law that takes into account strain hardening, strain rate hardening and thermal softening Table 4.

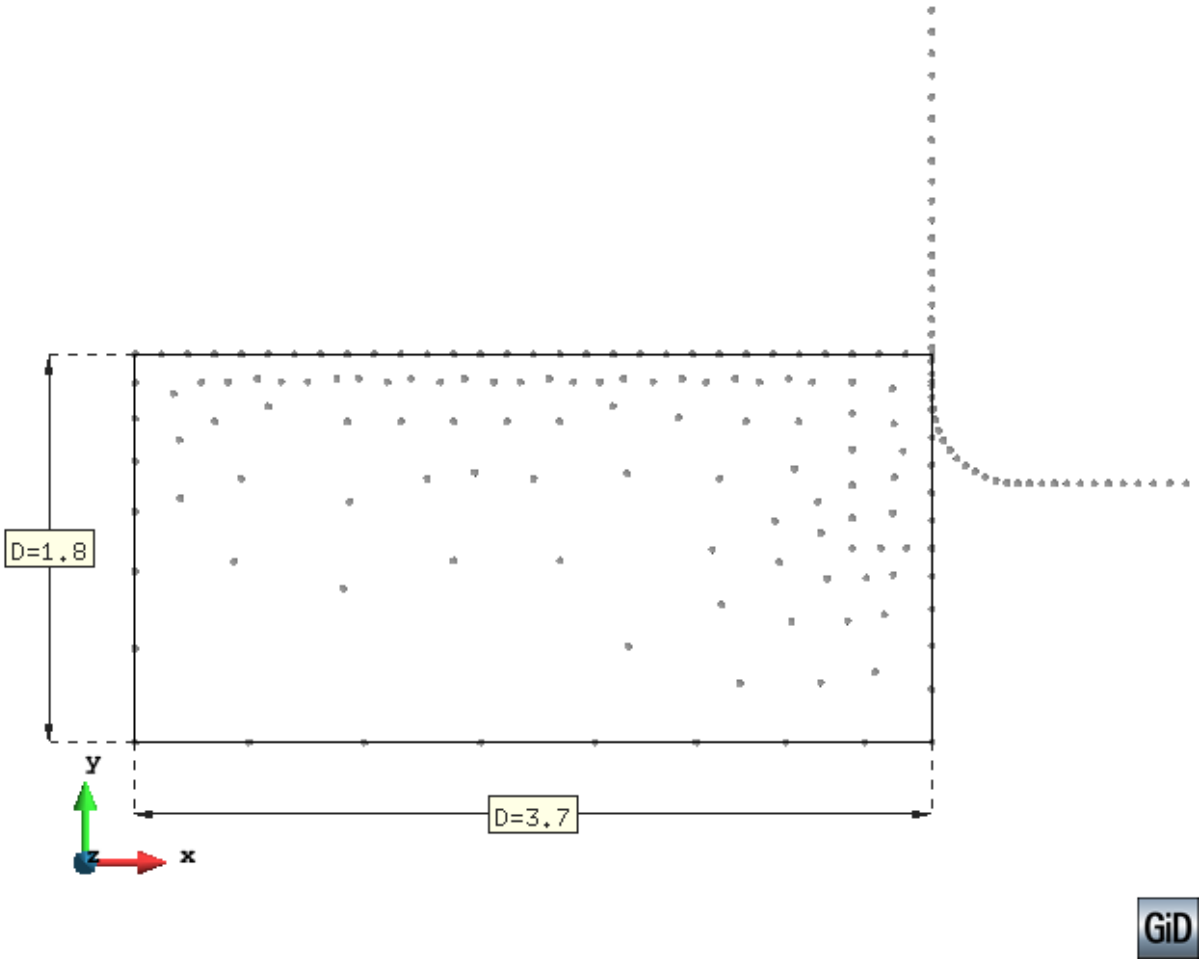


Figure 3. Workpiece dimensions and initial distribution of particles.

Property	**Value**
Elastic Modulus (MPa)	210000
Poisson ratio	0.3
Heat Capacity(N/Km^2)	3.5155×10^6
Conductivity(N/sK)	14
A (MPa)	860
B (MPa)	683
C	0.035
N	0.47
M	1
T_0 (K)	293.15
T_{melt} (K)	1793

Table 4. Materials properties used in the numerical simulations.

Conductivity and specific heat does not depend on temperature, we consider them constant (see Table 4). The following assumptions are made: First, the tool is supposed to be rigid and friction is neglected. Furthermore, the thermal exchange between the part and the tool are also neglected. The inertia of the part is neglected. A classical penalty method is considered for the contact constraint generated by the action of the rigid tool.

A staggered implicit time integration scheme was used. Time steps were of 1.33×10^{-6} s which takes 500 steps for a tool to travel 2.22 mm. The assumption that the tool is rigid is reasonable, since the deformation of the tool is negligible compared with the deformation of the workpiece.

Temperature, pressure, effective plastic strain rate and von Mises contours are presented in Figure 4 and Figure 5. Depicted pressure distribution shows the tension and compression zones. It is completed with the von Mises stress shown, which demonstrates that relatively high stresses arise in the primary shear zone and at the tool chip interface. The localization of this zone agrees with simplified models. It is also important to note the presence of residual stresses at and below the produced new surface and in the upper part of the chip, especially near the tool-chip interface where unloading due to curling of the chip occurred. The effective strain rate in the primary and the secondary shear zone is of the order of 18000 and it has its highest value close to the tool tip. Finally, temperature distribution is also shown in the workpiece.

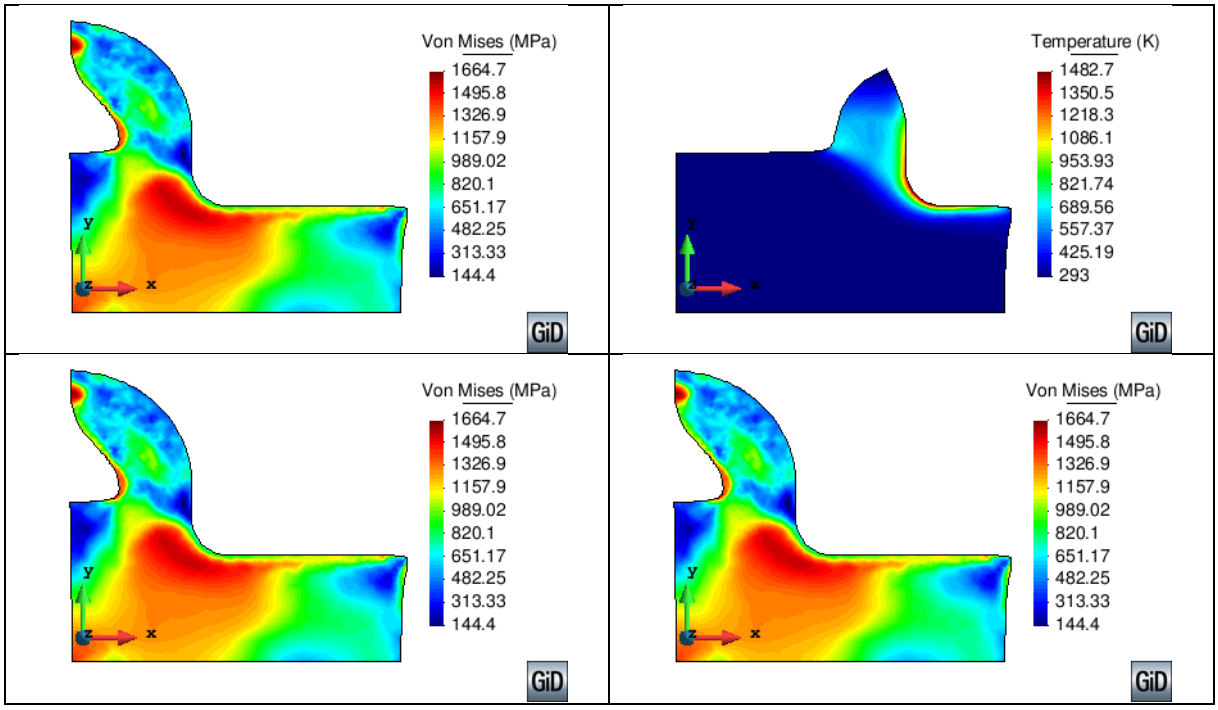


Figure 4. Continuous chip formation using the Johnson-Cook constitutive model: von Mises (MPa) and Temperature (K).

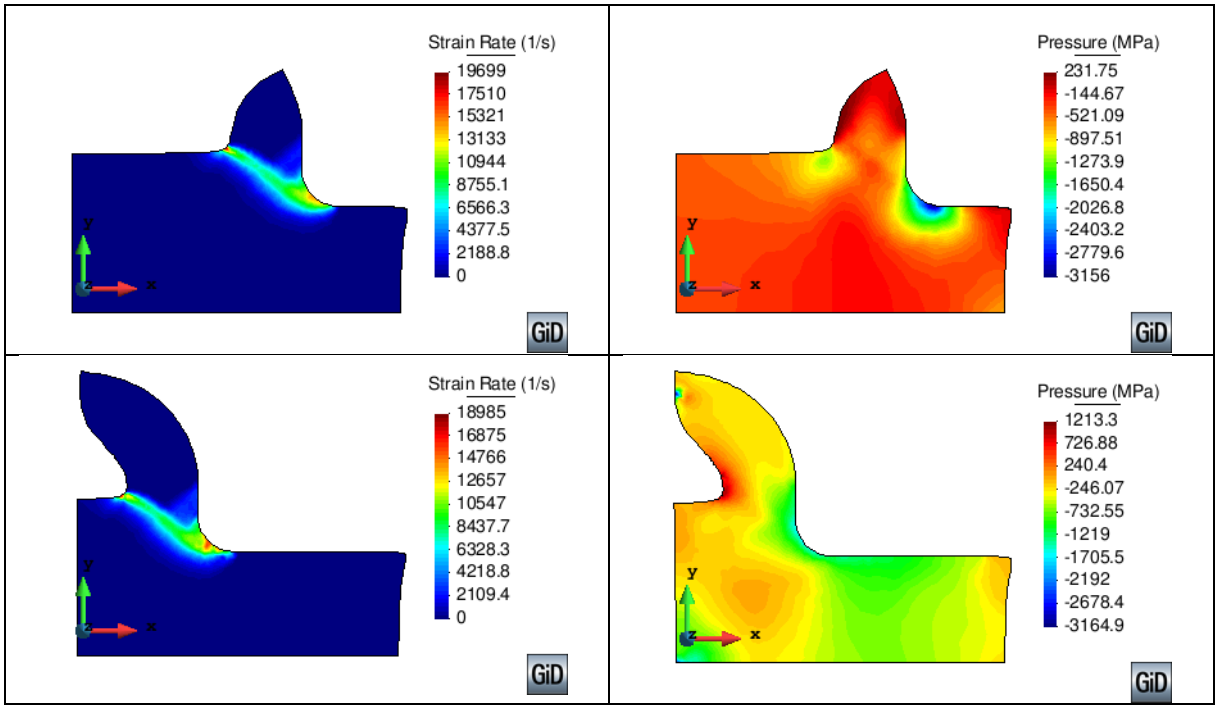


Figure 5. Continuous chip formation using the Johnson-Cook constitutive model: Strain rate (1/s) and Pressure (MPa).

Figure 6 depicts the cutting and feed forces applied on the tool, which are obtained from the simulation. The cutting and feed force reach a steady state as expected in the numerical simulation of continuous chip formation.

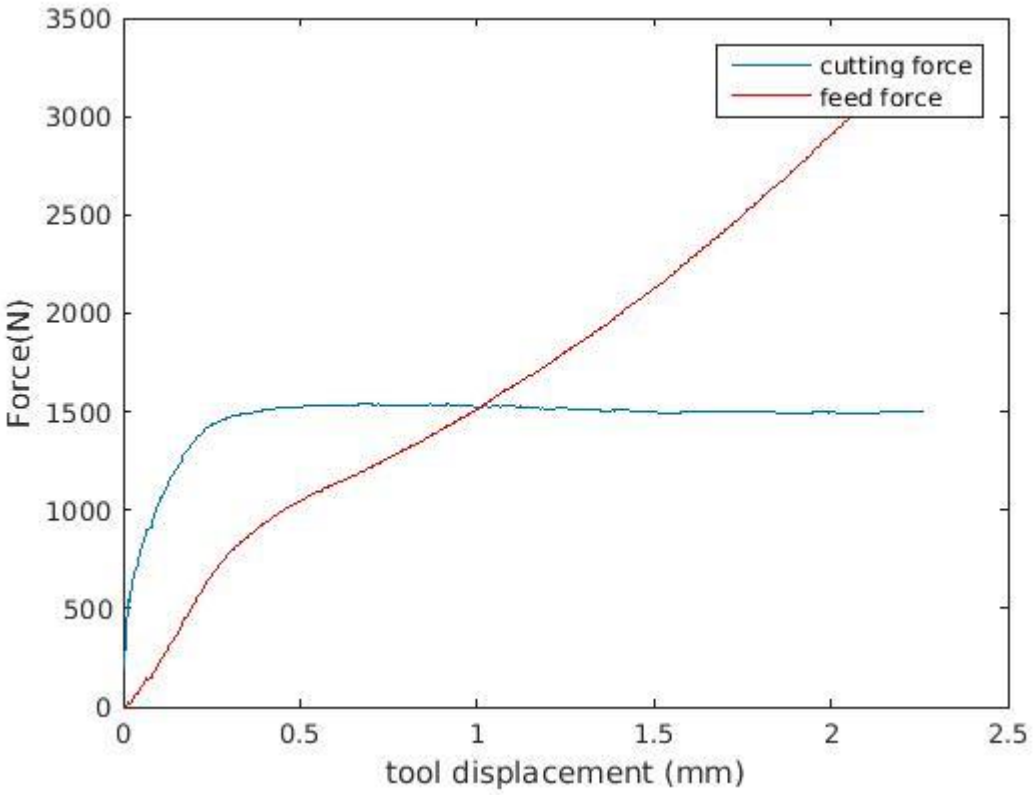


Figure 6. Cutting and thrust force vs. stool displacement.

5.2 Orthogonal cutting test: 3D approach

The material properties for the workpiece, the cutting conditions are the same used in example 5.1. Also, the displacement in the x-direction of front and back faces of the workpiece is restricted to reproduce a plain strain conditions (see example 5.1). The initial finite element meshes for the workpiece and the tool are shown in Figure 7. The workpiece is modeled with 569 particles and the tool with 1592 triangle elements. For saving calculation time and memory space, insertion and removal of particles is used during the simulation. As a consequence, the finer and smaller elements are used in the contact area between the workpiece and the tool because very large gradients in strains and temperature will occurs in this zone. The coarser and the larger elements are allocated in the area of the workpiece not affected by the cutting processes.

Figure 8 shows a three dimensional view of the chip geometry predicted when the tool advances from 0 to 2.22 mm. Temperature distribution is presented in Figure 9. Because temperature increase is primary induced by plastic deformation, the more severe a deformation zone is, the greater the temperature increase.

The cutting force for the 2D and 3D simulations are show in Figure 10. The results for the 2D and 3D are qualitatively the same, the small different between both is due to the larger distance between particles used in the 3D numerical simulations.

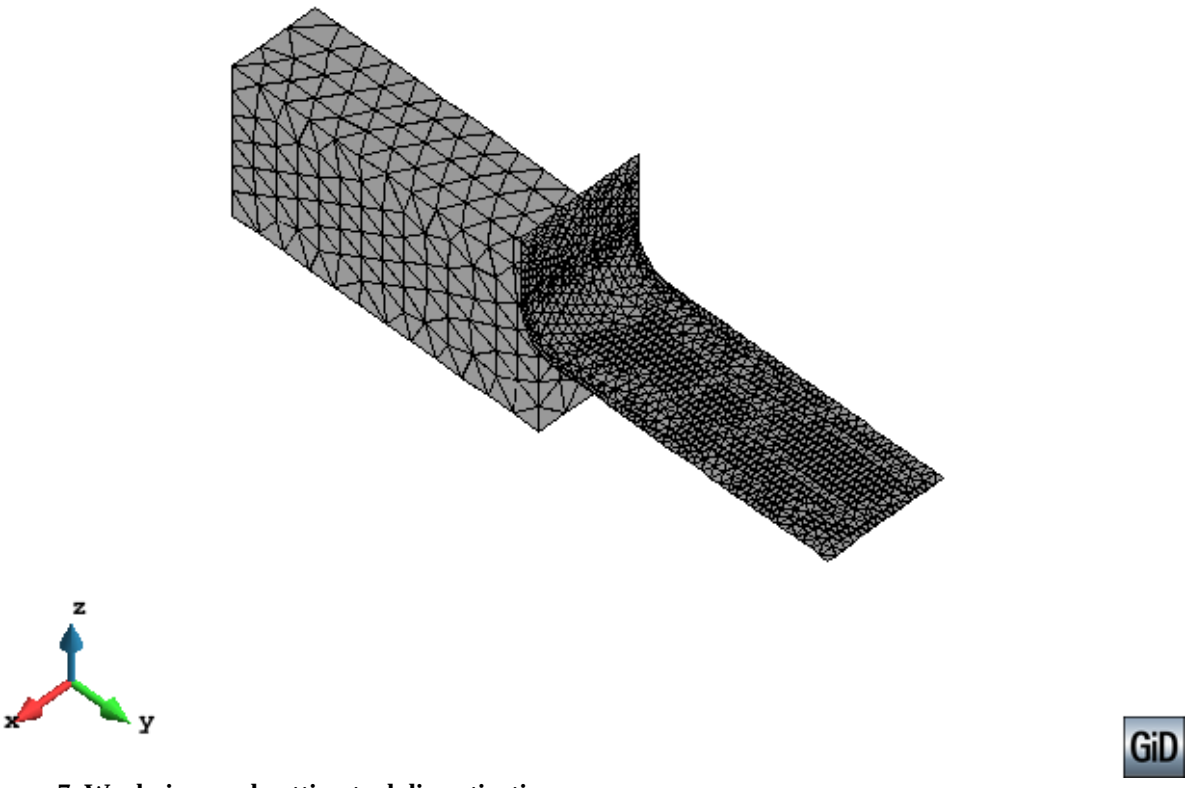


Figure 7. Workpiece and cutting tool discretization.

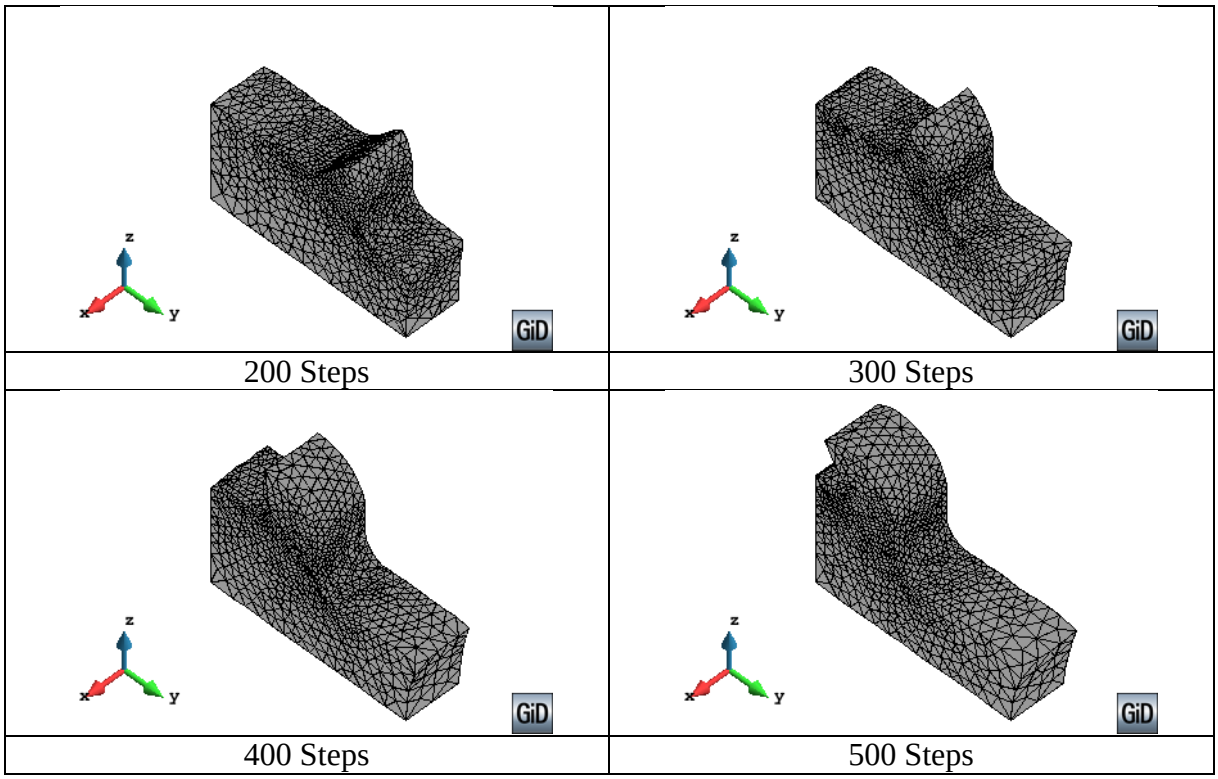


Figure 8. Chip geometries at different time steps during orthogonal cutting.

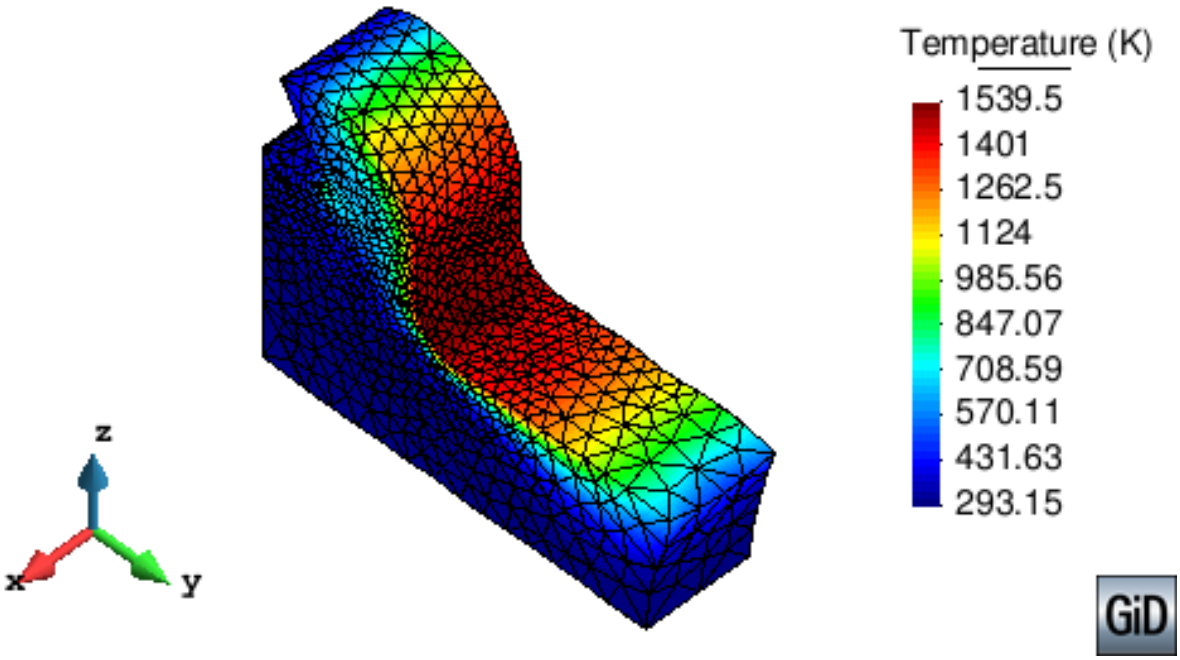


Figure 9. Temperature distribution (K) in the workpiece an the chip (the cutting tool is 2.22 mm)

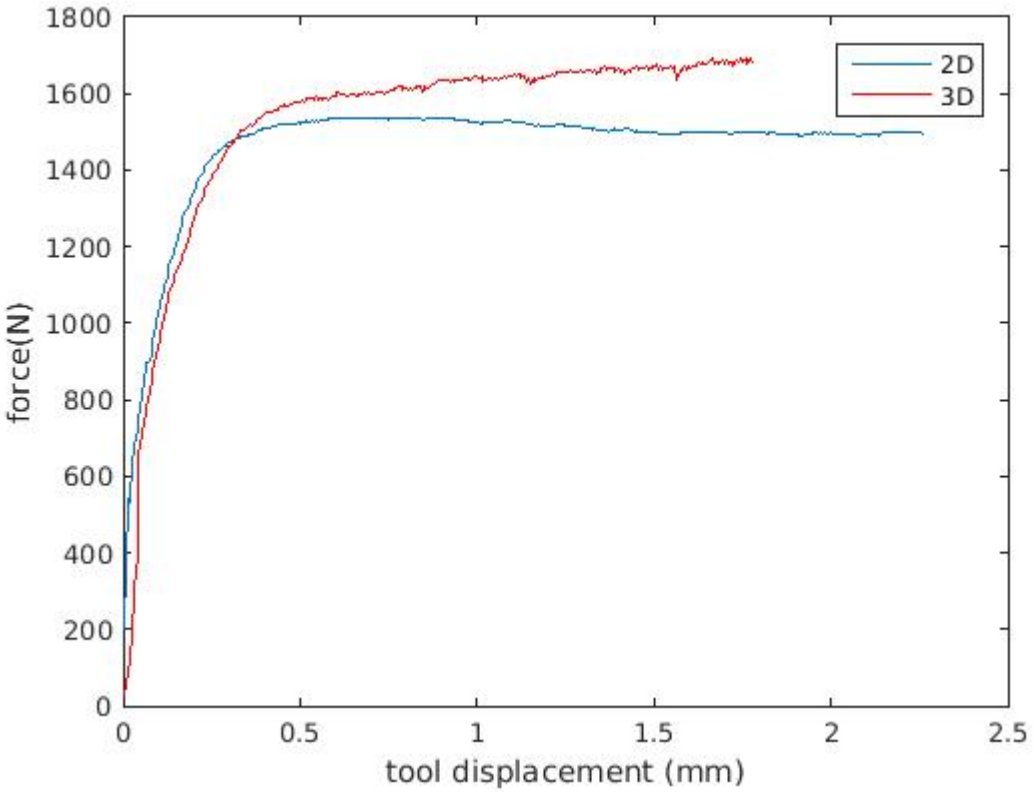


Figure 10. Force displacement curve for orthogonal cutting for 2D and 3D case.

6 CONCLUSIONS

A Lagrangian formulation for analysis of metal cutting processes involving thermally coupled interactions between deformable continua is presented. The governing equations for the generalized continuum are discretized using elements with equal linear interpolation for the displacement and the temperature. The merits of the formulation in terms of its general applicability have been demonstrated in the solution two representative numerical simulations of orthogonal cutting using the PFEM both in 2D and 3D. Results are promising for future numerical simulations of metal cutting processes

REFERENCES

[1] T. D. Marusich and M. Ortiz, "Modelling and simulation of high-speed machining," *International Journal for Numerical Methods in Engineering*, vol. 38, pp. 3675–3694, 1995.

[2] G. S. Sekhon and J. L. Chenot, "Numerical simulation of continuous chip formation during non-steady orthogonal cutting simulation," *Engineering Computations*, vol. 10, 1993.

[3] L. Illoul and P. Lorong, "On some aspects of the CNEM implementation in 3D in order to simulate high speed machining or shearing," *Computer and Structures*, vol. 89, pp. 940–958, 2011.

[4] F. Chinesta, P. Lorong, D. Ryckelynck, M. Martinez, E. Cueto, M. Doblaré, G. Coffignal, M. Touratier, and J. Yvonnet, "Thermomechanical cutting model discretiation: eulerian or lagrangian, mesh or meshless?," *International Journal of Forming Processes*, vol. 7, pp. 83-97, 2004.

[5] F. Greco, L. Filice, C. Peco, and M. Arroyo, "A stabilized formulation with maximum entropy meshfree approximants for viscoplastic flow simulation in metal forming," *INTERNATIONAL JOURNAL OF MATERIAL FORMING,* vol. 8, pp. 341-353, 2015/07/01 2015.

[6] A. Raczy, M. Elmadagli, W. J. Altenhof, and A. T. Alpas, "An eulerian finite-element model for determination of deformation state of a copper subjected to orthogonal cutting," *Metalurgical and Materials Transactions A,* vol. 35, pp. 2393-2400, 2004.

[7] L. Olovsson, L. Nilsson, and K. Simonsson, "An ALE formulation for the solution of two-dimensional metal cutting problems," *Computers & Structures,* vol. 72, pp. 497–507, 1999.

[8] R. Rakotomalala, P. Joyot, and M. Touratier, "Arbitrary Lagrangian-Eulerian thermomechanical finite-element model of material cutting," *Communications in Numerical Methods in Engineering,* vol. 9, pp. 975–987, 1993.

[9] M. Movahhedy, M. S. Gadala, and Y. Altintas, "Simulation of the orthogonal metal cutting process using an arbitrary Lagrangian–Eulerian finite-element method," *Journal of Materials Processing Technology,* vol. 103, pp. 267-275, 2000.

[10] E. Uhlmann, R. Gerstenberger, M. Graf von der Schulenburg, J. Kurnert, and A. Mattes, "The Finite-Pointset-Method for the Meshfree Numerical Simulation of Chip Formation," presented at the 12 Cirp Conference on Modelling of Machining Operations, San Sebastian, Spain, 2009.

[11] E. Uhlmann, R. Gerstenberger, and J. Kuhnert, "Cutting Simulation with the Meshfree Finite Pointset Method," *Procedia CIRP,* vol. 8, pp. 391-396, 2013.

[12] C. R. Dohrmann and P. B. Bochev, "A stabilized finite element method for the Stokes problem based on polynomial pressure projections," *International Journal for Numerical Methods in Fluids,* vol. 46, pp. 183–201, 2004.

[13] J. M. Rodriguez, J. M. Carbonell, J. C. Cante, and J. Oliver, "The particle finite element method (PFEM) in thermo-mechanical problems," *International Journal for Numerical Methods in Engineering,* pp. n/a-n/a, 2015.

[14] H. Edelsbrunner, E. P. M, #252, and cke, "Three-dimensional alpha shapes," *ACM Trans. Graph.,* vol. 13, pp. 43-72, 1994.

[15] J. M. Rodríguez, J. C. Cante, and X. Oliver, "On the numerical modelling of machining processes via the Particle finite Element method (PFEM)," vol. 156, ed. Barcelona: CIMNE, 2015, p. 186.

[16] J. C. Simo, "A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition: part I. continuum formulation," *Computer Methods in Applied Mechanics and Engineering archive,* vol. 666, pp. 199-219, 1988.

[17] J. C. Simo, "A framework for finite strain elastoplasticity based on maximum plastic dissipation and the multiplicative decomposition. Part II: Computational aspects," *Computer Methods in Applied Mechanics and Engineering,* vol. 68, pp. 1-31, 1988.

[18] J. C. Simo and C. Miehe, "Associative coupled thermoplasticity at finite strains: Formulation, numerical analysis and implementation," *Computer Methods in Applied Mechanics and Engineering,* vol. 98 pp. 41–104, 1992.

[19] G. H. Johnson and W. H. Cook, "A constitutive model and data for metals subjected to large strains high strain rates and high temperatures," *Proceedings of the 7th symposium on ballistics,* 1983.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DYNAMICS OF DUST PARTICLES NEAR THE ROUND BELL SUCTION INLET

O. A. Averkova[1], K. I. Logachev[1], A. K. Logachev[1], and E. I. Tolmacheva[1]

[1] Belgorod State Technological University named after V.G. Shukhov,
308012 Belgorod, Russia, Kostukova str., 46
e-mail: {olga _19572004, kilogachev, artlogachev, lenka_110290}@mail.ru

Keywords: local ventilation exhaust, extreme trajectories, discrete vortex method

Abstract. *The main problem of aspiration is the efficient localization of dust-and-gas emissions at the lowest energy consumption. One of the ways to reduce the energy consumption of the local open ventilation exhausts is their profiling by the found outline of the separated-flow region. The round local ventilation exhaust, made in the form of the bell, is the simplest and the most popular installation for catching dust-and-gas emissions. The purpose of this work is determining the influence of the expansion angle and the length of bell exhaust on the separated-flow region's characteristics, formed at the entrance to the bell, on the dynamics of dust particles and their collection efficiency. There was developed a mathematical model, computational algorithm and computer program for calculating separation flows at the entrance to round bell exhausts and for researching the dust particles behavior in their range of action.*
The characteristic sizes of vortex regions at the entrance to the round bell exhaust have been determined. The analytic formula for calculating the separated-flow region at the entrance to the long round bell exhaust has been received. Profiling bells by the found outline of the separated-flow region allows improving the aerodynamic and acoustic properties of local ventilation exhausts. The extreme trajectories of dust particles of various sizes in the range of action of the round bell exhaust have been plotted. The efficiency criterion of a local ventilation exhaust in the form of the collection rate of a given dust particle, which amounts to the ratio of aspiration volume of an exhaust with bell and without it, has been suggested. The computational algorithms of calculating extreme trajectories of dust particles of the local ventilation bell exhaust's range of action are presented. The aspiration volumes and collection rates of dust particles of various fractions depending on the length and expansion angle of the bell are determined. A conclusion about the most efficient round bell exhaust is made. The characteristic sizes of separated-flow region at the entrance to bell outlet, located above a leakproof surface, have been determined. There was suggested an analytic formula for determining the separation line at the entrance to the bell with the expansion angle 180 degrees, located above a leakproof surface, profiling by which will allow to improve the aerodynamic and acoustic properties of a local ventilation exhaust.

1 INTRODUCTION

Local venting suction socket is simplest and most spread element of exhaust ventilation systems. Interesting is determination of the boundaries of the gas separated flows area at inlets of the suction channels. Profiling on these boundaries improve the aerodynamic and acoustic properties of socket. Similar research for the slot-type suction-socket were first presented in a series of articles [1-3]. There used of complex variable theory methods. Separated flows for round-type suction socket investigated using the non-stationary discrete vortex method [4-5]. Characteristic dimensions of the vortex region inlet to the socket (Figure 1) at a fixed point in time were obtained. But it was not taken into account, that these dimensions pulsate with time. The paper [6] was an attempt to average over time value of the characteristic dimensions. Averaging was carried out by finding arithmetic mean of the values of 20 random points in time. The calculation was quite laborious. But 20 points in time is not sufficient to obtain convergent results in some cases. Therefore, it became necessary to calculate the gas separated flows area by the stationary discrete vortex model. Idea of this model was first described in the article [7], and then developed in the works [8-11,16]. Moreover, it is of interest determination such angles and length of suction socket, for which local venting suction socket more efficient.

2 EFFICIENCY CRITERIA OF SUCTION SOCKET

There are various approaches to determining of socket efficient. One of them is the criterion irregularity of suction which is determined by speed ratio at the center of the socket to the average velocity of suction $\eta = v_c / v_0$. For duct without the socket according to the data [12] $\eta = 0{,}95 - 0{,}97$; for $\alpha = 60^\circ$ – $\eta = 1{,}03 \div 1{,}09$; for $\alpha = 90^\circ$ – $\eta = 1{,}65 \div 1{,}75$. The studies were conducted by Fialkovskaya, T.A. [13] and Baturin, V.V. [14]. It determines field of velocities in the zone of exhaust hood square and rectangular shape with aspect ratio of 1:2, 1:3, 1:4 and at different angles between defining lines from 50° to 90° (obviously it is the angle α shown in Figure 1).

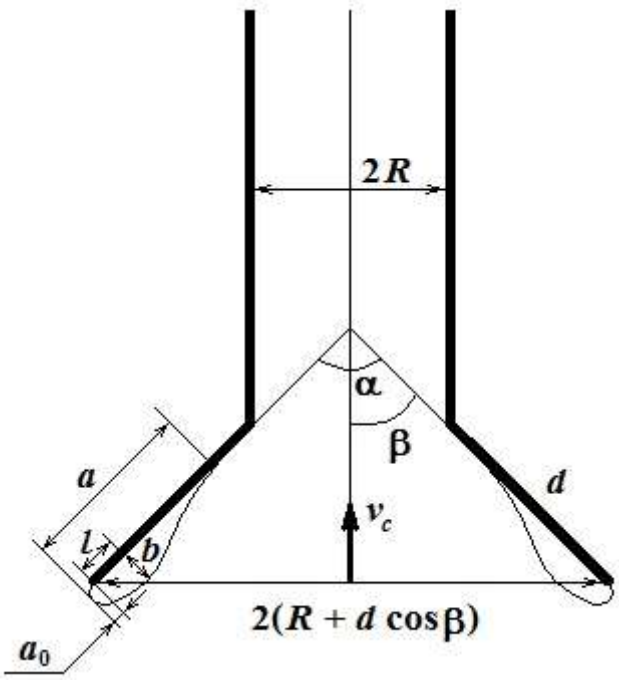


Figure 1: Round local ventilation suction - bell

About the length of socket information is not accounted. About suctions of round-socket is also no mentioned. It was concluded, that because hoods with opening angle $\alpha = 60^\circ$ the relative velocity is nearly to the average velocity in the center of hood hole the suction hole (flow rate per area), then the apex angle of hood must be less than $\alpha = 60^\circ$. Dimensions of the inlet hood must be greater than dimensions of the heat source.

Further, approval about most efficient $\alpha = 60^\circ$ published in numerous of books. For example in the book [15] it indicated that in order to hood working by full air intake cross section, opening angle of hood accept usually not more than $\alpha = 60^\circ$. At larger angles actually acts only the central part of hood, and on the periphery appeared dead zones.

There are also other positions. It stated that the hood should be done with the opening angle $\beta = 60^\circ$ in the Posohin V.N. book [15]. It was noted that in this case the dimension of the vortex area at the inlet of the socket is minimal and accordingly the "efficient dimension" hood is maximum. But here under the opening angle is meant $\beta = \alpha / 2$. The researches were conducted with the help of Conformal Mapping Method and [1-3] considered slot-socket suction. At here width of effective suction accepted as efficiency criterion.

Other efficiency criterion is highest rate of suction in the suction socket axis. It is shown by numerical and analytical studies [17] that the axial velocity reaches a maximum value when $\beta = \alpha / 2 = 90^\circ$. It used Conformal Mapping Method and discrete vortex Method too. Also it considered slot-type and round-type suction socket.

It can specify the aerodynamic drag as one more efficiency criterion of suction socket. By the graphs given in [18] we conclude, that the suction socket has smallest local resistance coefficient (l.r.c.) at $\alpha = 50^\circ - 100^\circ$. It depends on the length of the socket. If the socket length is smaller, then the change of l.r.c. from the opening angle is approaching to linear law.

It proposed to introduce another criterion of efficiency: non-dimensional capture coefficient – the ratio of the aspiration volume of the suction-socket to aspiration volume of air duct without socket, because the local ventilation suckings [16, 19, 21-25] used not only for capture the excess of heat but also for various types of dust capture, which working with abrasive-metallic dust, wood dust, cement, sand, welding fumes and etc. Aspiration volume - is the volume of space being in which the dust particle with preset geometrical and physical properties will be caught by the local ventilation suction.

The main goal of the paper is the following: to elaborate a mathematical model and an algorithm of calculation of separation flow region at the inlet round-type suction socket with the use of the stationary discrete vortex method; to reveal the regularities in changes of the parameters of vortex region at socket inlet and to make a research of the efficiency of the local suction socket with the use of the suggested criterion.

3 DEVELOPMENT OF A MATHEMATICAL MODEL AND A COMPUTATIONAL ALGORITHM

In order to develop a mathematical model of the separated flow at the inlet to the round suction-bell used discrete infinitely thin vortex rings. Border suction funnel discretize attached set of vortex rings and control points (arbitrary points on a circle, embracing the suction-bell). The free surface of current begins on the sharp edge of the bell, the initial approximation is given for it. A discrete model in the meridian half-plane is shown in Figure 2. The crosses are depicted from control points, in which fulfils the condition of impenetrability - velocity along the direction of the normal is zero. The shaded circles -It is the attached ring vortices. Hollow circles - free vortex rings.

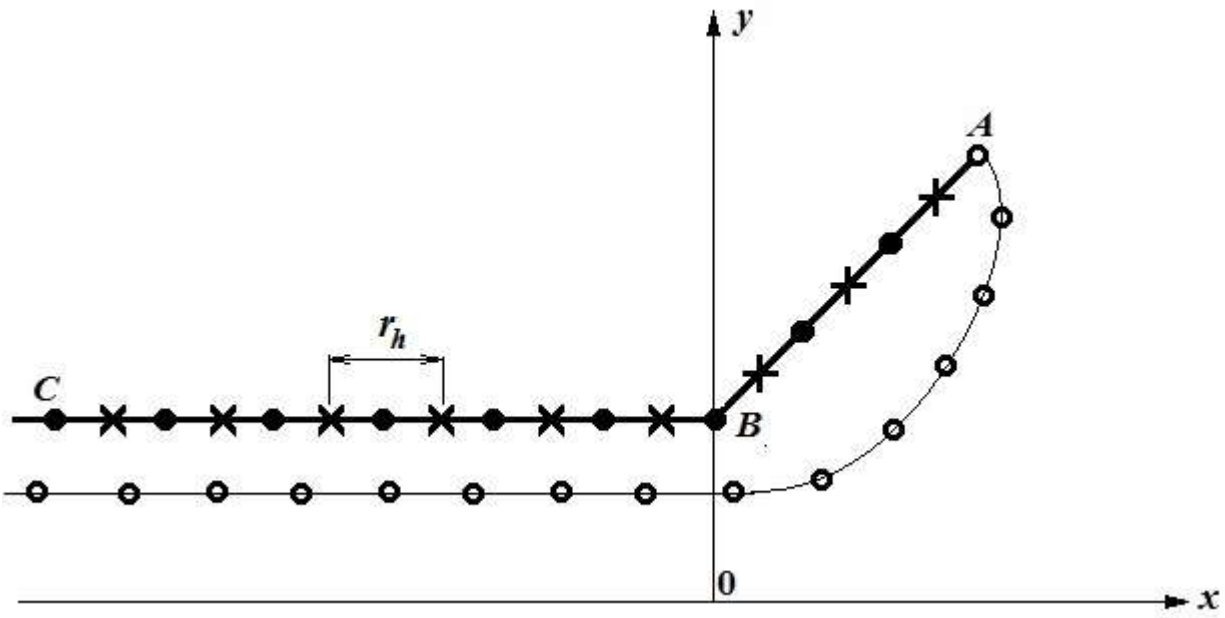


Figure 2: Discrete model of separated flow at the inlet to the suction-bell

Let N – number of attached vortex rings ; N_s – the number of free vortex rings; x^p - control point, $p = 1, 2, \ldots, N$.

The velocity at any point x along the direction $\vec{n}$ is calculated using the formula:

$$v_n(x) = \sum_{q=1}^{N} \Gamma(\xi^q) G(x, \xi^q) + \gamma \sum_{q=1}^{N_s} G(x, \zeta^q), \qquad (1)$$

Where ξ^q - the location of q-th attached vertex ring with circulation $\Gamma(\xi^q)$, $\gamma = \text{const}$ - given free circulation of the vertex ring, ζ^q - location of the point q-th free vortex ring.

Function $G(x, \xi)$ expresses the influence of that point x of vortex ring with single loop, located at the point ξ

$$\begin{cases} G(x,\xi) = \dfrac{(A_1 b + A_2 a)}{b} \cdot \dfrac{4}{(a-b)\sqrt{a+b}} E(t) - \dfrac{A_2}{b} \cdot \dfrac{4}{\sqrt{a+b}} F(t) & \text{at } b \neq 0, \\ G(x,\xi) = \dfrac{\xi_2^2 \cdot n_1}{2a\sqrt{a}} & \text{at } b = 0, \end{cases}$$

$2x_2\xi_2 = b > 0$, $a = (x_1 - \xi_1)^2 + \xi_2{}^2 + x_2{}^2 > 0$, $A_1 = \dfrac{\xi_2{}^2 n_1}{4\pi}$, $A_2 = \dfrac{\xi_2}{4\pi}\left[(x_1 - \xi_1)n_2 - x_2 n_1\right]$,

$F(t) = \int_0^{\pi/2} \dfrac{d\theta}{\sqrt{1 - t^2 \sin^2\theta}}$, $E(t) = \int_0^{\pi/2} \sqrt{1 - t^2 \sin^2\theta}\, d\theta$;

$t = \dfrac{2b}{a+b}$; $F(t) = \sum_{i=0}^{4} c_i (1-t)^i + \sum_{i=0}^{4} d_i (1-t)^i \ln\dfrac{1}{1-t}$; $E(t) = 1 + \sum_{i=1}^{4} c_i (1-t)^i + \sum_{i=1}^{4} d_i (1-t)^i \ln\dfrac{1}{1-t}$;

c_i, d_i taken from the tables [20].

If the distance from the point x to the point ξ smaller than the steps of discreteness r_h, than this function is calcuated as follows:

$$G(x,\xi)=\frac{(x_1-\xi_1)n_2-(x_2-\xi_2)n_1}{2\pi r_h^2},$$

where r_h - step of discreteness (distance between adjacent vortex rings).

In case $x=\xi$, function $G(x,\xi)=0$.

The numerical algorithm is constructed as follows. After specifying the location of the attached vortex points and the initial approximation for free vortices forms matrix.

$$G^{pq}=G(x^p,\xi^k);\ p=1,2,\ldots,N;\ q=1,2,\ldots,N.$$

Further begins the first intergation.

Forms column of free terms:

$$v^p=-\gamma\sum_{q=1}^{N_s}G(x^p,\zeta^q),\ p=1,2,\ldots,N.$$

Solves systems of linear algebraic equations:

$$\sum_{q=1}^{N}\Gamma^q G^{pq}=v^p,\ p=1,2,3,\ldots N,$$

From where determines unidentified quantities $\Gamma^q=\Gamma(\xi^q)$.

Constructed free current lines, starting with a sharp edge A. Using the formula (1), at $\vec{n}=\{1,0\}$ calculated velocity components v_x, at $\vec{n}=\{0,1\}$ calculated velocity componets v_y. Following point (x',y') is determined from the previous (x,y) using the given formula:

$$x'=x+\Delta t v_x/\sqrt{v_x^2+v_y^2},\quad y'=y+\Delta t v_y/\sqrt{v_x^2+v_y^2},$$

where Δt - step, which is chosen sufficiently small.

Integration process for building free surfaces of current finishes, if the position of the free surface of current stops to change with precisión to the desired accuracy.

After the parameters of separated flow región were determined(Figures 3-7).

Proposed formula for determining the boundary of the vortex at the inlet to the suction-bell.

$$\bar{x}(t)=a\frac{1-e^{\frac{t-2\pi-\beta}{3,1t-3(\pi+\beta)}}}{R(t-\pi-\beta)^b}\cos t+\bar{d}\cos\beta;\quad \bar{y}(t)=a\frac{1-e^{\frac{t-2\pi-\beta}{3,1t-3(\pi+\beta)}}}{R(t-\pi-\beta)^b}\sin t+1+\bar{d}\sin\beta;$$

where $\pi+\beta<t\le 2\pi+\beta$; $\bar{x}(t)=x(t)/R$; $\bar{y}(t)=y(t)/R$;

$$a=(0,03\bar{d}-0,034)\frac{\beta}{\pi}+0,01\bar{d}+0,032;\quad \bar{d}=d/R;\quad b=0,38-0,48\frac{\beta}{\pi}.$$

The formula is valid for long sockets, where $15^\circ\le\beta\le 90^\circ$, $l/B\ge 2$; when $l/B=1$ and $45^\circ\le\beta\le 90^\circ$. The calculation is carried out with $x>0$.

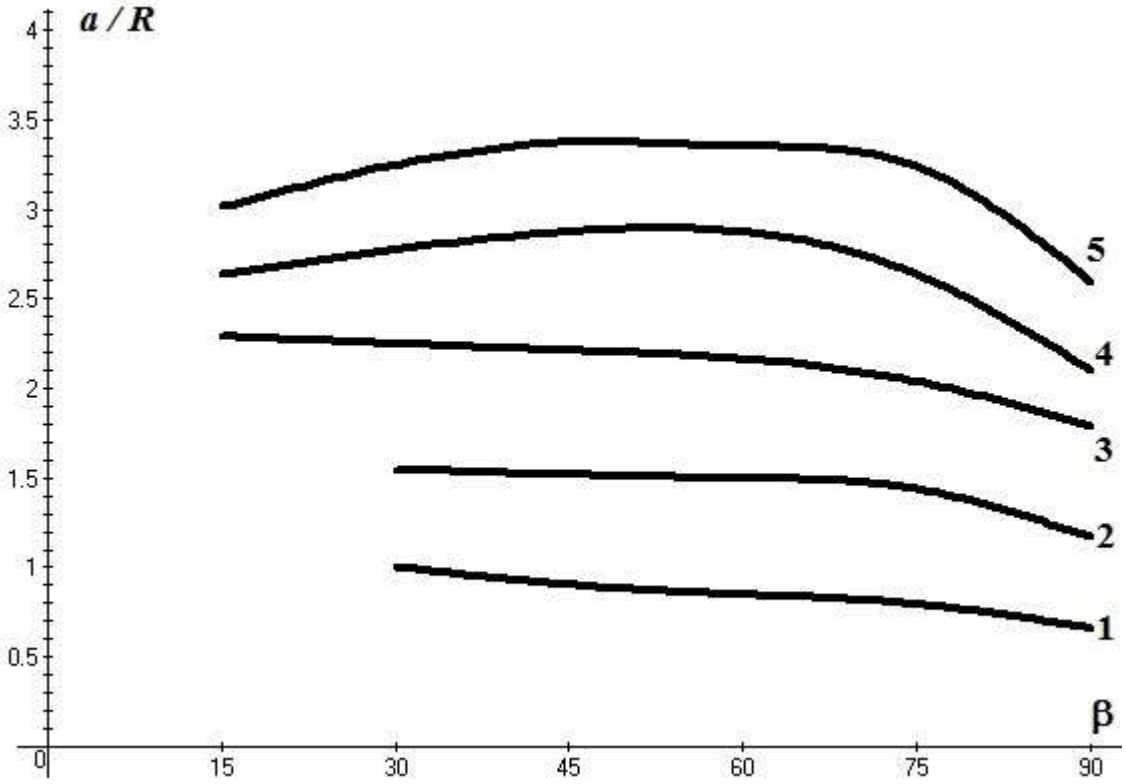


Figure 3: The dependence of the vortex field length a at the entrance to the suction-bell

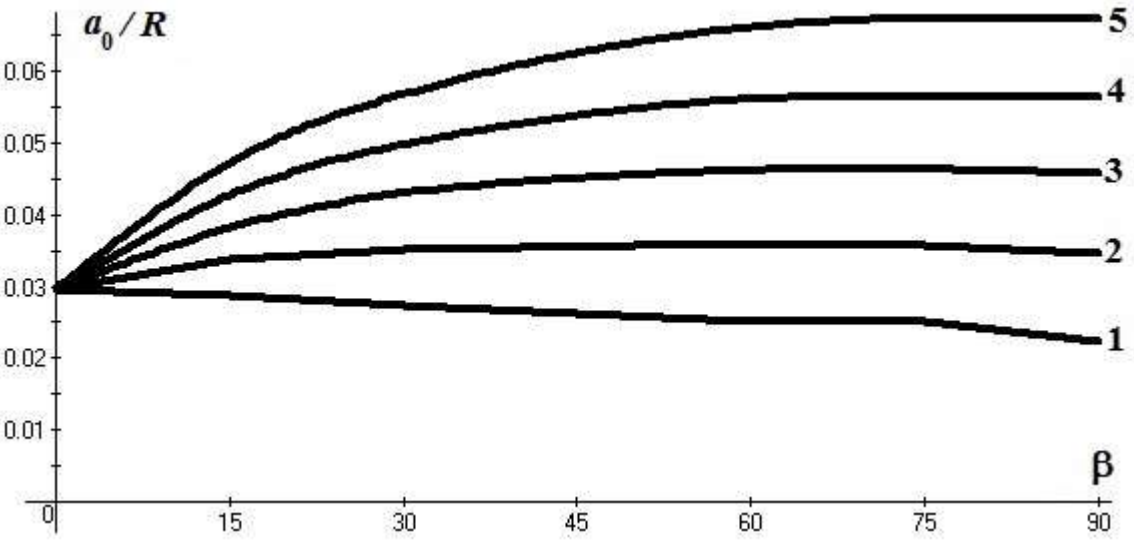


Figure 4: The change in the value a_0 of removal of the separation region behind the socket

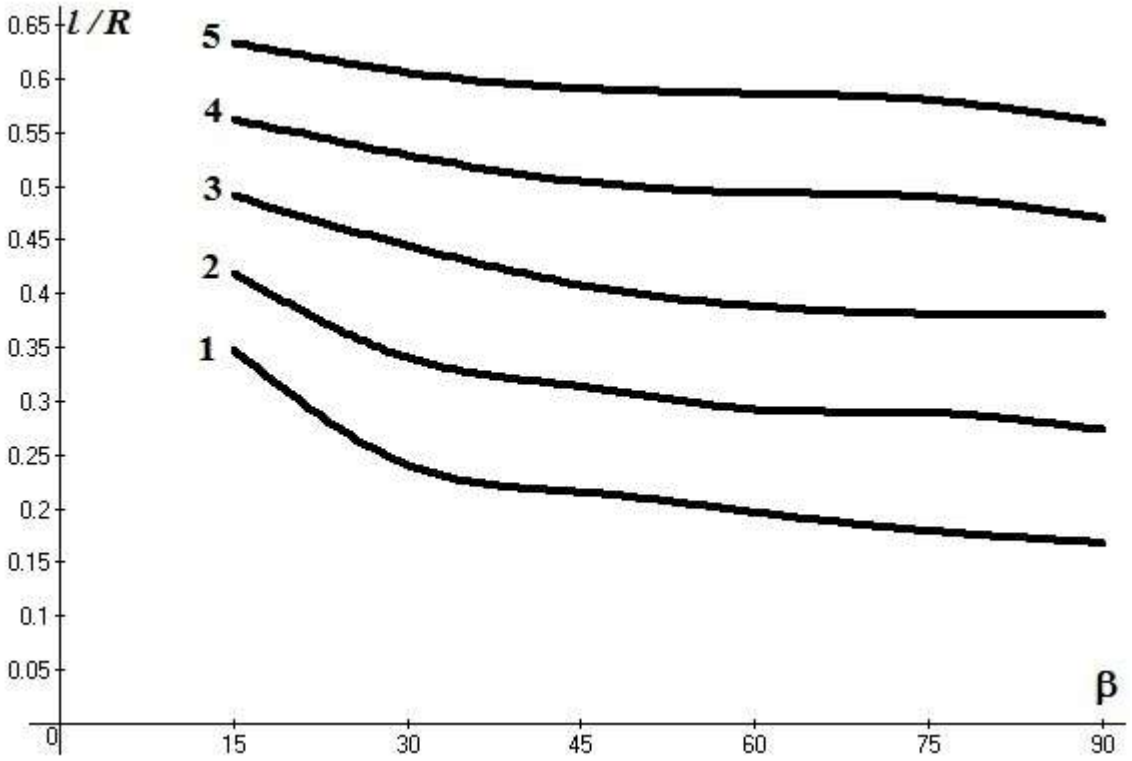


Figure 5: The change of parameter l from the angle of the funnel opening

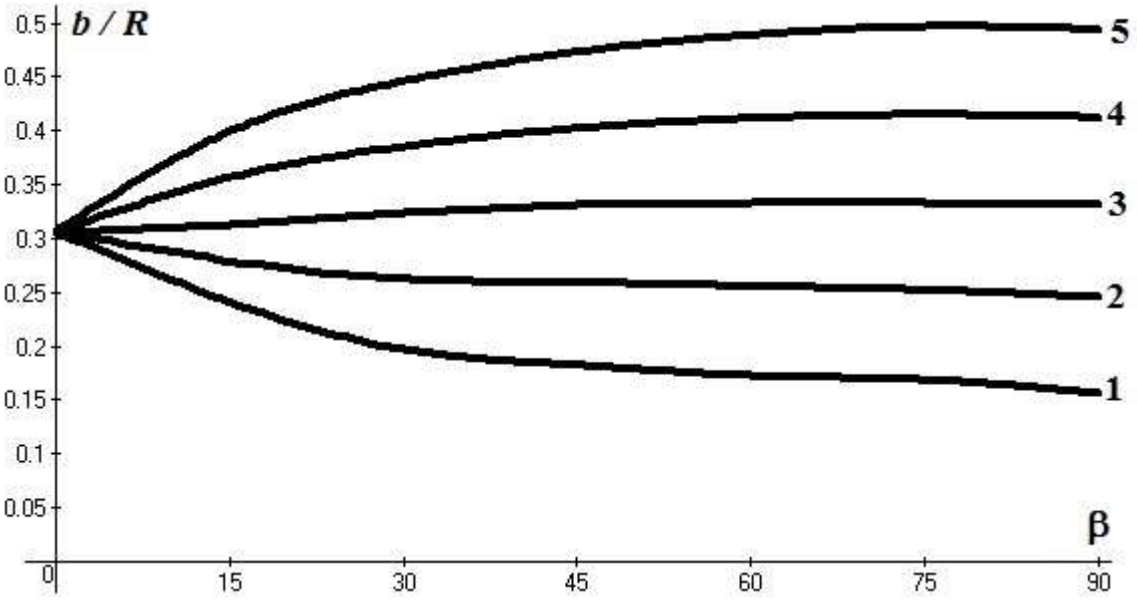


Figure 6: The change of the maximum thickness b of separation region

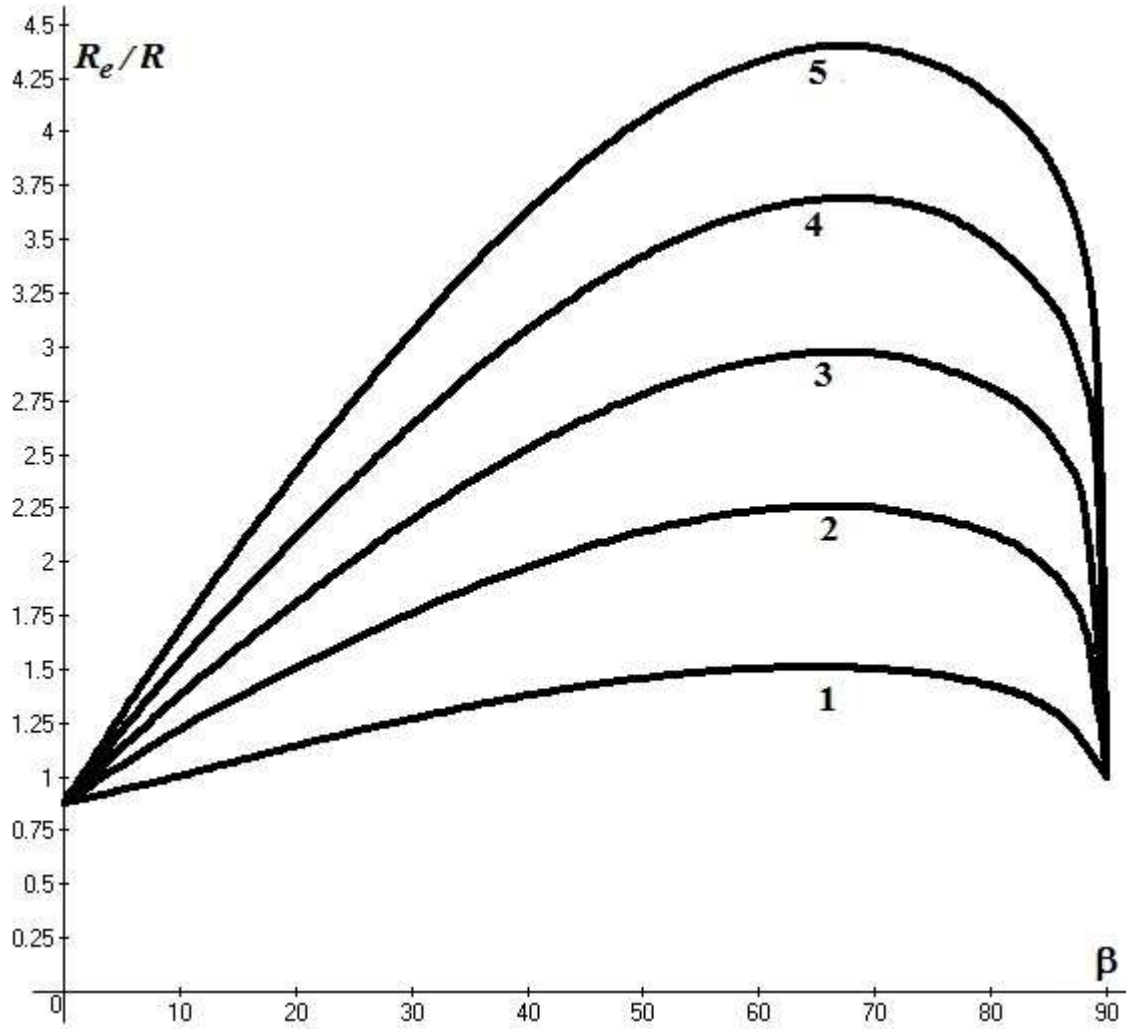


Figure 7: The change of radius of the effective absorption

4 THE DYNAMICS OF DUST PARTICLES

The trajectories of dust particles, build with the help of differential equation[17]:

$$\frac{\pi d_e^{\,3}}{6}\rho_p \frac{d\vec{v}_p}{dt} = \frac{\pi d_e^{\,3}}{6}\rho_p \vec{g} - \psi\chi S_m \rho_a \frac{\left|\vec{v}_p - \vec{v}_a\right|\cdot(\vec{v}_p - \vec{v}_a)}{2},$$

where $\vec{v}_a$ – velocity of the air; ρ_a – air denstity; $\vec{v}_p$ – the particles velocity; ρ_p –the particles density; d_e – equivalent diameter of the particles; $\vec{g}$ – accleration of free fall; $S_m = \pi d_e^2/4$ – mid-section area of the particle; χ – coefficient of its dynamic form; ψ – coefficient of air resistance:

$$\psi = \begin{cases} 24/\mathrm{Re}, & \text{at} \quad \mathrm{Re} < 1, \\ 24\left(1 + 1/6 \cdot \mathrm{Re}^{2/3}\right)/\mathrm{Re}, & \text{at} \quad 1 \le \mathrm{Re} < 10^3, \\ 24/\mathrm{Re}\cdot\left(1 + 0{,}065\,\mathrm{Re}^{2/3}\right)^{1.5}, & \text{at} \quad \mathrm{Re} \ge 10^3, \end{cases}$$

where $\mathrm{Re} = \dfrac{\rho_a \left|\vec{v}_p - \vec{v}_a\right| d_e}{\mu}$, μ – coefficient of dynamic viscosity of air.

Limited trajectories of the dust particles have diameter 1-120 microns (Figure 8).

Volume of aspiration- It is the volume between the plane A-A (Figure 8), passing through the entrance into the bell and the surface of limited trajectory of dust particles.

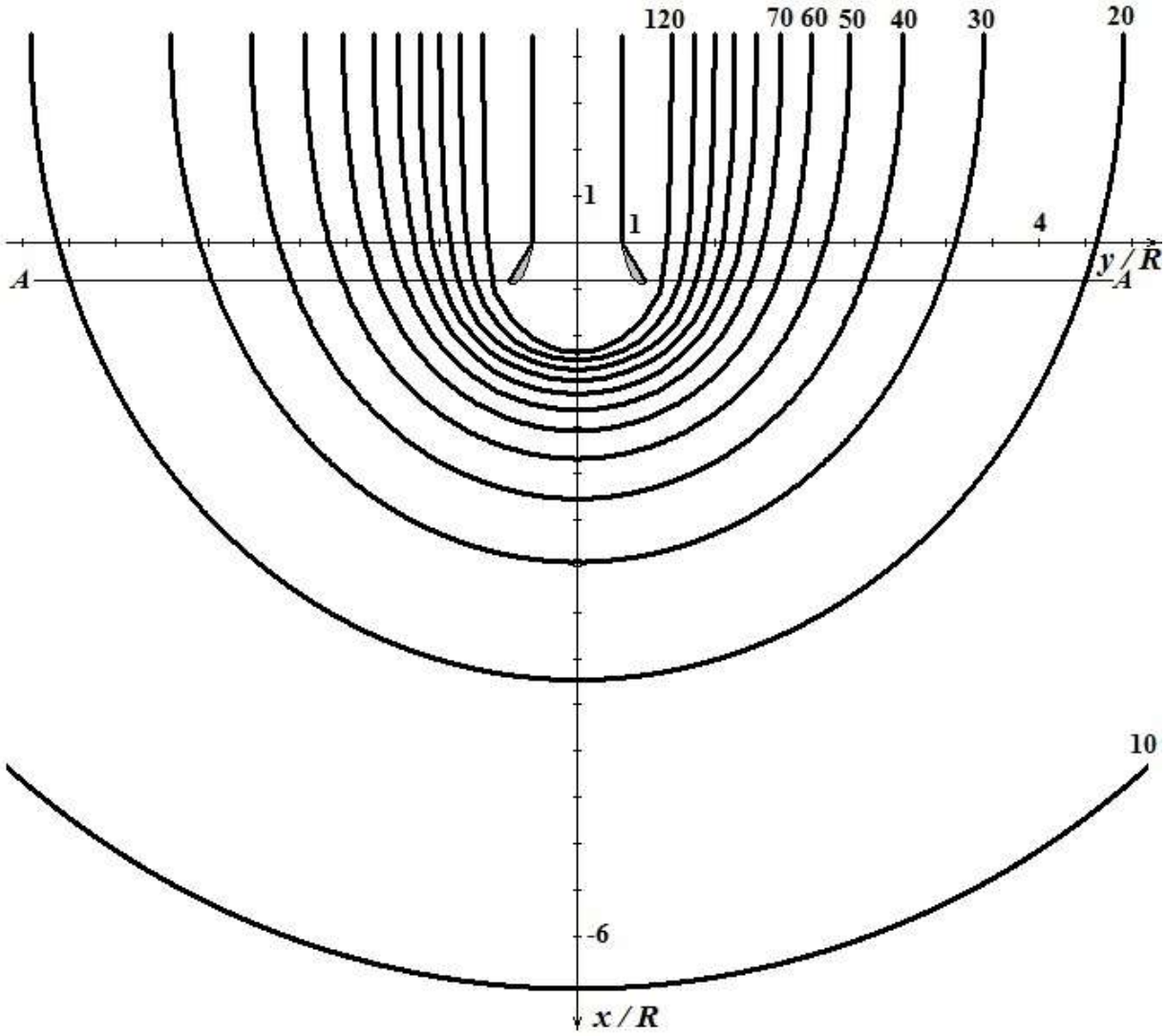


Figure 8: The limit trajectory of dust particles in the spectrum of action circular suction-bell length of 1gauge with an opening angle $\alpha = 60^{\circ}$ when the middle suction speed is 10 m / s

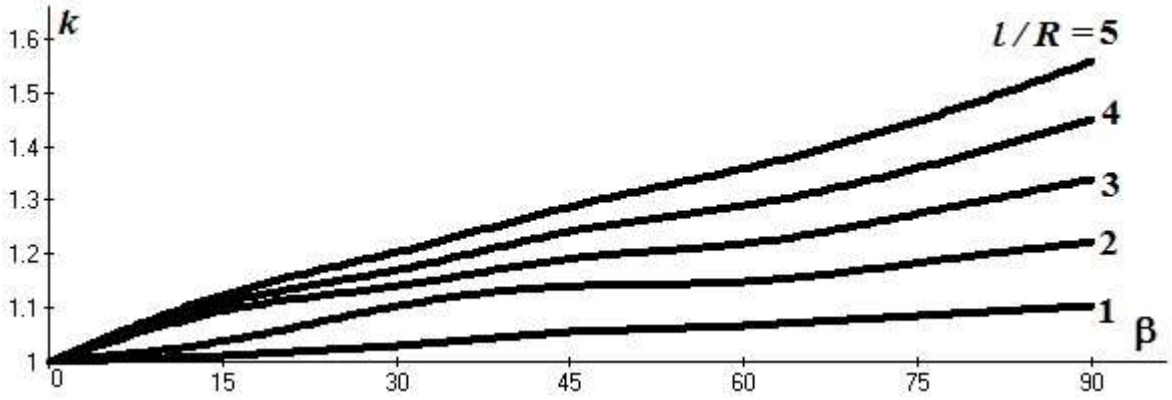
Under the criterion of efficiency of local ventilation of suction-bell for dust particles, we understand the collection coefficient.

$$k = V_a / V_0,$$

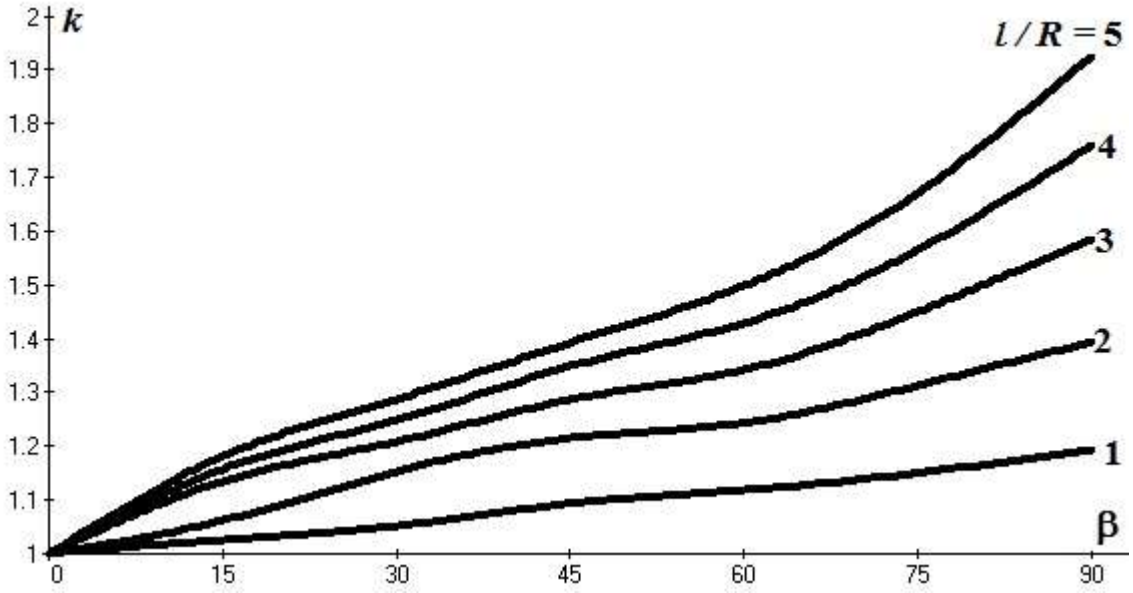
V_0 - volume of aspiration suction with the socket; V_0 - volume of the aspiration suction without socket.

In calculations: $\rho_p = 2500\,\text{kg/m}^3$; $\chi = 1$; $g = 9{,}81\,\text{m/s}^2$; $\rho_a = 1{,}2\,\text{kg/m}^3$; $\nu = 1{,}5\cdot 10^{-5}\,\text{m}^2/\text{s}$; $v_0 = 10$ m/s.

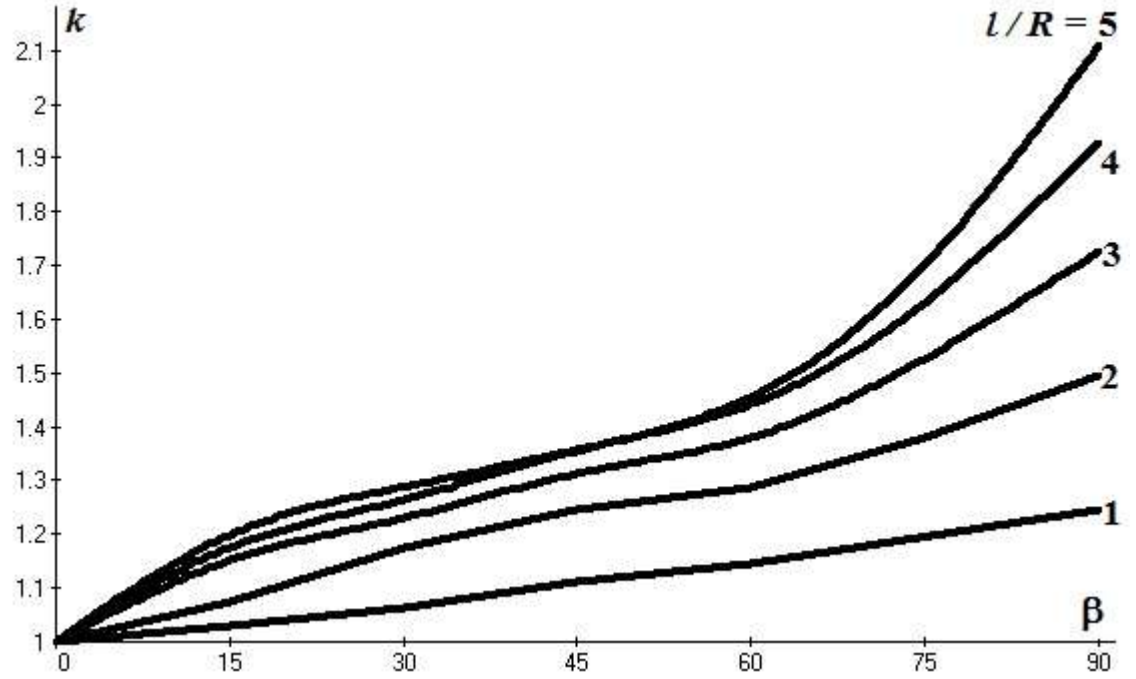
For particles with diameters 20 mkm, 40 mkm, 60 mkm, 80 mkm the collection coefficient is highest when $\beta = 90^\circ$ (Figure9).



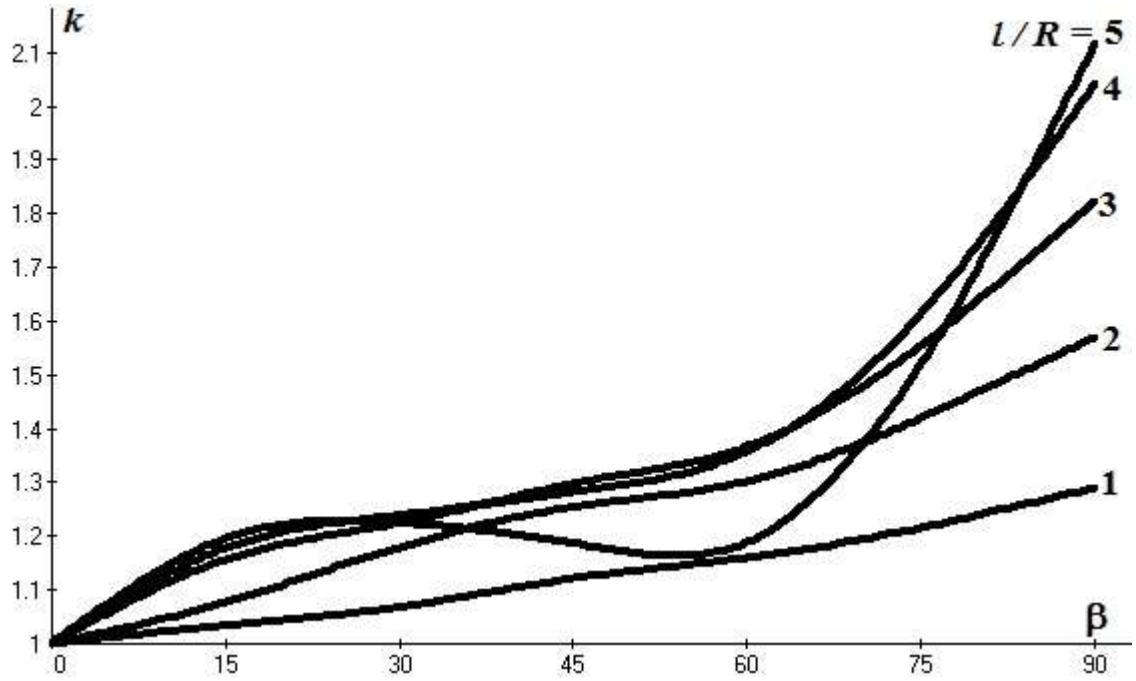
a) 20 mkm



b) 40 mkm



c) 60 mkm



d) 80 mkm

Figure 9: The change in the coefficient capture from the angle for different lengths of the funnel

The effect of collection coefficient on the length of the socket is shown in figure 10.

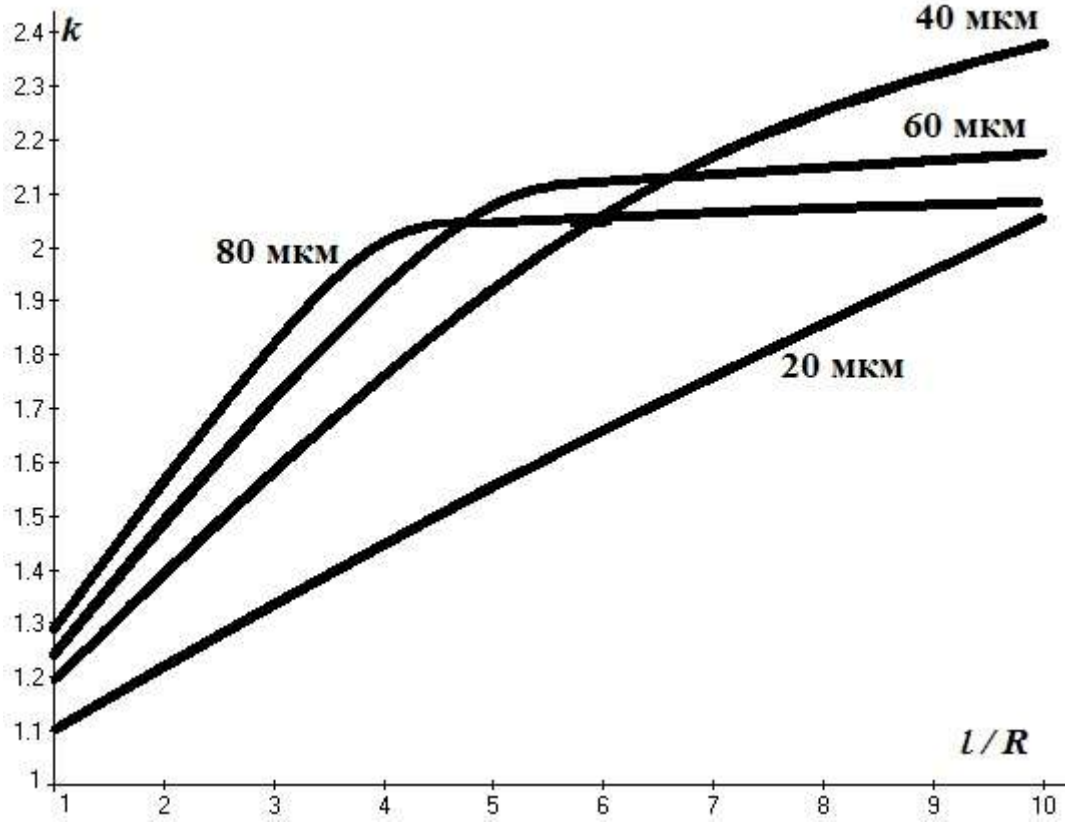


Figure 10: The change in coefficient capture according to the length of the socket when $\beta = 90^{\circ}$.

5 CONCLUSIONS

- A mathematical model is designed, computational algorithm and a computer program to calculate separated flows at the entrance to the round suction-bells and to study the behavior of dust particles in the spectrum of their activities.
- Identified characteristic dimensions of the vortex areas at the entrance to the round suction-bell. Found an analytical formula for calculating the boundaries of the separation region at the entrance to a long suction-bell. Profiling sockets found on the outline of the tear region will improve the aerodynamic and acoustic properties of the local air suction.
- Built limiting trajectories of dust particles of different sizes in the range of action of the round suction-bell. Proposed the criterion on the effectiveness of local suction ventilation in the form of collection coefficient of the given dust particles, which is equal to the ratio of volumes of aspiration suction with socket and without it.
- The most effective suction is the funnel with an opening angle $\alpha = 180^{\circ}$ and length of at least 4 gauge.

5 ACKNOWLEDGEMENTS

The reported study was funded by RFBR according to the research project № 16-08-00074*а и* and Russian President Scholarship СП-121.2016.1.

REFERENCES

[1] V.N. Posokhin, N.B. Salimov, K.I. Logachev, A.M. Zhivov, Regarding analysis of flows around a slotted bell-shaped hood inlet. Message 1. *News of Higher Educational Institutions. Construction*, **8**, 70-76, 2002

[2] V.N. Posokhin, N.B. Salimov, K.I. Logachev, A.M. Zhivov, Regarding analysis of flows around a slotted bell-shaped hood inlet. Message 2. *News of Higher Educational Institutions. Construction*, **9**, 80-85, 2002

[3] V.N. Posokhin, N.B. Salimov, K.I. Logachev, A.M. Zhivov, Regarding analysis of flows around a slotted bell-shaped hood inlet. Message 2. *News of Higher Educational Institutions. Construction*, **10**, 81-85, 2002.

[4] K.I. Logachev, A.I. Puzanok, V.N. Posokhin, Analysis of flows around a slotted bell-shaped hood inlets using the discrete vortex method. *Izv. vuzov. Problems of energetics*, **7-8**, 61-69, 2004.

[5] K.I. Logachev, V.N. Posokhin, A.I.Puzanok, Geometrical characteristics of flows at the entrance of hood-shaped exhausts. *Engineering systems. AVOK North-West*, **1**, 12-14, 2005.

[6] O.A. Averkova, A.E. Kanar, E.I. Tolmacheva, I.V. Khodakov, Average characteristics of eddy regions at the entrance of a round bell duct. *The book includes: Energy saving, information technologies and sustainable development. Electronic scientific publications: digest of the International research and application internet conference. FSBEI "IzhGTU named after M.T. Kalashnikov"*, Izhevsk, 162-167, 2014

[7] K.I. Logachev, V.N. Posokhin, Analysis of flows around intake pipes of round cross-section. *Izv. vuzov. Aeronautical engineering*, **1**, 29-32, 2004.

[8] O.A. Averkova, I.N. Logachev, K.I. Logachev, Modeling of potential flows with unknown boundaries on the basis of stationary discrete vortices. *Computational methods and programming*, **12**, 213-219, 2011.

[9] O.A. Averkova, I.N. Logachev, K.I. Logachev, Modeling of flow separation at the inlet of a suction channel in regions with cuts. *Computational methods and programming*, **13**, 298-306, 2012.

[10] O.A. Averkova, I.N. Logachev, K.I. Logachev, *Separation flows in the exhaust ducts spectra*. Moscow-Izhevsk: IKI, 2012.

[11] I.N. Logachev, K.I. Logachev, O.A. Averkova. *Local Exhaust Ventilation: Aerodynamic Processes and Calculations of Dust Emissions*. Boca Raton: CRC Press, 2015.

[12] P.N. Kamenev. *Heating and ventilation. Part 2. Ventilation*. Moscow: State publishing house of building, architecture and building materials literature. - 1959.

[13] T. A. Fialkovskaya, *Exhaust hoods and fume chambers*. Moscow: Stroyizdat, 1947.

[14] V. V. Baturin, The fundamentals of industrial ventilation. Moscow: Profizdat, 1990.

[15] E.A. Shtokman, V.A. Shilov, E.E. Novgorodsky, I.I. Savvidi, T.A. Skorik, V.V. Pashkov, *Ventilation, conditioning and air purification at food industry enterprises. Ed. by E.A.Shtokman*. Moscow: ASV publishing house, 2001.

[16] V. N. Posokhin. *Ventilation aerodynamics*. Moscow: AVOK-PRESS, 2008.

[17] I.N. Logachev, K.I. Logachev, O.A. Averkova. *Local Exhaust Ventilation: Aerodynamic Processes and Calculations of Dust Emissions*. Boca Raton: CRC Press, 2015.

[18] I.E. Idelchik. *Hydraulic resistance guide book.* Moscow: Mashinostroeniye, 1977.

[19] V.G. Shaptala, V.V. Shaptala, A.V. Gavrilenko, Numerical simulation of cement-air mixture in the pneumochamber pump. *Bulletin of BSTU named after V.G. Shukhov*, **2**, 159-161, 2015.

[20] *Special functions reference book*. ed. by M.Abramovits and I.Stigan. M.:Nauka, 1979.

[21] V.V. Shaptala, Modeling and calculation systems of cleaning dust emissions of cement production. *Bulletin of BSTU named after V.G. Shukhov*, 6, 205-209, 2015.

[22] V.G. Shaptala, V.V. Shaptala, The method of calculating efficiency of centrifugal collectors sticky dust in the production of building materials. *Bulletin of BSTU named after V.G. Shukhov*, 3, 58-62, 2014.

[23] V.A.Minko, *Dust control of construction materials production processes*. Voronezh: VSU Publishing, 1981.

[24] I.V. Khodakov, Modeling of detached flow at the entry polygonal suction hole. *Bulletin of BSTU named after V.G. Shukhov*, 2, 10-14, 2016.

[25] I.V. Khodakov, Numerical and experimental research of detached flow at the inlet in the suction duct with the mechanical screen. *Bulletin of BSTU named after V.G. Shukhov*, 3, 6-12, 2016.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

RECOVERY OF DIFFERENTIATION/INTEGRATION COMPATIBILITY OF MESHLESS OPERATORS VIA LOCAL ADAPTATION OF THE POINT CLOUD IN THE CONTEXT OF NODAL INTEGRATION

Gabriel Fougeron[1,2], Guillaume Pierrot[1] and Denis Aubry[2]

[1]ESI-Group
99 Rue des Solets, Parc d'affaires SILIC, 94513 RUNGIS CEDEX, FRANCE
e-mail: {gabriel.fougeron,guillaume.pierrot}@esi-group.com

[2]CentraleSupélec Laboratoire MSSMat - UMR CNRS 8579
Grande Voie des Vignes 92290 CHATENAY-MALABRY, FRANCE
e-mail: denis.aubry@ecp.fr

Keywords: Meshless methods, Nodal integration, SPH, Advection, Diffusion.

Abstract. *Standard SPH methods are very attractive to simulate complex flow problems thanks to their robustness and ease of implementation, but often suffer from poor precision regarding point-wise stress computation or stability issues characterized by a loss of quality of the point cloud (particle clumping, formation of voids).*

In this paper, we recall an operator-based framework for the description of meshless discretizations using nodal integration. Using this framework, we are able to define a dual gradient inspired by the integration by parts formula, clarify its role in the discretization of a simple diffusion problem and specify sufficient conditions to satisfy the patch test. For symmetric discretizations, we show that one of these conditions reduces to a discrete version of Stokes' theorem, which we call differentiation/integration compatibility of meshless operators.

Our earlier work [22] focused on the recovery of these compatibility conditions via the modification of gradient coefficients. In a companion presentation [23], we extend the analysis to the concept of element-based integration and investigate the possibility of a one-shot solution algorithm, jointly solving for both compatibility and pressure.

The present work however is an attempt at achieving compatibility while keeping the computational simplicity of SPH. Instead of solving a global linear system for compatibility, the position of SPH nodes are adapted so that compatibility conditions are naturally enforced using classical and well-known SPH gradients.

In the context of Lagrangian simulation, this means that the SPH nodes are not advected with the physical velocity, but with a corrected velocity. The idea is not new: it can be traced back to Monaghan and his XSPH formulation [19], and appeared more recently in the context of fluid dynamics with the works of Adami, Hu and Adams ([1] and [16] for instance). In this work, we try to incorporate their corrections to our framework, generalize their method and give a novel interpretation for their results.

1 INTRODUCTION

Smoothed Particle Hydrodynamics (SPH) is a family of meshless simulation methods initiated in 1977 with the independent contributions of Lucy ([17]) and Gingold and Monaghan ([10]). SPH was initially applied to astrophysical problems and still enjoys success in this field (see [25] for a recent review), but its scope of applications soon widened to cover very complex flow problems ([11] for fluid-structure interaction for instance, [2] regarding chemically reacting viscous flows), elasticity and plasticity problems ([18] for instance), as well as complex magneto-hydrodynamic problems ([24]). Thanks to its robustness, SPH seems ideally suited to give an answer to geometrically or physically complex problems, where all other methods usually fail. Comparatively little work has been dedicated to solve well established simple test-cases with analytical solutions (let us cite [13] and [3] as exceptions).

However, SPH has long suffered from stability issues characterized by local loss of quality of the point cloud (particle clumping, formations of voids. See for instance [26] where the infamous term "tensile instability" originated) Unsurprisingly, the main effort in the SPH community has been to propose several remedies to control the apparition of instability (see [19] with the XSPH formulation and [20] introducing artificial stress formulation). More recently, significant improvements to stability as well as point-wise accuracy have been achieved with Adami, Hu and Adams' "transport-velocity" formulation (see [1] and [16]).

In previous works ([21] and [22]), the present authors have investigated the influence of the properties of discrete meshless differentiation operators (or meshless "gradients") on the accuracy of the solution of a simple diffusion problem in the context of nodal integration. In section 2, we very briefly recall the symmetric meshless discretization of the diffusion problem and derive sufficient conditions to pass the patch test, which we call "compatibility conditions". Our earlier work [22] focused on the recovery of these conditions via the modification of gradient coefficients and the introduction of a corrected gradient at the cost of solving a global linear system. Section 2.6 however is an attempt to achieve compatibility while keeping the computational simplicity of SPH: Instead of solving a global linear system for compatibility, the position of SPH nodes are adapted so that compatibility conditions are naturally enforced using classical and well-known SPH gradients. We show that this technique yields second order convergence for the diffusion problem.

Finally, we revisit the contributions of Adami, Hu and Adams and propose a slightly different procedure to limit the inconsistencies of the classical SPH gradient. This modification allows an easier interpretation and exploration of its consequences, which we undertake in section 3.

2 DISCRETIZATION OF THE DIFFUSION PROBLEM

2.1 The Meshless Framework

In SPH, every node i of the point cloud $\mathcal{C}$ (including those of the "boundary" cloud $\partial\mathcal{C}$) located at coordinates $\mathbf{x}_i$ is attributed a positive volume usually denoted V_i. These volumes play the role of weights to define a discrete scalar product between fields $f, g : \mathcal{C} \to \mathbb{R}$, approximating the classical L^2 scalar product between functions as:

$$(f, g)_{\mathcal{C}} = \sum_{i \in \mathcal{C}} V_i f_i g_i \tag{1}$$

A discrete meshless differential operator, or "meshless gradient" is a linear operator $\mathbb{\nabla} : (\mathcal{C} \to \mathbb{R}) \to (\mathcal{C} \to \mathbb{R}^d)$ that maps a scalar field $f : \mathcal{C} \to \mathbb{R}$ to a vector field $\mathbb{\nabla} f : \mathcal{C} \to \mathbb{R}^d$ (d is

the number of space dimensions). Such an operator can always be written at point i as:

$$V_i \mathbb{\nabla}_i f = \sum_{j \in \mathcal{C}} \mathbf{A}_{i,j} f_j \tag{2}$$

where the $\mathbf{A}_{i,j}$ are vector coefficients (the coefficient V_i was added for convenience, see equation 4). There is *a priori* (and in general) no relation between $\mathbf{A}_{i,j}$ and $\mathbf{A}_{j,i}$.

A meshless gradient is said to have n-th order consistency if it is exact on all monomials of order at most n. Mimicking integration by parts theorem, given a "primal" meshless gradient $\mathbb{\nabla}$, we can introduce a "dual" meshless gradient $\mathbb{\nabla}^*$, as:

$$\forall f, g : \mathcal{C} \to \mathbb{R} \mid f_{\partial \mathcal{C}} = 0, \quad (\mathbb{\nabla} f, g)_{\mathcal{C}} + (f, \mathbb{\nabla}^* g)_{\mathcal{C}} = 0 \tag{3}$$

Remark: We only consider functions that are identically null on the border $\partial\mathcal{C}$ not to have to include a boundary term in the right-hand side of equation 3. Boundary nodes will thus need to receive a special treatment in what follows.

From equation 3, we can easily derive that the coefficients $\mathbf{A}^*_{i,j}$ of the dual gradient will in general read:

$$\forall i, j \in \mathcal{C} \quad \mathbf{A}^*_{i,j} = -\mathbf{A}_{j,i} \tag{4}$$

The (skew-)symmetry of this formula entails $\mathbb{\nabla}^{**} = \mathbb{\nabla}$. This underlines the fact that $\mathbb{\nabla}$ and $\mathbb{\nabla}^*$ are actually dual to each other. There is *a priori* no reason to prefer one gradient over the other.

2.2 Classical examples of meshless gradients

Given a kernel function W_h, let us list three well-known meshless gradient from the SPH literature (see [14]) as well as their dual counterparts:

- The "classical" SPH gradient, which is equal to its dual because of the symmetry properties of the kernel function:

$$\mathbb{\nabla}^{\mathrm{CLASS}}_i f = -\sum_{j \in \mathcal{C}} V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_i) f_j \tag{5}$$

- The renormalized gradient of order 0, which is zeroth-order consistent:

$$\mathbb{\nabla}^{\mathrm{R0}}_i f = -\sum_{j \in \mathcal{C}} V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_i)(f_j - f_i) \tag{6}$$

  Its dual gradient can be written as:

$$\mathbb{\nabla}^{\mathrm{R0}*}_i f = -\sum_{j \in \mathcal{C}} V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_i)(f_j + f_i) \tag{7}$$

  In particular, the dual renormalized 0 gradient of a constant scalar field is twice the classical gradient of this constant : $\mathbb{\nabla}^{\mathrm{R0}*} 1 = 2\mathbb{\nabla}^{\mathrm{CLASS}} 1$.

- The renormalized gradient of order 1:

$$\mathbb{\nabla}^{\mathrm{R1}}_i f = -\sum_{j \in \mathcal{C}} V_j \mathbb{B}_i \nabla W_h(\mathbf{x}_j - \mathbf{x}_i)(f_j - f_i) \tag{8}$$

Where $\mathbb{B}_i$ is the $d \times d$ matrix set as to recover first order consistency:

$$\mathbb{B}_i^{-1} = \mathbb{\nabla}_i^{\mathrm{R0}}\mathbf{x} = -\sum_{j \in \mathcal{C}} V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_i) \otimes (\mathbf{x}_j - \mathbf{x}_i) \tag{9}$$

In the above formula, $\otimes$ denotes the classical tensor product. The dual gradient $\mathbb{\nabla}_i^{\mathrm{R1}*}$ can be written:

$$\mathbb{\nabla}_i^{\mathrm{R1}*} f = -\sum_{j \in \mathcal{C}} V_j (\mathbb{B}_i f_i + \mathbb{B}_j f_j) \nabla W_h(\mathbf{x}_j - \mathbf{x}_i) \tag{10}$$

2.3 Strong discretization of the diffusion problem

In the following sections, regularity conditions which should be put on continuous fields are knowingly ignored because they are irrelevant to the discretization procedure.
Let us consider the following strong form of the diffusion problem with Dirichlet boundary conditions on $\Omega \subset \mathbb{R}^d$:

$$\begin{cases} -\Delta u = s \\ u|_{\partial\Omega} = g \end{cases} \tag{11}$$

If we bluntly replace continuous operators with discrete ones, we find the following strong SPH discretization:

$$\begin{cases} -\mathbb{\nabla}_i \cdot \mathbb{\nabla} u = s_i & \forall i \in \mathcal{C} \\ u_i = g_i & \forall i \in \partial\mathcal{C} \end{cases} \tag{12}$$

Note that this is not the only possible meshless discretization of the strong form given in equation 11. Equation 12 defines the following strong discrete Laplace operator:

$$(-\mathbb{\Delta})_{i,j} = (-\mathbb{\nabla} \cdot \mathbb{\nabla})_{i,j} = -\sum_{k \in \mathcal{C}} \frac{\mathbf{A}_{i,k} \cdot \mathbf{A}_{k,j}}{V_i V_k} \tag{13}$$

This operator is not symmetric in general, has been observed to yield nearly singular systems, and thus would need stabilization techniques in order to give satisfactory results. A sufficient condition for the discrete strong laplacian of a linear function to identically vanish (which is a sufficient condition for the patch test) is the first order consistency of the meshless gradient $\mathbb{\nabla}$.

2.4 Weak discretization of the diffusion problem

The corresponding weak form of the diffusion problem can be written as follows: $\forall v : \Omega \to \mathbb{R} \mid v|_{\partial\Omega} = 0$,

$$\begin{cases} \displaystyle\int_\Omega \nabla u \cdot \nabla v = \int_\Omega sv \\ u|_{\partial\Omega} = g \end{cases} \tag{14}$$

Again bluntly replacing continuous operators with discrete ones, we find the following weak SPH discretization: $\forall v : \mathcal{C} \to \mathbb{R} \mid v|_{\partial\mathcal{C}} = 0$,

$$\begin{cases} (\mathbb{\nabla} u, \mathbb{\nabla} v)_{\mathcal{C}} = (s, v)_{\mathcal{C}} \\ u|_{\partial\Omega} = g \end{cases} \tag{15}$$

The coefficients of the corresponding (self-adjoint !) weak laplacian discrete operator then read:

$$(-\mathbb{\Delta})_{i,j} = (-\mathbb{\nabla}^* \cdot \mathbb{\nabla})_{i,j} = \sum_{k \in \mathcal{C}} \frac{\mathbf{A}_{k,i} \cdot \mathbf{A}_{k,j}}{V_i V_k} \tag{16}$$

Sufficient conditions for the discrete weak laplacian of a linear function to be null (which is a necessary condition for the patch test) are the first order consistency of the primal gradient $\mathbb{\nabla}$ **and** the zeroth-order consistency of the dual gradient $\mathbb{\nabla}^*$. This second condition, which we call compatibility is paramount: it involves both discrete integration and discrete differentiation and can be re-stated (via equation 3) as a discrete version of Stokes' theorem. None of the gradients described in section 2.2 naturally enforces this condition on general clouds of points. In the meshless literature, the notion of compatibility appeared in the work of Bonet (see [3]), where compatibility is sought by means of kernel correction. The second notable contribution is Chiu's in [4]. However, Chiu does not discriminate between primal and dual gradients since he *a priori* enforces skew-symmetry of his gradient coefficients.

Our previous work [22] showed that using a corrected version of the renormalized 1 gradient, it is possible to experience second order convergence rate for the diffusion problem with sources. More importantly, it also suggested that exact compatibility is not strictly necessary for second-order convergence: it seems sufficient to bound the term $\|\mathbb{\nabla}^* 1\|$ with a constant as the point cloud is refined instead of the worst-case $\|\mathbb{\nabla}^* 1\| = \mathcal{O}(h^{-1})$ (Here, h denotes a typical inter-particular distance in the point cloud. It is chosen proportional to the maximum distance $\|\mathbf{x}_j - \mathbf{x}_i\|$ for which either $\mathbf{A}_{i,j}$ or $\mathbf{A}_{j,i}$ is non-null). We call these gradients "quasi-compatible".

In section 2.6, we will once again demonstrate that using a first order consistent and quasi-compatible gradient yields ultimate second order convergence for the diffusion problem.

2.5 Another symmetric discretization of the diffusion problem

Inspired by finite volume discretizations on orthogonal meshes, other discretizations of the diffusion operator have also been proposed in the meshless literature (see [7] for a review SPH discretizations of the diffusion operator). The details vary, but the most promising (according to [7]) reads in coefficients:

$$(-\mathbb{\Delta})_{i,j} = 2V_j \frac{\nabla W_h(\mathbf{x}_j - \mathbf{x}_i) \cdot (\mathbf{x}_j - \mathbf{x}_i)}{\|\mathbf{x}_j - \mathbf{x}_i\|^2} \quad \forall i \neq j \in \mathcal{C} \tag{17}$$

This operator is self-adjoint and the related bi-linear form reads:

$$\langle f, g \rangle_{\mathcal{C}} = \sum_{i,j \in \mathcal{C}} V_i V_j \frac{\nabla W_h(\mathbf{x}_j - \mathbf{x}_i) \cdot (\mathbf{x}_j - \mathbf{x}_i)}{\|\mathbf{x}_j - \mathbf{x}_i\|^2} (f_j - f_i)(g_j - g_i) \tag{18}$$

The conditions for the patch test can once again be restated as compatibility conditions. Indeed in the case of the above laplacian (using a kernel with radial symmetry), a sufficient condition for the laplacian of a linear field to vanish is that the classical (dual) gradient of the unit field $\mathbb{\nabla}^{\mathrm{CLASS}} 1$ vanishes.

2.6 Looking for quasi-compatible configurations and results

The generation of point clouds well-suited for meshless simulations is somehow similar to the meshing procedure in mesh-based methods: it should not be performed regardless of the numerical method that follows. Generating point cloud has recently been an active area of

research in the computer graphics community, where harmonious yet unstructured point clouds are sought. Several methods to build these "blue noise" configuration bear striking similarities with the method exposed here. See [8], [12] and [5] for more details.

In [1] and [16], Adami, Hu and Adams add a "background pressure" source term to their momentum equation as the gradient of a constant field. This term was demonstrated to have a regularizing effect on the point cloud, which gave better approximation properties to the underlying kernel estimation of (constant in their incompressible setting) density.
In order to better understand this source term, let us define the following function of all the node coordinates:

$$\mathcal{V} = \sum_{i,j \in \mathcal{C}} V_i V_j W_h(\mathbf{x}_j - \mathbf{x}_i) \tag{19}$$

Let us remark that $\mathcal{V}$ is non-negative and thus admits at least a minimum on every compact.

The gradient of $\mathcal{V}$ with respect to $\mathbf{x}_k$ can be written:

$$\frac{\partial \mathcal{V}}{\partial \mathbf{x}_k} = -2V_k \sum_{j \in \mathcal{C}} V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_k) = V_k \mathbb{\nabla}^{R0*} 1 \tag{20}$$

Hence, this constant pressure source term (eq. 17 of [16]) can be re-written using our notations:

$$2 \sum_j V_i V_j \nabla W_h(\mathbf{x}_j - \mathbf{x}_i) P_c = -V_i \mathbb{\nabla}_i^{R0*} P_c = -P_c \frac{\partial \mathcal{V}}{\partial \mathbf{x}_i} \tag{21}$$

Since their second term is dissipative (it is the discretization of viscosity forces), their procedure can be thought of as a generalized gradient descent trying to minimize $\mathcal{V}$. They remain unclear concerning the boundary condition that were used, but we can safely assume that there were either fixed nodes surrounding the discretized domain or that periodic boundary conditions were used. In any case, the $\mathbf{x}_j$ are bound to rest in a compact region, and a local minimum of $\mathcal{V}$ is ultimately reached. This local minimum has zero gradient, hence it defines a point cloud where the discrete dual meshless gradient of a constant function is null. With such node positions, the classical gradient is compatible (and consistent, and coincides with the renormalized 0 gradient).

Similarly, we generated several clouds of points of different sizes using a gradient descent algorithm. The algorithm was stopped when the norm of the gradient of a unit field reached a given threshold. We call this threshold "Stokes' default" since it is a measure of how much Stokes' formula fails to be satisfied. We also tried the same procedure using the renormalized 1 gradient. Even if we could not find any function similar to $\mathcal{V}$ which partial derivatives with respect to positions give the dual gradient of a constant function, we still have been able to reduce $\|\mathbb{\nabla}^* 1\|$ by three orders of magnitude or more using this technique.

In our tests, the nodes were initially distributed as a low discrepancy Halton sequence (see [9] for details about low discrepancy sequences), and two layers of boundary nodes ensured that no particle escaped the simulation domain. These particles were later used to enforce Dirichlet boundary conditions for the diffusion equation. Node volumes were chosen all equal, and so that their sum reproduced the volume of the simulation domain. This is in agreement with the Quasi-Monte Carlo philosophy associated with low discrepancy sequences. Typical node arrangements are displayed on figures 1 to 4.

Concerning the diffusion equations, the source s and the Dirichlet border conditions are set so that an analytical solution is known. Thus, it was possible to perform an h-convergence analysis for each presented gradient. The corresponding curves are shown in figures 7 - 9. On

each curve, the Stokes default is kept below the following thresholds: 100, 50, 30, 10, 5, 3, 1, 0.5 and 0.3. It should be clear from these curves that enforcing compatibility conditions is a good means to improve accuracy of diffusion simulations. Indeed, the lower the Stokes default, the best the accuracy. With poor compatibility, very poor accuracy was consistently obtained (error $\geq 100\%$) for all gradients.
It can also be noted that increasing the order of consistency of the primal gradient was always beneficial to accuracy. Finally let us remark that only the first order consistent gradient produced straight-looking accuracy curves. A linear least-squares fit of the three last curves (Stokes default 1, 0.5 and 0.3) gave approximate order of convergence of 2.03, 2.08 and 2.10. This is why we claim we achieve ultimate second order convergence if we keep the Stokes default under a low enough threshold.

3 APPROXIMATE ADVECTION AND COMPATIBILITY

In the context of Lagrangian simulation, the point cloud is evolving at every time step. In this section, we will show on a concrete example that exact advection of the nodes along the physical velocity field is not a desirable feature of a SPH simulation. The idea is not new, and several authors have proposed regularized velocity fields better suited for advection ([19] and his XSPH formulation is probably the most famous). To a large extend, artificial viscosity can also be regarded as an additional source in the momentum equation.

It is often not clear how the addition of those regularizing source terms impact the kinematics of the fluid. This section quantatively explores the implications of the use of an ALE velocity on the mean level of Stokes default.

3.1 An ALE source term

Corrective procedures proposed in [19] (XSPH), [1] (transport-velocity) and [15] (Fick-diffusion based regularization) are difficult to analyze because they are directly formulated at the time-discretized level. Hence, time discretization and space discretization are already coupled, and it is quite often not clear how they really interact. This is the reason why we propose an Arbitrary Lagrangian Eulerian (see [6] for a good introduction to ALE formulations) velocity formulation independent of the time discretization. In the results that we show, high order time discretization as well as low time steps were used so that time discretization errors are completely negligible.

Let us suppose that the space discretization is initially very good, and that the computed velocity field $\mathbf{v}(\mathbf{x})$ is extremely close to the physical velocity. In order to go to the next time-step, the simulation procedure should carry the SPH nodes along the streamlines of the velocity for the duration of a time-step. Figure 10 shows the advection of initially regularly arranged nodes after advection with a Taylor-Green flow (which is a solution of the incompressible Navier-Stokes equations). Even if the flow is extremely regular (in our example it is smooth and its divergence is zero, thus not creating voids or excess of fluid in any area at the continuous level), we can see that, at a discrete level, holes have been created in some areas, and nodes have clumped in others. Concurrently, the Stokes default (as plotted on figure 15) quickly rises from a null value (the cartesian arrangement of nodes is compatible), to a size-dependant small range of values (40-50 in the example). At mid-time, the flow velocity is reversed and the flow travels back towards the initial state. Symmetry of figure 15 demonstrates the time-reversibility of the approach.

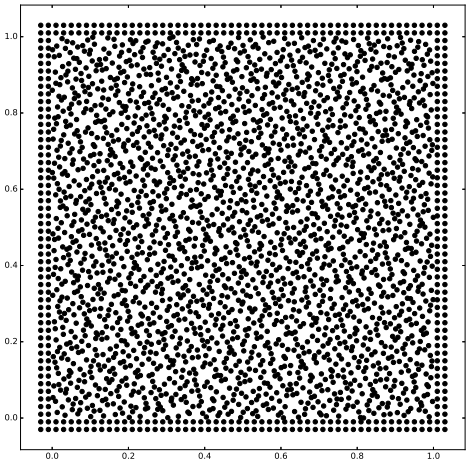


Figure 1: Halton sequence of 2500 nodes
$\|\mathbb{\nabla}^{\mathrm{CLASS}*}1\| \approx 14.5$
$\|\mathbb{\nabla}^{\mathrm{R0}*}1\| \approx 29.1$
$\|\mathbb{\nabla}^{\mathrm{R1}*}1\| \approx 36.3$

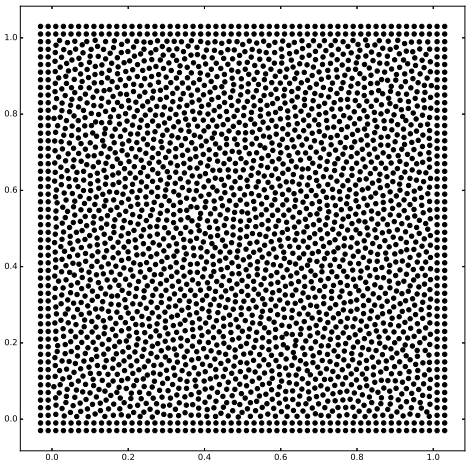


Figure 2: Modified point cloud.
$\|\mathbb{\nabla}^{\mathrm{CLASS}*}1\| \approx 1$

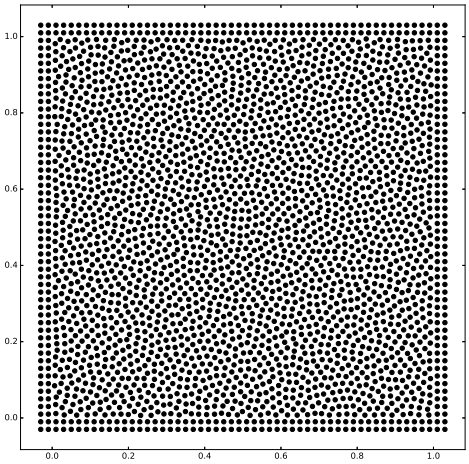


Figure 3: Modified point cloud.
$\|\mathbb{\nabla}^{\mathrm{R0}*}1\| \approx 1$

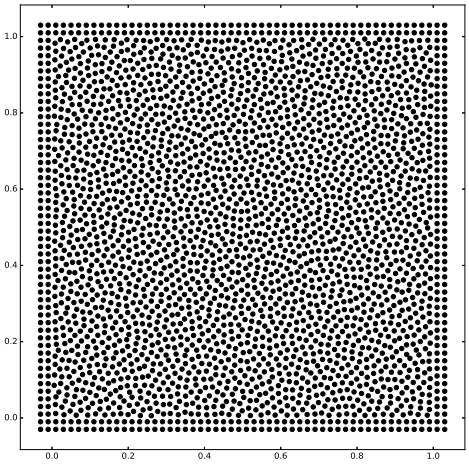


Figure 4: Modified point cloud.
$\|\mathbb{\nabla}^{\mathrm{R1}*}1\| \approx 1$

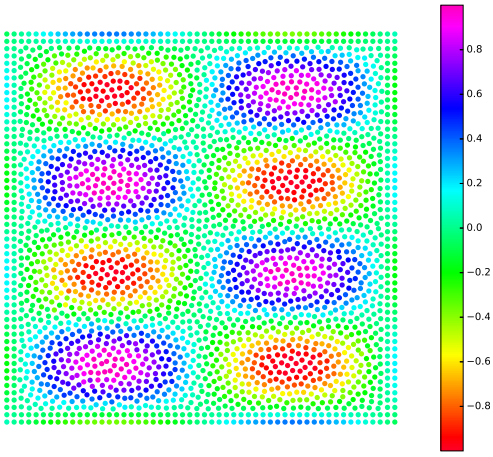


Figure 5: SPH solution using the renormalized 1 gradient

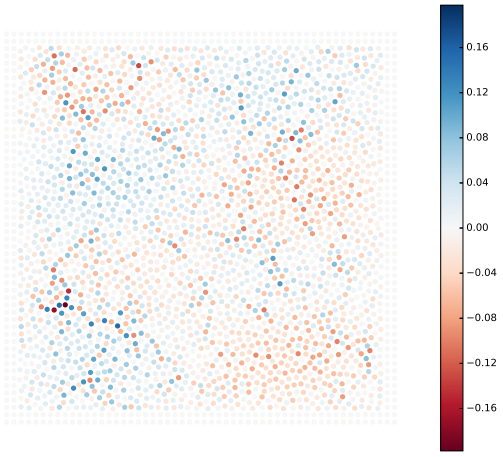


Figure 6: Signed error to exact solution

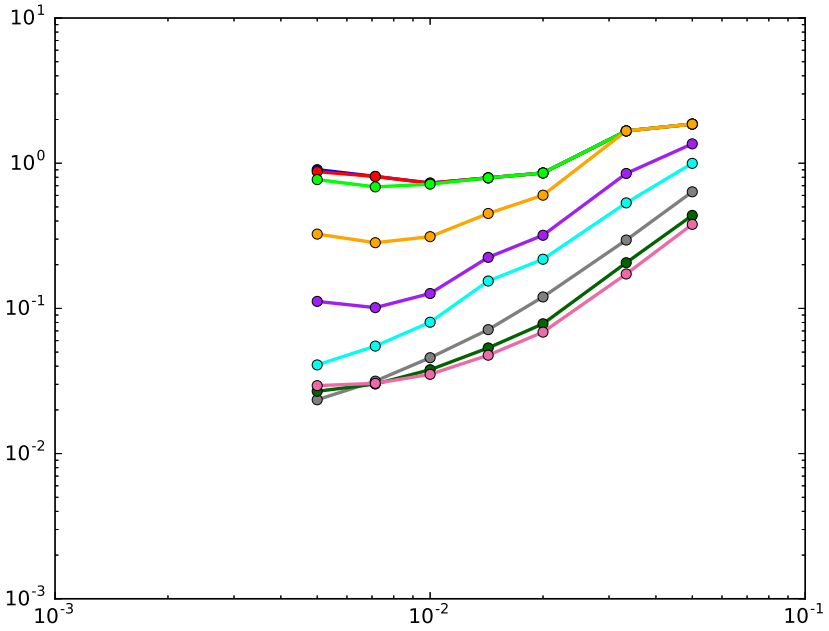


Figure 7: Relative error of the laplacian system as a function of discretization length for the classical gradient with Stokes default kept below 100, 50, 30, 10, 5, 3, 1, 0.5 and 0.3

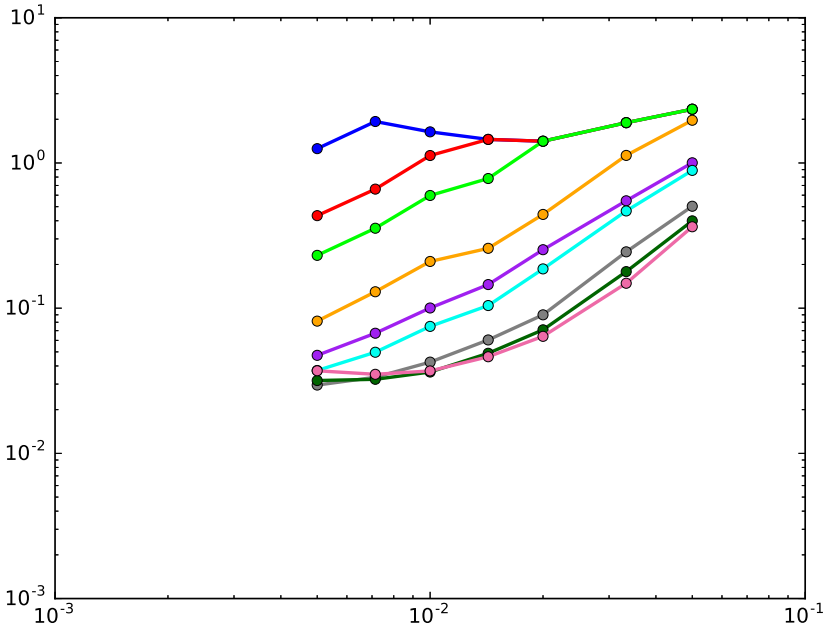


Figure 8: Relative error of the laplacian system as a function of discretization length for the renormalized 0 gradient with Stokes default kept below 100, 50, 30, 10, 5, 3, 1, 0.5 and 0.3

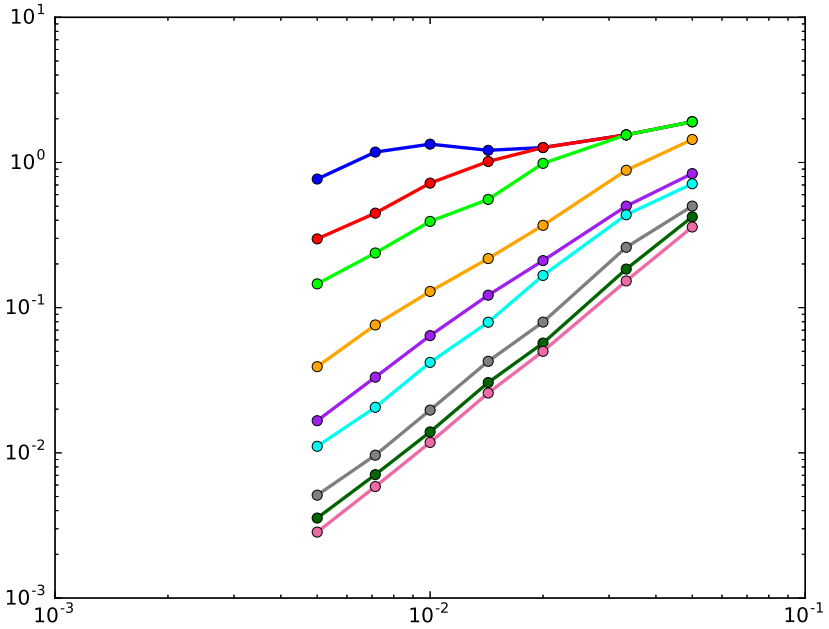


Figure 9: Relative error of the laplacian system as a function of discretization length for the renormalized 1 gradient with Stokes default kept below 100, 50, 30, 10, 5, 3, 1, 0.5 and 0.3.

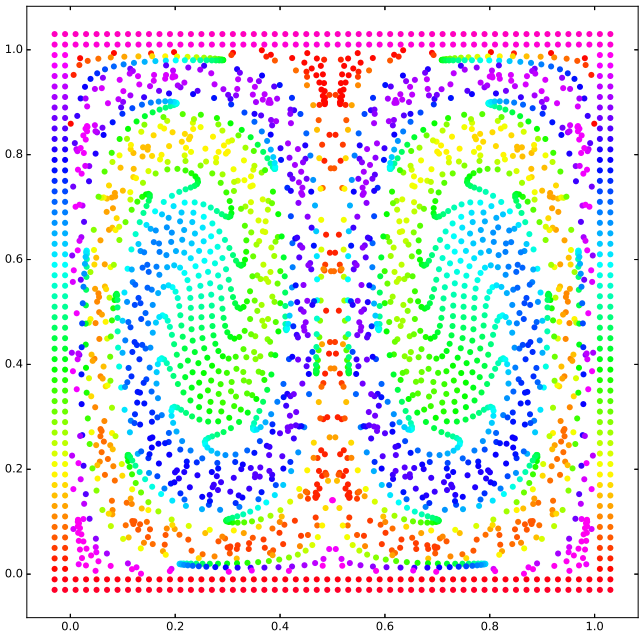


Figure 10: Advection of initially regular cloud of 2500 nodes by the Taylor-Green vortex flow. The color of each particle is set as a function of is initial y-coordinate. The two outer layers of nodes were kept fixed

In order to create figure 10, the following uncoupled system of ODEs was solved:

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = \mathbf{v}(\mathbf{x}_i) \quad \forall i \in \mathcal{C} \tag{22}$$

While equation 22 defines the solution of the exact node advection, we could instead seek a "close" solution to the advection problem while keeping a better mesh quality (i.e. lower Stokes default). For instance, we could solve the following system (where $\mathbf{c}_i$ is a "small" velocity field designed so as to keep the Stokes default under a given threshold).

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = \mathbf{v}(\mathbf{x}_i) - \mathbf{c}_i \quad \forall i \in \mathcal{C} \tag{23}$$

In 23, $\mathbf{c}_i$ can immediately be interpreted as the ALE velocity defined in [6] in the context of point cloud advection (as opposed to mesh advection).
Inspired by the quasi-gradient descent method of [16], we propose the following easy to compute expression for $\mathbf{c}_i$ (where γ is a user-provided constant):

$$\mathbf{c}_i = \gamma \mathbb{\nabla}_i^* 1 \tag{24}$$

All the presented results used the classical SPH gradient $\mathbb{\nabla}^{\mathrm{CLASS}}$, but extremely similar results were obtained with the renormalized 1 gradient $\mathbb{\nabla}^{\mathrm{R1}}$. Figure 17 shows the time evolution of the mean Stokes default as a function of time when integrating system 23 - 24 with several values of the parameter γ. The velocity field $\mathbf{v}$ is again chosen to be the Taylor-Green vortex flow (reversed after $t = 50$). We can see that the value of the mean Stokes default quickly stabilizes at a near-constant level. This level depends on the total number of simulation nodes as well as the parameter γ as pictured in figures 18 - 20.

Least-square fitting in figures 18 and 20 reveal the following approximate dependencies: $\|\mathbb{\nabla}^{\mathrm{CLASS}*}1\| \propto \gamma^{-0.65} h^{0.35}$ when $\gamma \neq 0$, whereas the ALE speed grows as: $\|\mathrm{c}_i\| = \gamma \|\mathbb{\nabla}^{\mathrm{CLASS}*}1\| \propto \gamma^{0.35} h^{0.35}$. Of course, these exponents are probably case-dependent, but they still seem to indicate that one could choose a value $\gamma = \mathcal{O}(\sqrt{h})$ in order to keep $\|\mathbb{\nabla}^{\mathrm{CLASS}*}1\| = \mathcal{O}(1)$. This choice would ensure that the advection velocity goes to zero as the cloud is refined as $\|\mathrm{c}_i\| = \mathcal{O}(\sqrt{h})$, so that the method is asymptotically Lagrangian instead of fully ALE. Of course, other compromises could also be made.

As a final remark, let us note that using formulation 23 - 24 with gradient $\mathbb{\nabla}^{\mathrm{CLASS}*}$ lowers the value of $\|\mathbb{\nabla}^{\mathrm{R1}*}\|$ for a given value of the parameter γ, but this value still increases as h increases as shown in figure 19. This fact slightly tempers the good results presented in [15].

4 CONCLUSIONS AND FUTURE WORK

In section 2.1 we presented a meshless framework for the discretization of partial differential equations and gave the example of the diffusion equation. In section 2.4, we studied the patch test using a symmetric discretization of the diffusion equation underlining the key role of the so-called compatibility conditions, a discrete counterpart of Stokes' formula. In section 2.6 , we showed that good accuracy of the solution to the discrete laplacian problem could not be reached unless the compatibility conditions were met at least approximately.

In section 3 we linked previously proposed stabilization techniques with the recovery of a quasi-compatible gradient. We also proposed a new procedure to enforce approximate compatibility in ALE simulations.

In future works, we would like to investigate the role of the discretization of the boundary (meshed boundaries, ghost particles, ...) on the ALE formulation we proposed and extend the

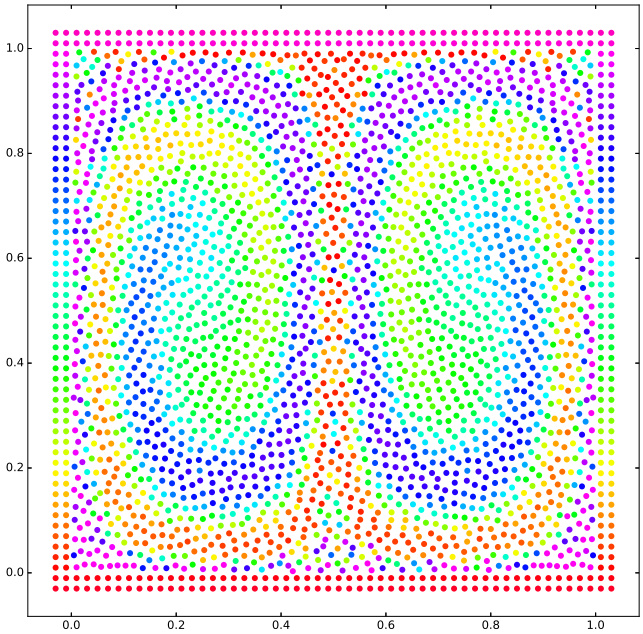


Figure 11: Advection of initially regular cloud of 2500 nodes by the Taylor-Green vortex flow using formulation 23 - 24 with $\gamma = 0.003$ at time $t = 50$ (mid-time). The color of each particle is set as a function of is initial y-coordinate. The two outer layers of nodes were kept fixed

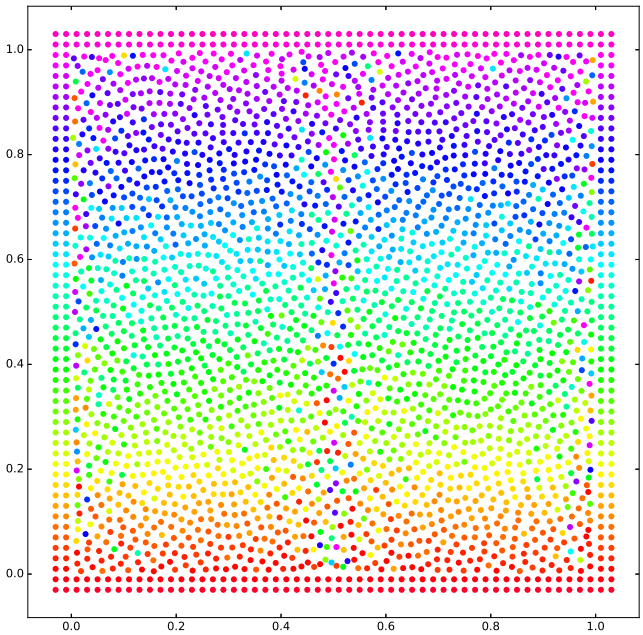


Figure 12: Advection of initially regular cloud of 2500 nodes by the Taylor-Green vortex flow using formulation 23 - 24 with $\gamma = 0.003$ at time $t = 100$ (end-time). The color of each particle is set as a function of is initial y-coordinate. The two outer layers of nodes were kept fixed. The error compared to exact avection can be visually appreciated with the non-uniformity of the color field.

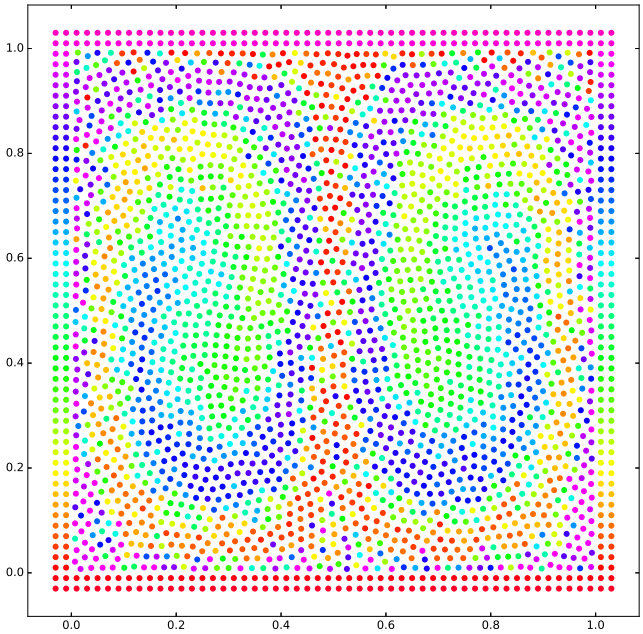


Figure 13: Advection of initially regular cloud of 2500 nodes by the Taylor-Green vortex flow using formulation 23 - 24 with $\gamma = 0.1$ at time $t = 50$ (mid-time). The color of each particle is set as a function of is initial y-coordinate. The two outer layers of nodes were kept fixed

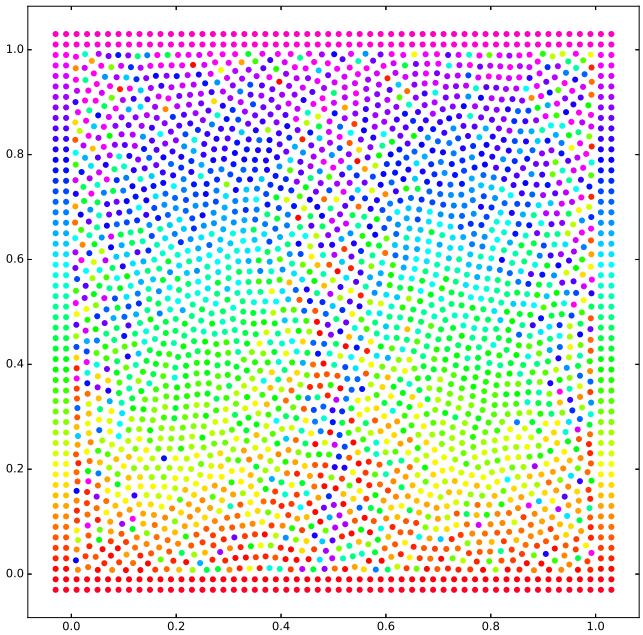


Figure 14: Advection of initially regular cloud of 2500 nodes by the Taylor-Green vortex flow using formulation 23 - 24 with $\gamma = 0.1$ at time $t = 100$ (end-time). The color of each particle is set as a function of is initial y-coordinate. The two outer layers of nodes were kept fixed. The error compared to exact avection can be visually appreciated with the non-uniformity of the color field.

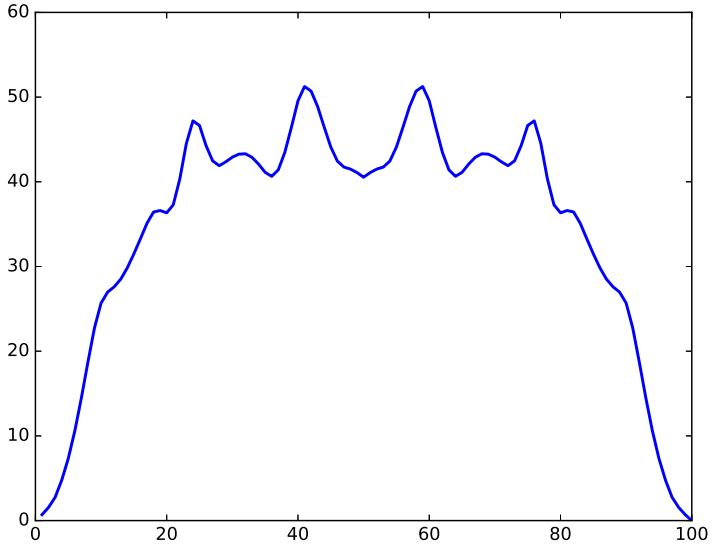


Figure 15: Mean Stokes default of an initially regular cloud of 2500 nodes advected by the Taylor-Green vortex flow as a function of time. At time $t = 50$, the flow is reversed (hence the symmetry of the curve)

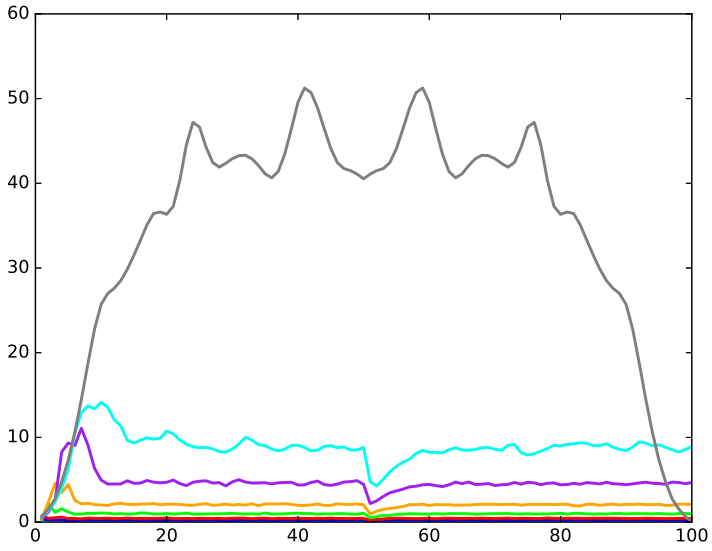


Figure 16: Mean Stokes default of an initially regular cloud of 2500 nodes advected by the Taylor-Green flow, and relaxed by formulations 23 - 24 as a function of time. At time $t = 50$, the Taylor-Green flow is reversed, but the ODE system is no longer time-reversible. The values of γ used are: 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 and 0.

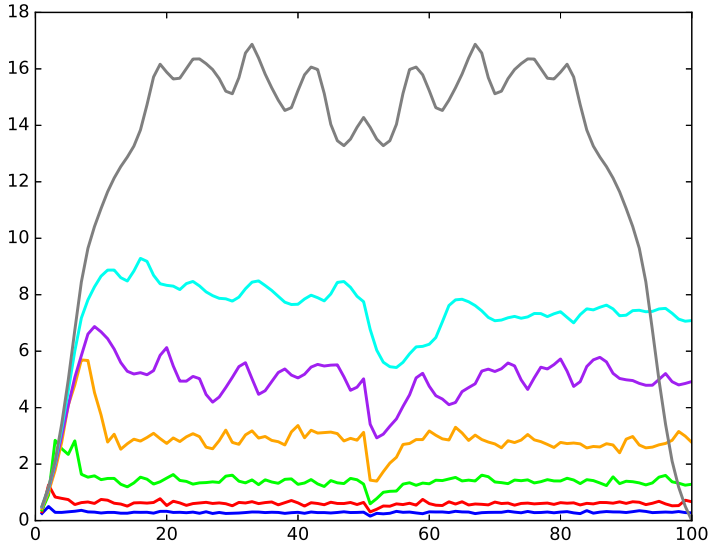


Figure 17: Mean Stokes default of an initially regular cloud of 400 nodes advected by the Taylor-Green flow, and relaxed by formulations 23 - 24 as a function of time. At time $t = 50$, the Taylor-Green flow is reversed, but the ODE system is no longer time-reversible. The values of γ used are: 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 and 0.

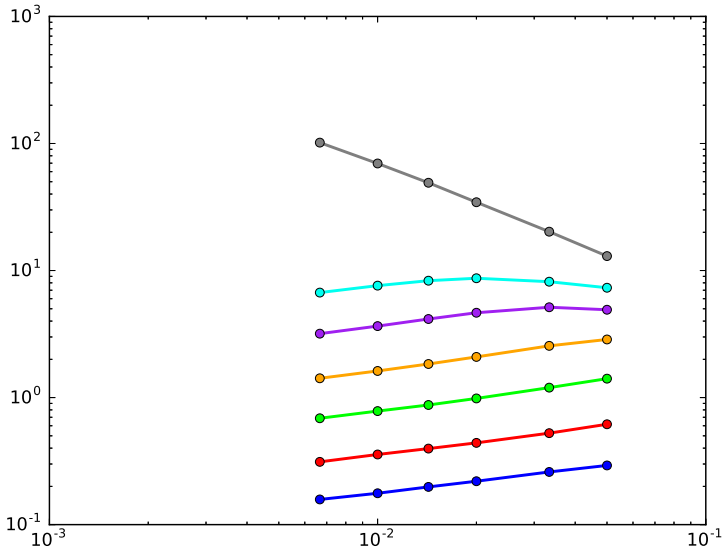


Figure 18: Averaged mean Stokes default as a function of discretization length in the case of the Taylor-Green vortex flow relaxed by formulation 23 - 24. The values of γ used are: 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 and 0.

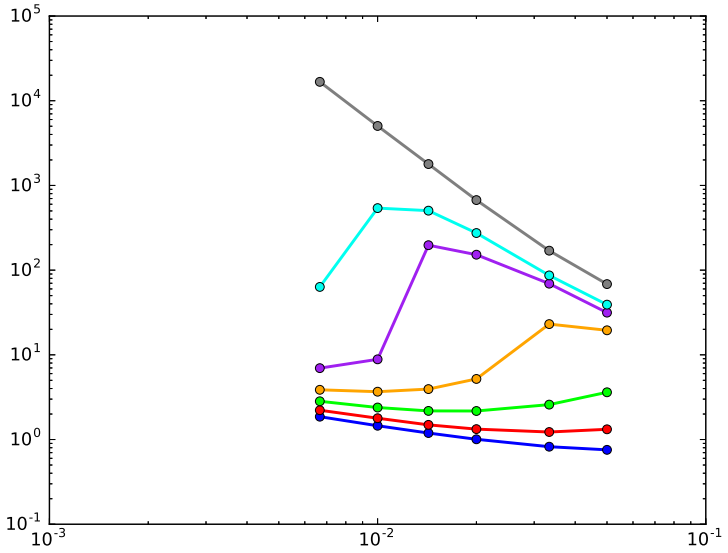


Figure 19: Averaged mean Stokes default of gradient $\mathbb{\nabla}^{\mathrm{R1}}$ as a function of discretization length, when using gradient $\mathbb{\nabla}^{\mathrm{CLASS}}$ in 23 - 24. The values of γ used are: 0.3, 0.1, 0.03, 0.01, 0.003, 0.001 and 0.

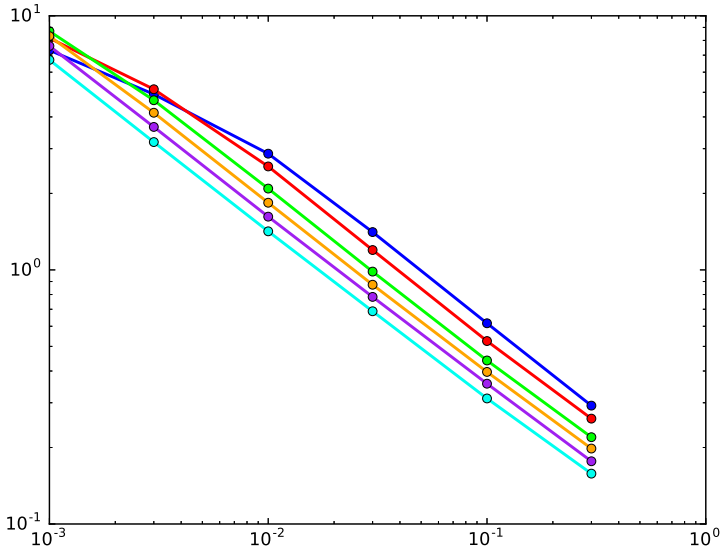


Figure 20: Averaged mean Stokes default as a function of the parameter γ in the case of the Taylor-Green vortex flow relaxed by formulation 23 - 24. The number of nodes used are: 400, 900, 2500, 4900, 10000 and 22500

analysis to more complicated simulations (fluid flows for instance, but also structural mechanics). Additionally, we would like to clarify the role of the functional $\mathcal{V}$ defined in equation 19 and generalize its definition to other gradients.

REFERENCES

[1] S. Adami, X.Y. Hu, and N.a. Adams. A transport-velocity formulation for smoothed particle hydrodynamics. *Journal of Computational Physics*, 241:292–307, 2013.

[2] a.K. Chaniotis, D. Poulikakos, and P. Koumoutsakos. Remeshed Smoothed Particle Hydrodynamics for the Simulation of Viscous and Heat Conducting Flows. *Journal of Computational Physics*, 182(1):67–90, 2002.

[3] J Bonet and S Kulasegaram. Correction and stabilization of smooth particle hydrodynamics methods with applications in metal forming simulations. *International Journal for Numerical Methods in Engineering*, 47(July 1998):1189–1214, 2000.

[4] Edmond Kwan-yu Chiu. *A Conservative Meshless Framework for Conservation Laws with Applications in Computational Fluid Dynamics*. PhD thesis, Stanford University, 2011.

[5] Fernando de Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. Blue noise through optimal transport. *ACM Transactions on Graphics*, 31(6):1, 2012.

[6] J Donea, Antonio Huerta, J Ponthot, and a Rodr. Arbitrary Lagrangian Eulerian Methods. *Encyclopedia of Computational Mechanics*, pages 1–25, 1999.

[7] R Fatehi and M T Manzari. Error estimation in smoothed particle hydrodynamics and a new scheme for second derivatives. *Computers Math.\ Applic.*, 61:482–498, 2011.

[8] Raanan Fattal. Blue-noise point sampling using kernel density model. *ACM Transactions on Graphics*, 30(4):1, 2011.

[9] Henri Faure, Peter Kritzer, and Friedrich Pillichshammer. From van der Corput to modern constructions of sequences for quasi-Monte Carlo rules. *Indagationes Mathematicae*, 26(5):760–822, 2015.

[10] R. A. Ginghold and J.J. Monaghan. Smoothed particle hydrodynamics : theory and application to non-spherical stars. *Monthly Notices of the Royal Astronomical Society*, 181:375–389, 1977.

[11] PM Guilcher. *Contribution au développement d'une méthode SPH pour la simulation numérique des interactions houle-structure*. PhD thesis, Université de Nantes, 2008.

[12] Min Jiang, Yahan Zhou, Rui Wang, Richard Southern, and Jian Jun Zhang. Blue noise sampling using an SPH-based method. *ACM Transactions on Graphics*, 34(6):1–11, 2015.

[13] Nathalie Lanson. *Etude des méthodes particulaires renormalisées. Applications aux problèmes de la dynamique rapide.* PhD thesis, INSA, 2001.

[14] Nathalie Lanson and Jean-Paul Vila. Renormalized Meshfree Schemes I: Consistency, Stability, and Hybrid Methods for Conservation Laws, 2008.

[15] S. J. Lind, R. Xu, P. K. Stansby, and B. D. Rogers. Incompressible smoothed particle hydrodynamics for free-surface flows: A generalised diffusion-based algorithm for stability and validations for impulsive flows and propagating waves. *Journal of Computational Physics*, 231(4):1499–1523, 2012.

[16] S. Litvinov, X.Y. Hu, and N.a. Adams. Towards consistence and convergence of conservative SPH approximations. *Journal of Computational Physics*, 301:394–401, 2015.

[17] L. B. Lucy. A numerical approach to the testing of the fission hypothesis. *The astronomical journal*, 82(12):1013–1024, 1977.

[18] Bertrand Maurel. *Modélisation par la méthode SPH de l'impact d'un réservoir rempli de fluide*. PhD thesis, INSA Lyon, 2008.

[19] J.J. Monaghan. On the problem of penetration in particle methods. *Journal of Computational Physics*, 82(1):1–15, 1989.

[20] J.J. Monaghan. SPH without a Tensile Instability. *Journal of Computational Physics*, 159(2):290–311, apr 2000.

[21] Guillaume Pierrot. Meshless methods: a unified framework and analysis. In *ESI Expert Seminar on Meshless Methods*, 2011.

[22] Guillaume Pierrot and Gabriel Fougeron. Enforcing differentiation/integration compatibility in meshless methods with nodal integration. In *Particle-based Methods IV Fundamentals and Applications*, 2015.

[23] Guillaume Pierrot and Gabriel Fougeron. Discrete gradient theorem and element-based integration in meshless methods. *VII European Congress on Computational Methods in Applied Sciences and Engineering*, 2016.

[24] Daniel J Price. Smoothed Particle Hydrodynamics and Magnetohydrodynamics. *J. Comp. Phys.*, 231:759–794, 2012.

[25] Volker Springel. Smoothed Particle Hydrodynamics in Astrophysics. *ARAA*, 48(1), 2010.

[26] JW Swegle, DL Hicks, and SW Attaway. Smoothed Particle Hydrodynamics Stability Analysis. *Journal of Computational Physics*, 116(1):123–134, 1995.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELING OF NON-SPHERICAL, ELONGATED PARTICLES FOR INDUSTRIAL SUSPENSION FLOW SIMULATION

Jakob D. Redlinger-Pohn[1], Lisa M. König[1], Christoph Kloss[2], Christoph Goniva[2], and Stefan Radl[1]

[1] Institute of Process and Particle Engineering, Graz University of Technology
Inffeldgasse 13/III, 8010 Graz, Austria
e-mail: redlinger-pohn@tugraz.at, lisa.koenig@tugraz.at, radl@tugraz.at

[2] DCS Computing GmbH
Industriezeile 35, 4020 Linz, Austria
e-mail: christoph.kloss@dcs-computing.com, christoph.goniva@dcs-computing.com

Keywords: Open-Source Code, Euler-Lagrange Simulation, Elongated Particles, Suspension Flow, Particle Segregation

Abstract. *Euler-Lagrange (EL) simulations of particulate suspension flow are an important tool to understand and predict multiphase flow in nature and industrial applications. Unfortunately, solid-liquid suspensions are often of (mathematically) stiff nature, i.e., the relaxation time of suspended particles may be small compared to relevant flow time scales. Involved particles are typically in the size range from µm to mm, and of non-spherical shape, e.g., elongated particles such as needle-shaped crystals and/or natural and man-made fibres. Depending on their aspect ratio and bending stiffness, those particles can be treated as rigid, or flexible. In this paper we present a recent implementation into the open-source LIGGGHTS® and CFDEM® software package for the simulation of systems involving stiff non-spherical, elongated particles. A newly implemented splitting technique of the coupling forces and torques, following the ideas of Fan and Ahmadi (J. Aerosol Sci. 26, 1995), allows significantly larger coupling intervals, leading to a substantial reduction in the computational cost. Hence, large-scale industrial systems can be simulated in an acceptable amount of time. We first present our modeling approach, followed by the verification of our code based on benchmark problems. Second, we present results of one-way coupled CFD-DEM simulations. Our simulations reveal segregation of fibres in dependence on their length due to fibre-fluid interaction in torus flow.*

1 INTRODUCTION

Understanding the transport of suspended particles is of great importance for the rational design of industrial equipment. Early studies concentrated on suspension of spherical particles due to their relative simplicity. However, such systems are rarely found in industrial application and rather limited to academic cases, i.e. spherical glass beads in suspensions. Thus the application of simulation results for the design of industrial equipment might not be advisable. In the last two decades, efforts were made to model and simulate elongated particles, i.e. cellulose fibres, needle shaped crystals, in suspension flow. We will review selection of recent studies in the following chapter. We then introduce our simulation approach which is based entirely on open-source code, i.e., OpenFOAM, LIGGGHTS®, and CFDEM®. The capabilities of our code will be demonstrated on hand of an examples of suspension flow of cellulose fibres.

1.1 Approaches to simulate Elongated Particles

Most of the previous studies [1] use an Euler-Lagrange approach to simulate the suspension flow of elongated particles. It allows the tracking of individual particle and thus to describe the segregation and orientation of single particles in suspension flow. For example, Soldati and Marchioli extended their simulation studies of particle deposition in turbulent boundary layer from spherical particles to elongated particles of different aspect ratio and density [2–4]. Thus, varying the particle response time and particle inertia. Rigid elongated particles were regarded as point particles. Collisions between particles were neglected, and collisions of particles with the wall was not fully resolved, i.e., only the wall-normal velocity of a particle was reversed upon impact. Creeping flow around the particle is assumed, and the calculation of hydrodynamic force and torque follows the analytical solution presented by Kim and Karrila [5]. Comparing non-spherical particles to spherical particles, they found that the wallward drift increased with the particle length, and thus the longest particles deposited near the wall at a higher rate.

Lindström, and later Andric extended fibre modelling from rigid fibres to flexible fibres [6,7]. In their approach they followed earlier work of Schmid et.al [8,9] who studied the flocculation behavior of flexible fibres in shear flow. Ca. 7 to 10 spheroids, i.e., rigid elements were linked together in the model for one flexible fibre. This, however, increased the simulation effort. Their model was capable to consider inertial effects and included particle-particle interaction. Back-coupling of forces from the particle exerted on the fluid was considered in the code, however not actively used in their simulation work. Findings of their study were that the specific viscosity for non-straight but curved fibres increased with the fibre curvature. In suspension flow, flexible fibres tended to align with the vorticity axis. This was in difference to rigid fibres which tended to align in the shear-plane [10].

Zhao and van Wachem extended their simulation studies and included effects of particle motion on the fluid motion [11]. Particles were coupled to the fluid by adding a source term which represented the sum of the drag force of particles in one fluid cell averaged over the volume of the fluid cell. They also considered particle-particle interaction and particle-wall interaction. Contact detection was performed by placing small spheres at the hull of one spheroid and thus using common contact detection algorithms for spherical particles. The repelling force was calculated using a Hertzian-like force model. Comparing their findings to findings of Marchioli and Soldati [4], they found that the turbulent kinetic energy decreased by 4.8% for suspensions of spheroids at a volume concentration of 0.0219%. Elongated particles were orientated randomly in the channel center. Closer to the wall particle orientation was preferen-

tially into the flow direction. They found the effect of the particle-wall interaction more pronounced, which was attributed to a more realistic interaction model.

Common in the above studies is that the authors used the same test geometry: a rectangular channel (dimension: 4π, 2, 2π, in x, y, z) which is wall bounded in the *y*-direction, and cyclic in the *x*-, and *z*- direction. The test case was based on studies of turbulent channel flow performed by Kim et.al. [12]. Following their settings, Marchioli and Soldati [3] used a semi-spectral method to solve the fluid motion. Lindström and Uesaka [6], Andric et.al. [13], as well as Zhao and van Wachem [11] used a finite-volume solver to simulate the Navier-Stokes equation.

Differently, the research group of Salcudean and Gartshore focused on simulation of fibre suspension flow in process equipment, i.e., the head-box of a paper machine, a hydrocyclone, and in a slotted channel [14–16]. To solve for the fluid motion, they used LES and RANS approaches, i.e., not resolving all flow structures. Elongated particles were modelled as stiff and flexible fibres. Flexible fibres were again modelled as chain of 4 spheroids linked together. A wall model was developed to account for fibre-wall interaction. Specifically, the wall was resolved by a fine mesh, smaller than the spheroid elements. Calculating the normal and tangential force, a force was applied on the spheroid element in contact with the wall being large enough to prevent the fibre from penetrating the wall. Hydrodynamic drag and torque interaction followed the work of Kim and Karilla [5] for creeping flow around the spheroid. Mass and lift forces were neglected. They found an accumulation of fibres in the vicinity of the wall. For a fibre-suspension in a wall bounded channel flow, they found an accumulation of fibres at a distance of the wall equal to their half-length.

2 MODELING APPROACH

We use an open-source CFD-DEM code to solve for the particle motion in a dilute fibre-suspension. The Navier-Stokes equations are solved using OpenFOAM (The OpenFOAM Foundation, 2014). Particles are placed into the developed flow field. Particle motion is solved with LIGGGHTS®. The coupling to the fluid is performed with CFDEM®coupling [17], which was extended to account for non-spherical, elongated stiff particles.

The translational and rotational motion of the fibre (in an inertial frame of reference) can be described by Newton's Equation of motion:

$$m_p \, d_t \mathbf{v} = \mathbf{f}^h + \mathbf{f}^w + \mathbf{f}^c \ , \tag{1}$$

$$\frac{d}{dt}\left(\mathbf{I}\cdot\boldsymbol{\omega}\right) = \mathbf{t}^h + \mathbf{t}^c \, . \tag{2}$$

Here, m_p is the mass of the prolate spheroid, and $\mathbf{I}$ is the moment of inertia tensor for a prolate spheroid. $\mathbf{f}^h$ and $\mathbf{t}^h$ are force and torque exerted on the particle by hydrodynamic interaction. $\mathbf{f}^c$ and $\mathbf{t}^c$ are interaction force and torque due to particle collision with the wall and/or other particles. $\mathbf{f}^w$ are body forces on the fibre. On the particle side (LIGGGHTS®), the elongated particle is represented as a spherocylinder, i.e., a cylinder with to half-spheres on each end of the cylinder (Figure 1a). The geometries are described by the cylinder length l and the cylinder diameter c. The radius of the half-sphere is thus $c/2$. It is possible to account for additional particle roughness, i.e., fibrils of a cellulose fibres. The height of the roughness is defined by δ_{rough} in our model.

For computing the fluid interaction forces and torques, the elongated particle is represented as a spheroid (Figure 1b). There is an analytical solution for spheroids under creeping flow conditions presented by Kim and Karrila [5], which we follow in the hydrodynamic interaction.

The major semi-axis *a* of the spheroid equals half the length of the cylinder, $a = l/2$. The minor semi-axis of the spheroid *b* is defined such that the hydrodynamic properties of the cylinder and the spheroid in shear-flow are identical. According to Cox [18], *b* calculates to:

$$b = \frac{1}{2.48} c \sqrt{\ln\left(\lambda_{cylinder}\right)} \quad . \tag{3}$$

$\lambda_{cylinder}$ is the aspect ratio of the spherocylinder, defined as $\lambda_{cylinder} = l/c$.

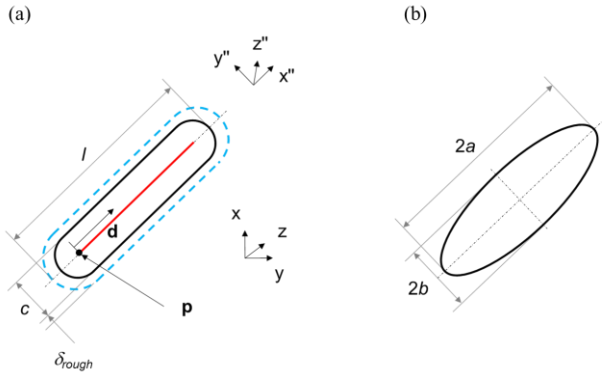


Figure 1: Schematic representation of the spherocylinder (a), and the prolate spheroid (b). *l* and *c* are the length and the diameter of the spherocylinder. **p** and **d** are the origin of the line segment and its (normalized) direction vector. δ_{rough} represents the particle surface roughness. *a*, and *b* are the major and minor semi-axis of the spheroid respectively. The global and particle associated Cartesian coordinate systems are indicated by (x,y,z) and (x",y",z") respectively.

Point-volume particles are assumed, and the fluid translational velocity **u**, as well as the local rotation rate **Ω**, are taken from the fluid cell coinciding with particle center, defined as $\mathbf{p} + \left(l - c\right)\mathbf{d}/2$.

2.1 Hydrodynamic Interactions

Our model for the hydrodynamic interaction coincides with the model presented by Lindström and Uesaka [6]. Assuming creeping flow, hydrodynamic force $\mathbf{f}^h$ and hydrodynamic torque $\mathbf{t}^h$ will be dominated by viscous effects modeled as:

$$\mathbf{f}^v = \mathbf{A}^v \cdot \left[\mathbf{u} - \mathbf{v}\right] \quad . \tag{4}$$

$$\mathbf{t}^v = \mathbf{C}^v \cdot \left[\mathbf{\Omega} - \boldsymbol{\omega}\right] + \mathbf{H}^v \cdot \dot{\boldsymbol{\gamma}} \, . \tag{5}$$

Here **v** is the fibre velocity at its center of mass position. $\dot{\boldsymbol{\gamma}}$ is the rate of strain tensor $\dot{\boldsymbol{\gamma}} = \left[\nabla\mathbf{u} + \left(\nabla\mathbf{u}\right)^T\right]/2$. $\mathbf{A}^v$, $\mathbf{C}^v$, and $\mathbf{H}^v$ are hydrodynamic resistance tensors defined as:

$$\mathbf{A}^v = 3\pi\eta l \left[Y^A \boldsymbol{\delta} + \left(X^A - Y^A\right)\mathbf{d} \otimes \mathbf{d}\right] \, . \tag{6}$$

$$\mathbf{C}^v = \pi\eta l^3 \left[Y^C \boldsymbol{\delta} + \left(X^C - Y^C\right)\mathbf{d} \otimes \mathbf{d}\right] . \tag{7}$$

$$\mathbf{H}^v = \pi\eta l^3 Y^H \left(\boldsymbol{\varepsilon}\mathbf{d}\right) \otimes \mathbf{d} \, . \tag{8}$$

δ and **ε** denote the unit tensor and the permutation tensor. The parameters X^A, Y^A, X^C, Y^C, and Y^H depend on the eccentricity $e = \left(1 - b^2/a^2\right)^{1/2}$ of the prolate spheroid [6]. Previous work [6] also introduced a term considering inertial effects, i.e., an inertial drag force $\mathbf{f}^I$ and torque $\mathbf{t}^I$.

The resulting total hydrodynamic force $\mathbf{f}^h$ and hydrodynamic torque $\mathbf{t}^h$ on an elongated particle thus calculates to:

$$\mathbf{f}^h = \left(\mathbf{A}^v + \mathbf{A}^I\right)\cdot\left[\mathbf{u}-\mathbf{v}\right]. \tag{9}$$

$$\mathbf{t}^h = \left(\mathbf{C}^v + \mathbf{C}^I\right)\cdot\left[\mathbf{\Omega}-\boldsymbol{\omega}\right] + \left(\mathbf{H}^v + \mathbf{H}^I\right)\cdot\dot{\boldsymbol{\gamma}}. \tag{10}$$

As described by Fan and Ahmadi [19], the equations for the particle motion (eqn. 1, and eqn. 2) are of stiff nature, and need to be solved using a forward-differencing scheme. To do so, the fluid-particle interactions terms in the governing equations must be decomposed into and an explicit (indicated by subscript '*exp*' below) and an implicit contribution. Furthermore, it is useful to solve the translational equation of motion in the inertial reference frame, and that for rotational motion in a reference frame aligned with the principal axes of the particles. The latter is indicated by primed variables in what follows. In our work we also consider a simple model for added mass effects (due to liquid rotating with the particles). In analogy to Fan and Ahmadi [19], the equations of motion read:

$$\left(1+C_{addM}\,\rho_f/\rho_p\right)m_p d_t\mathbf{v} = \mathbf{f}_{\exp} + \mathbf{K}_{sl}\cdot\left(\mathbf{u}-\mathbf{v}\right), \tag{11}$$

$$\mathbf{I}_p^{'}\cdot d_{t^{'}}\left(\boldsymbol{\omega}^{'}\right) + \boldsymbol{\omega}^{'}\times\left(\mathbf{I}_p^{'}\cdot\boldsymbol{\omega}^{'}\right) = \mathbf{t}_{\exp}^{'} + \mathbf{K}_{sl,rot}^{'}\cdot\left(\mathbf{\Omega}^{'}-\boldsymbol{\omega}^{'}\right), \text{ with} \tag{12}$$

$$\mathbf{I}_p^{'} = \mathbf{I}^{'}\left(1+C_{add}\,\rho_f/\rho_p\right) \tag{13}$$

Here $\mathbf{K}_{sl}$ and $\mathbf{K}_{sl,rot}$ are the diagonal part of the translational and rotational drag resistance tensors defined above, respectively. Note that $\mathbf{I}$' is the moment of inertia tensor in the body-fixed coordinate system, which is constant for each particle.

We now apply a Crank-Nicholson forward discretization scheme (with CN_α being the Crank-Nicholson blending factor) to arrive at the following explicit evolution equations for the translational and rotation speed of each particle at timestep n+1:

$$\mathbf{v}^{\{n+1\}} = \frac{\mathbf{v}^{\{n\}} + \dfrac{\Delta t}{\left(1+C_{addM}\,\rho_f/\rho_p\right)m_p}\left[\mathbf{f}_{\exp}^{\{n\}} + \mathbf{K}_{sl}^{\{n\}}\cdot\left(\mathbf{u}^{\{n\}} - \left(1-CN_\alpha\right)\mathbf{v}^{\{n\}}\right)\right]}{1+\Delta t\,\mathbf{K}_{sl}^{\{n\}}\,CN_\alpha/\left[\left(1+C_{addM}\,\rho_f/\rho_p\right)m_p\right]}, \tag{14}$$

$$\boldsymbol{\omega}_i^{'\{n+1\}} = \frac{\boldsymbol{\omega}_i^{'\{n\}} + \dfrac{\Delta t}{\mathbf{I}_{p,ii}^{'}}\left(\mathbf{t}_{tot,i}^{'\{n\}} - \left[\boldsymbol{\omega}_i^{'\{n\}}\times\left(\mathbf{I}_p^{'}\cdot\boldsymbol{\omega}_i^{'\{n\}}\right)\right]_i + \mathbf{K}_{sl,rot,i}^{'\{n\}}\,CN_\alpha\,\boldsymbol{\omega}_i^{'\{n\}}\right)}{1+\Delta t\,\mathbf{K}_{sl,rot,i}^{'\{n\}}\,CN_\alpha/\mathbf{I}_{p,ii}^{'}}, \text{ with} \tag{15}$$

$$\mathbf{t}_{tot}^{'\{n\}} = \mathbf{t}_{\exp}^{'\{n\}} + \mathbf{K}_{sl,rot}^{'\{n\}}\cdot\left(\mathbf{\Omega}^{'\{n\}} - \boldsymbol{\omega}^{'\{n\}}\right) \tag{16}$$

We have verified our implementation against the analytical solution of Jeffery [20]. Thus, an elongated particle (of zero mass) in a linear shear flow will rotate with a time period:

$$\tau = \frac{2\pi}{\dot{\gamma}_p} = \frac{2\pi\left(\lambda+1/\lambda\right)}{\dot{\gamma}}. \tag{17}$$

$\dot{\gamma}_p$ is the flipping frequency of the particle, and $\dot{\gamma}$ is the fluid shear rate. λ is the aspect ratio of the spheroid defined as, $\lambda = a/b$. A characteristic of the rotation of elongated particle in laminar shear flow is, that the particles remain in their initial orbit of rotation, i.e., the initial

orientation of the particle rotational axis with respect to the shear plane remains constant [5,21]. The orbit constant C is defined as

$$C = \lambda \tan(\theta)\left(\lambda^2 \sin^2(\phi) + \cos^2(\phi)\right)^{1/2}, \text{with} \tag{18}$$

$$\tan(\phi) = \lambda \tan\left(\frac{\dot{\gamma} t}{\lambda + 1/\lambda}\right), \tag{19}$$

$$\tan(\theta) = \frac{C\lambda}{\left(\sin^2(\phi) + \lambda^2 \cos^2(\phi)\right)^{1/2}}. \tag{20}$$

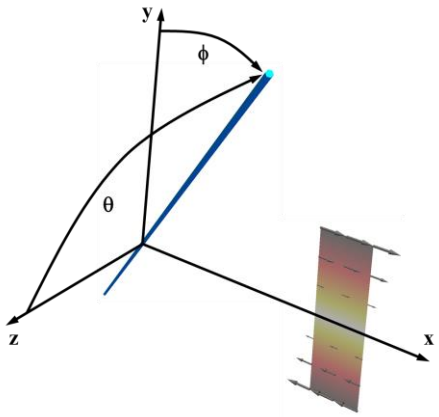


Figure 2: Coordinate system and definition of the Cartesian and polar coordinates with respect to the shear plane.

Figure 3a, compares the orientation angle ϕ from our simulation of an elongated particle, $\lambda = 10$, in a linear shear flow, $\dot{\gamma} = 666 s^{-1}$, to the analytical solution derived by Jeffery. The initial orientation is: $\phi = \pi/2$, and $\theta = \pi/2$, yielding an orbit constant of $C = 0.709$. In Figure 3b we plot the angular velocity, and observe that the simulation results perfectly represent the analytical solution for low density ratios. However, for increased particle density and thus increased particle inertia, Lundell and Carlson [22] found that the time-dependent flipping behavior of the elongated particle changes to a time independent rotational motion of the elongated particle. Figure 3c and d show the angular position and the angular velocity of a particle, initially orientated in shear plane ($C = \infty$) with a density ratio (particle to fluid) of 5000, respectively. We find, that the oscillation frequency increases and the particle actually is never at rest, i.e. $\omega > 0$. Thus, our code captures very well the transition from a time-dependent flipping to a time-independent rotational motion.

We note in passing that Rosén et.al. [21,23] performed detailed simulations of a single elongated particle, fully resolving the fluid-particle forces. He showed that at increased fluid inertia, elongated particles leave their initial orbit and drift towards a stable orbit. Unfortunately, our code is currently not able to account for this effect. Thus, simulations performed are yet limited to cases of small particle Reynolds number where we can assume creeping flow around the particle.

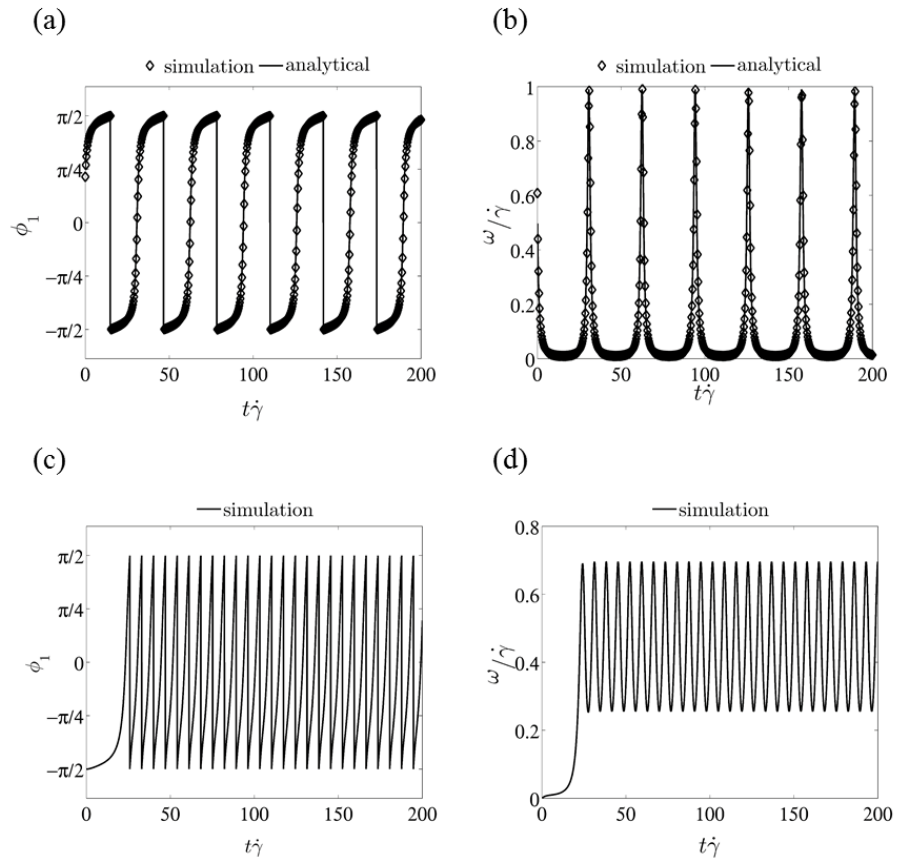


Figure 3: Time dependent orientation angle ϕ, and angular velocity ω, for an elongated particle, $\lambda = 10$, in linear shear flow. The particle relative density is 0.001 (panel a and b), and 5000 (panel c and d).

2.2 Contact Force and Torque

Calculation of contact force $\mathbf{f}^c$ and contact torque $\mathbf{t}^c$ follows a linear spring-dashpot model in normal and tangential direction [24,25]:

$$\mathbf{f}_p{}^c = \left(k_n \delta \mathbf{n}_{ij} - \gamma_n \mathbf{v}\mathbf{n}_{ij}\right) + \left(k_t \delta \mathbf{t}_{ij} - \gamma_t \mathbf{v}\mathbf{t}_{ij}\right) \ . \tag{21}$$

k_n and k_t are the normal and tangential spring stiffness, and γ_n and γ_t are the normal and tangential damping coefficient respectively. The vector $\mathbf{n}_{ij}$ and $\mathbf{t}_{ij}$ are the normal and tangential directions of the collision. Within the surface roughness region, the roughness adds to the total collision force, $\mathbf{f}^c = \mathbf{f}_p{}^c + \mathbf{f}_{rough}{}^c$. The formulation of $\mathbf{f}_{rough}{}^c$ is identical to the collision force of the particle. The torque is calculated as the cross product of the contact force $\mathbf{f}^c$ and the distance between the bodies $\mathbf{r}$.

$$\mathbf{t}^c = \mathbf{f}^c \times \mathbf{r} \ . \tag{22}$$

The contact detection algorithm follows the calculation of a line intersecting with another line (particle-particle), or a triangle (particle-wall, see Figure 4 for an illustration). The implementation follows the strategy proposed by Schneider and Eberly [26], and below the main idea of the algorithm is summarized. More details can be found in König [27].

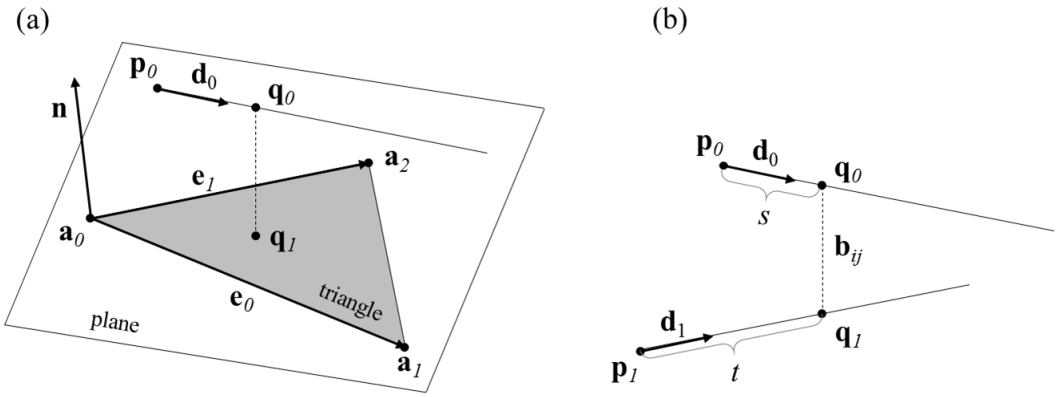


Figure 4: Graphical illustration of the (a) line-triangle interaction, and (b) line-line interaction [26,27].

The line segment is presented by its point of origin, the direction of orientation and its extension: $\mathbf{l}_0(s) = \mathbf{p}_0 + s\mathbf{d}_0$, and $\mathbf{l}_1(t) = \mathbf{p}_1 + t\mathbf{d}_1$. The two points on the line segment with the smallest distance are $\mathbf{q}_0$ and $\mathbf{q}_1$. The length of the vector $\mathbf{d}_{ij}$ presents the distance between both lines. The triangle segment is presented by $\mathbf{a}(u,v) = \mathbf{a}_0 + u\mathbf{e}_0 + v\mathbf{e}_1$ with $\mathbf{e}_0 = \mathbf{a}_1 - \mathbf{a}_0$, and $\mathbf{e}_1 = \mathbf{a}_2 - \mathbf{a}_0$. All points within the triangle are represented by $0 \le u$, $v \le 1$, subject to the constraint $u + v \le 1$. The code first checks if the elements, lines or line-triangle, are parallel to each other by calculating the dot product, $\mathbf{d}_0 \cdot \mathbf{n}$ and $\mathbf{d}_0 \cdot \mathbf{d}_1$, where $\mathbf{n}$ is the normal vector of the plane, $\mathbf{n} = \mathbf{e}_0 \times \mathbf{e}_1$. If the result is 0, both elements are parallel to each other. The minimum distance between the elements is found by minimizing the square distance function. In case the closest point of the two elements is of smaller distance than $c/2$ for line-triangle, or c for line-line interaction, the two elements are in contact with each other. In case the contact point is found on either end of the line segment, contact detection for spherical particles is applied.

While particle-particle interactions have been verified by König [27], selected results of the particle-wall interaction verification study are presented below. Specifically, the rebound translational v_x^+ and rotational ω_z^+ velocities were calculated following Kodam et.al. [28]:

$$\omega_z^+ = \frac{m \cdot v_x^- (1+\varepsilon) h' \cos(\alpha' + \theta)}{I_{zz} + m h'^2 \cos^2(\alpha' + \theta)} \tag{23}$$

$$v_x^+ = \omega_y^+ h' \cos(\alpha' + \theta) - \varepsilon v_x^- \tag{24}$$

Here the superscript + and – denote the post impact and pre impact quantities. v and ω are components of the translational and rotational velocity of the cylinder. ε is the coefficient of restitution, m is the mass of the prolate spheroid and I_{zz} is the spheroid inertia around the shorter semi-axis. h' and angle α' are geometrical values of the cylinder (Figure 5a). To apply the analytical solution, a rectangular cylinder corresponding to the spherocylinder had to be projected into the spherocylinder. Figure 5a illustrates the spherocylinder and its corresponding rectangular cylinder for a certain inclination θ. The shape of the projected cylinder (i.e., α', h', l', and c') changes with θ as follows:

$$c' = c \sin(\theta) \ . \tag{25}$$

$$l' = l - c(1 - \sin(\theta)) . \tag{26}$$

$$h' = \sqrt{c'^2 + l'^2}\,/\,2\,. \tag{27}$$

$$\alpha' = \arctan\left(c'/l'\right)\,. \tag{28}$$

Figure 5b compares the simulation results to the analytical solution, showing excellent agreement.

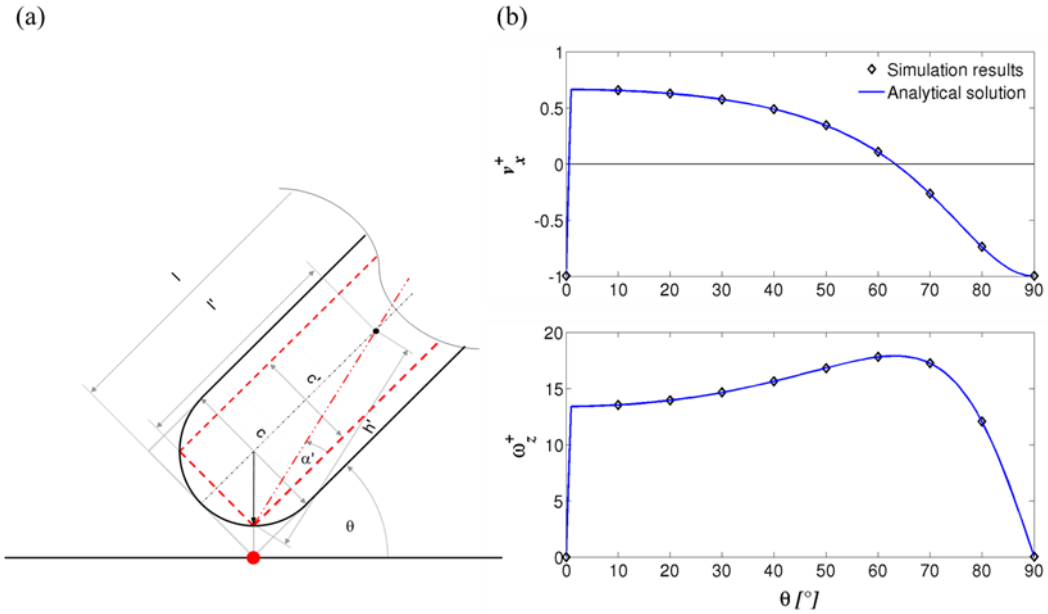


Figure 5: a: Fibre-wall impact presenting the sphero-cylinder and its corresponding rectangular cylinder. b: Comparison of simulation results to an analytical solution for rectangular cylinders [28]. Translational v_x^+ and rotational ω_z^+ velocity for different inclination angles θ.

3 RESULTS

In this section we illustrate our methodology by considering a general suspension flow test case, i.e., flow in a torus as previously considered by Redlinger-Pohn et al. [29]. The fibre-fluid density ratio is 1.3. The aspect ratio of the fibres was varied to model differently-sized fibres. The flow field was solved with OpenFOAM®, version 2.3.0. Time derivatives were approximated with a second-order Euler backwards scheme, and convective terms were approximated with the second-order Gauss linear interpolation scheme. The PISO algorithm was used to solve the equations. Settings for the fluid solver were in accordance with settings used by Andric et.al. [7].

3.1 The Motion of a Fibre Suspension in a Torus

Due to the curvature of coiled tubes the resulting flow field differs from the flow field in a straight pipe flow. In order to approximate such tubes, we here consider flow in a closed torus as an academic test case [30]. It is well known that centrifugal forces lead to a deflection of the fluid flow towards the outer bent, thus yielding a pressure gradient in the cross section [31,32]. Consequently, a secondary flow arises within the cross section from the outer bent *O* towards the inner bent *I* along the pipe wall, and from the inner bent to the outer bent along the equatorial mid plane. This secondary motion u_{sec} is called Dean flow, and the intensity is quantified by the Dean number *Da*:

$$\kappa = d/D\,. \tag{29}$$

$$u_{\text{sec}} = u_{bulk}\sqrt{\kappa}\ . \tag{30}$$

$$Re = \frac{u_{bulk}\cdot d}{\nu}\ . \tag{31}$$

$$Da = \frac{u_{\text{sec}}\cdot d}{\nu} = \text{Re}\cdot\sqrt{\kappa}\ \ . \tag{32}$$

The torus is sketched in Figure 6. d and D are the diameter of the tube cross section and the coil diameter of the torus respectively. u_{bulk} is the mean bulk velocity of the fluid, and ν is the kinematic viscosity of the fluid. The section above the equatorial mid-line is defined as upper half (*UH)*, and the section below is defined as lower half (*LH)*. *I* and *O* denote the inner and outer wall of the torus. Cartesian coordinates are donated as (x,y,z), and the corresponding toroidal coordinates as (r,φ,Θ).

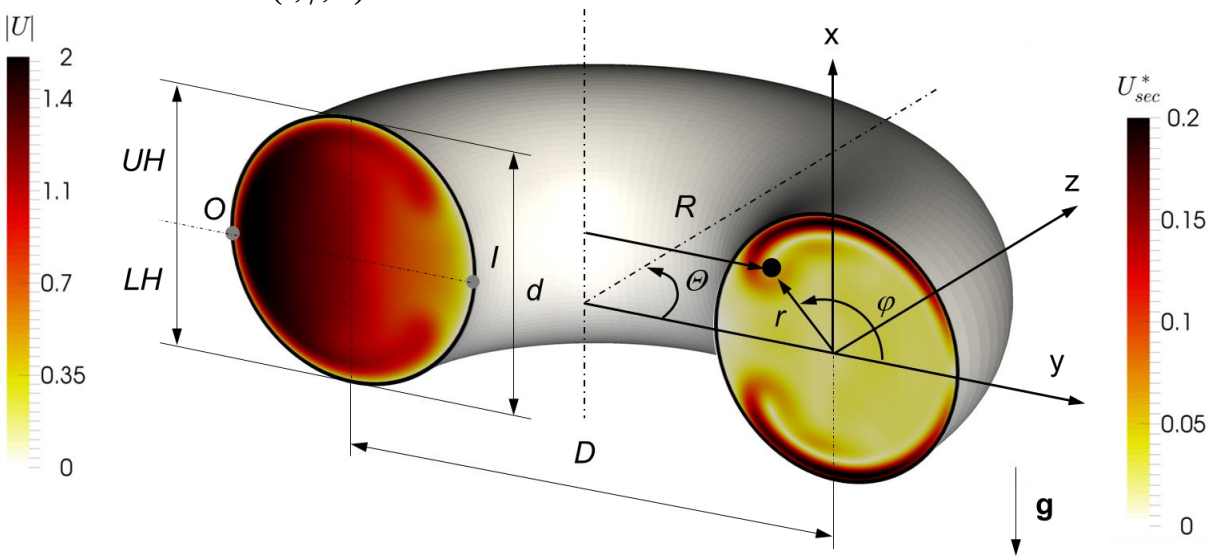


Figure 6: Schematic representation of the half torus in the Cartesian (x,y,z) and toroidal (r,φ,Θ) coordinate system. **g** denotes the direction of gravity. *D* and *d* present the torus and pipe diameter. *I* and *O* denote the inner and outer wall of the torus. Projected in the cross section is a typical flow profile of a laminar toroidal flow. Left cross section: flow in the stream wise direction, right cross section: secondary motion (*Da* = 1049, figure adapted from [29]).

Simulation of the toroidal flow is documented in Redlinger-Pohn et.al. [29]. The position of the Dean vortex is indicated by a black dot in Figure 6. At this position the secondary motion describes a rapid change in the flow direction.

Particles introduced into the flow field will be affected by the secondary motion. Depending on their mass and density, and thus their terminal settling velocity u_{set} particles will be (i) mixed in the cross section [33] or (ii) settle towards the *LH* and are pushed to the inner bent *I* by the secondary motion [34]. At the latter position we find particles in a region of low stream-wise fluid velocity.

In an extensive numerical study, fibres of different aspect ratio (ranging from 2 to 160) were introduced into the above shown flow profiles [29]. Here we document some additional findings from our previous simulation results that are illustrated in Figure 7: colored in yellow are regions where the (vertical) secondary motion exceeds the terminal settling velocity, and thus particles are lifted against gravity. Regions where the fibre settles in direction of the gravity are indicated in blue. We find, that regions in the *UH* where longer (and thus heavier)

fibres (λ = 160) are lifted are confined to the inner wall of the torus. Specifically, these regions are located at the outer bent where the secondary motion is strongest, and extend to the center of the Dean vortex. Regions where the smaller fibres (λ = 2) experience a lift are significantly larger, explaining the more homogeneous distribution of these fibres.

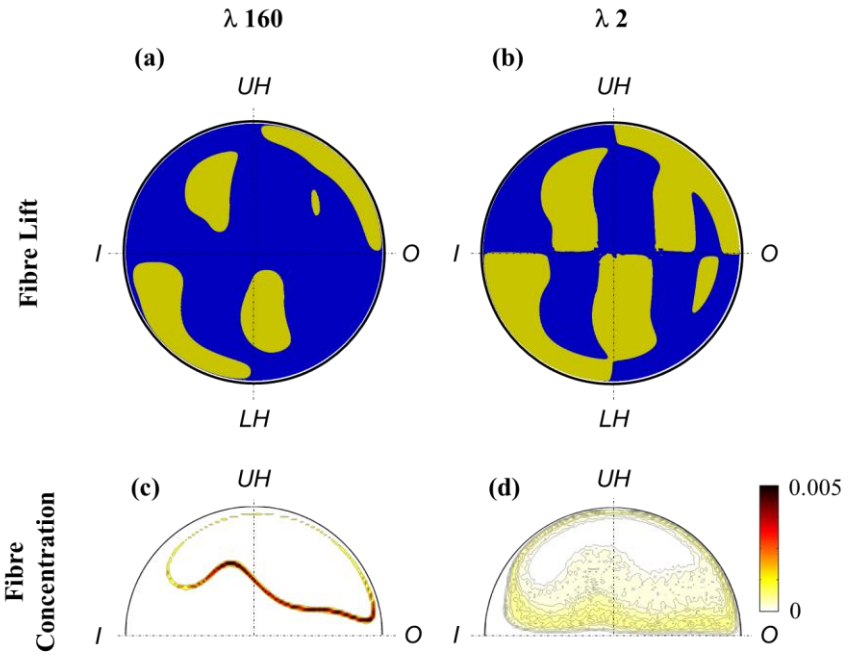


Figure 7: a and b: Sum of the secondary motion in direction of gravity and the fibres' terminal settling velocity. Yellow: secondary motion exceeding the settling velocity. c and d: Relative fibre concentration. *Da* = 1049 for this simulations.

The consequence of the fact that longer fibres only experience lift in certain regions explains the accumulation into a narrow band. Here the longer fibres orbit in the cross section in the *UH* (Figure 7c). Especially close to the equatorial midplane, where the fibres slowly move from the inner bent *I* to the outer bent *O*, we find a high concentration of fibres. Most likely, fibres will entangle in these regions and form a larger fibre-flock. Smaller fibres are more effectively mixed by the secondary motion, and are distributed more homogenously in the torus' cross section (see Figure 7d). The consequence is a fibre length-dependent segregation in flow direction, with the longer fibres moving (on average) faster than the fluid.

4 CONCLUSION AND OUTLOOK

We present details of an open-source code for the simulation of non-spherical, elongated particles suspended in a fluid. In our modeling approach we follow a strategy involving (i) rigid elongated particles, (ii) one-way coupling to the fluid, as well as (iii) fibre-fibre and fibre-wall interactions. For the particle interaction we represent the particle as spherocylinders, and detect the point of contact by a line-segment algorithm. Thus, we do not represent the particle surface by multiple spheres, allowing a comparably rapid calculation of the contact point. The code is successfully validated against standard cases for elongated particles, e.g., the Jeffery orbit.

Finally, we performed a simulation study of a fibre suspension in a torus. A fibre length-depended segregation was observed. The reason for this phenomena is that fibres are lifted by

the secondary motion in the upper half of the torus, i.e., they can move against gravity. This lifting effect differs for fibres of variable length. As a consequence we observe a localized high concentration of long fibres in the torus' cross section, while smaller fibres are rather homogeneously distributed. This segregation is certainly of key importance for the transport rate of fibres through the torus.

Future studies may include the simulation of fibre suspensions at higher Reynolds numbers, i.e., that exceeding the laminar flow regime. The investigation of more test cases used in literature (e.g., by Marchioli and Soldati [4], or Andric et.al. [7]) may pave the way towards an industrial application of our simulation tool. However, a more realistic flexible fibre model (than presented in our present work) should be used for these cases. This is because the formation of fibre aggregates certainly becomes a dominating effect at fibre concentrations relevant for industrial applications. Unfortunately, flexible fibre models require significantly more computational resources compared to the rigid fibre model in the present contribution.

5 ACKNOWLEDGEMENT

JDRP, LMK, and SR acknowledge partial support from FLIPPR (http://www.flippr.at/), as well from NAWI Graz by providing access to dcluster.tugraz.at. CK and CG acknowledge funding through the NanoSim project (http://www.sintef.no/projectweb/nanosim).

6 REFERENCES

[1] H.I. Andersson, A. Soldati, Anisotropic particles in turbulence: Status and outlook, Acta Mech. 224 (2013) 2219–2223. doi:10.1007/s00707-013-0926-y.

[2] C. Marchioli, A. Soldati, Rotation statistics of fibers in wall shear turbulence, Acta Mech. 224 (2013) 2311–2329. doi:10.1007/s00707-013-0933-z.

[3] C. Marchioli, M. Fantoni, A. Soldati, Orientation, distribution, and deposition of elongated, inertial fibers in turbulent channel flow, Phys. Fluids. 22 (2010) 1–14. doi:10.1063/1.3328874.

[4] A. Soldati, C. Marchioli, Physics and modelling of turbulent particle deposition and entrainment: Review of a systematic study, Int. J. Multiph. Flow. 35 (2009) 827–839. doi:10.1016/j.ijmultiphaseflow.2009.02.016.

[5] S. Kim, S.J. Karrila, Microhydrodynamics: Principles and Selected Applications, Dover Publications, 2005. https://books.google.at/books?id=_8llnUUGo0wC.

[6] S.B. Lindström, T. Uesaka, Simulation of the motion of flexible fibers in viscous fluid flow, Phys. Fluids. 19 (2007). doi:10.1063/1.2778937.

[7] J. Andrić, S.T. Fredriksson, S.B. Lindström, S. Sasic, H. Nilsson, A study of a flexible fiber model and its behavior in DNS of turbulent channel flow, Acta Mech. 224 (2013) 2359–2374. doi:10.1007/s00707-013-0918-y.

[8] C.F. Schmid, D.J. Klingenberg, Mechanical Flocculation in Flowing Fiber Suspensions, Phys. Rev. Lett. 84 (2000) 290–293. doi:10.1103/PhysRevLett.84.290.

[9] C.F. Schmid, L.H. Switzer, D.J. Klingenberg, Simulations of fiber flocculation: Effects of fiber properties and interfiber friction, J. Rheol. (N. Y. N. Y). 44 (2000) 781. doi:10.1122/1.551116.

[10] L. Jianzhong, Z. Weifeng, Y. Zhaosheng, Numerical research on the orientation distribution of fibers immersed in laminar and turbulent pipe flows, J. Aerosol Sci. 35 (2004) 63–82. doi:10.1016/S0021-8502(03)00388-4.

[11] F. Zhao, B.G.M. Van Wachem, Direct numerical simulation of ellipsoidal particles in turbulent channel flow, Acta Mech. 224 (2013) 2331–2358. doi:10.1007/s00707-013-0921-3.

[12] J. Kim, P. Moin, R. Moser, Turbulence statistics in fully developed channel flow at low Reynolds number, J. Fluid Mech. 177 (1987) 133. doi:10.1017/S0022112087000892.
[13] J. Andrić, S.B. Lindström, S. Sasic, H. Nilsson, Rheological properties of dilute suspensions of rigid and flexible fibers, J. Nonnewton. Fluid Mech. 212 (2014) 36–46. doi:10.1016/j.jnnfm.2014.08.002.
[14] Z. Wang, Numerical Simulation of Fiber Separation in Hydrocyclones, Master Thesis, The University of British Columbia, 2002.
[15] X. Zhang, Fiber Orientation in a Headbox, Master Thesis, The University of British Columbia, 2001.
[16] S. Dong, Modeling of Fiber Motion in Pulp and Paper Equipment, PhD Thesis, The University of British Columbia, 2002.
[17] C. Goniva, C. Kloss, N.G. Deen, J.A.M. Kuipers, S. Pirker, Influence of rolling friction on single spout fluidized bed simulation, Particuology. 10 (2012) 582–591. doi:10.1016/j.partic.2012.05.002.
[18] R.G. Cox, The motion of long slender bodies in a viscous fluid Part 1. General theory, J. Fluid Mech. 44 (1970) 791. doi:10.1017/S002211207000215X.
[19] F.-G. Fan, G. Ahmadi, A sublayer model for wall deposition of ellipsoidal particles in turbulent streams, J. Aerosol Sci. 26 (1995) 813–840. doi:10.1016/0021-8502(95)00021-4.
[20] G.B. Jeffery, The Motion of Ellipsoidal Particles Immersed in a Viscous Fluid, Proc. R. Soc. A Math. Phys. Eng. Sci. 102 (1922) 161–179. doi:10.1098/rspa.1922.0078.
[21] T. Rosén, M. Do-Quang, C.K. Aidun, F. Lundell, The dynamical states of a prolate spheroidal particle suspended in shear flow as a consequence of particle and fluid inertia, J. Fluid Mech. 771 (2015) 115–158. doi:10.1017/jfm.2015.127.
[22] F. Lundell, A. Carlsson, Heavy ellipsoids in creeping shear flow: Transitions of the particle rotation rate and orbit shape, Phys. Rev. E. 81 (2010) 016323. doi:10.1103/PhysRevE.81.016323.
[23] T. Rosén, F. Lundell, C.K. Aidun, Effect of fluid inertia on the dynamics and scaling of neutrally buoyant particles in shear flow, J. Fluid Mech. 738 (2014) 563–590. doi:10.1017/jfm.2013.599.
[24] P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies, Géotechnique. 29 (1979) 47–65. doi:10.1680/geot.1979.29.1.47.
[25] S. Luding, Introduction to discrete element methods. Basic of contact force models and how to perform the micro-macro transition to continuum theory, Rev. Eur. Génie Civ. 12 (2008) 785–826. doi:10.3166/ejece.12.785-826.
[26] P.J. Schneider, D.H. Eberly, Geometric Tools for Computer Graphics, Elsevier, 2003. doi:10.1016/B978-1-55860-594-7.50023-0.
[27] L.M. König, Particle Separation in a Virtual Microfluidic Channel, Master Thesis, Graz University of Technology, 2015.
[28] M. Kodam, R. Bharadwaj, J. Curtis, B. Hancock, C. Wassgren, Cylindrical object contact detection for use in discrete element method simulations, Part II-Experimental validation, Chem. Eng. Sci. 65 (2010) 5863–5871. doi:10.1016/j.ces.2010.08.007.
[29] J.D. Redlinger-Pohn, L.A. Jagiello, W. Bauer, S. Radl, Mechanisitc Understanding of Size-Based Fibre Separation in Coiled Tubes, Int. J. Multiph. Flow. submitted (2016).
[30] T.J. Hüttl, R. Friedrich, Influence of curvature and torsion on turbulent flow in helically coiled pipes, Int. J. Heat Fluid Flow. 21 (2000) 345–353. doi:10.1016/S0142-727X(00)00019-9.
[31] W.R. Dean, XVI. Note on the motion of fluid in a curved pipe, London, Edinburgh, Dublin Philos. Mag. J. Sci. 4 (1927) 208–223. doi:10.1080/14786440708564324.

[32] I. Di Piazza, M. Ciofalo, Transition to turbulence in toroidal pipes, J. Fluid Mech. 687 (2011) 72–117. doi:10.1017/jfm.2011.321.
[33] T.K. Palazoglu, K.P. Sandeep, Effect of tube curvature ratio on the residence time distribution of multiple particles in helical tubes, LWT - Food Sci. Technol. 37 (2004) 387–393. doi:10.1016/S0023-6438(03)00013-6.
[34] P. Tiwari, S.P. Antal, M.Z. Podowski, Three-dimensional fluid mechanics of particulate two-phase flows in U-bend and helical conduits, Phys. Fluids. 18 (2006) 043304. doi:10.1063/1.2189212.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A LAGRANGIAN FINITE ELEMENT METHOD FOR THE SIMULATION OF 3D COMPRESSIBLE FLOWS

M. Cremonesi[1], **and A. Frangi**[1]

[1]Department of Civil and Environmental Engineering
Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133, Milano, Italy
e-mail: {massimiliaano.cremonesi,attilio.frangi}@polimi.it

Keywords: compressible flow, Lagrangian approach, PFEM.

Abstract. *The numerical solution of compressible fluid flows is of paramount importance in many industrial and engineering applications. Compared to the classical fluid dynamics, the introduction of the fluid compressibility changes the formulation of the problem and consequently its computational treatment.*

Among the possible numerical solutions of compressible flow problems, the finite element method has always been privileged. However, the standard Eulerian approaches with fixed domain are not particularly suited to represent the strong shock waves and the significant movement of the external boundaries. On the contrary, in problems characterized by evolving surfaces, Lagrangian approaches can be very effective.

The governing equations of compressible flow problems are mass, momentum and energy conservation. These equations are discretized in the spirit of the Lagrangian Particle Finite Element Method (PFEM). The strong distortions of the mesh, typical of the Lagrangian approaches, are managed with a continuous remeshing of the computational domain. The nodal unknowns are velocities, density and internal energy. To fully exploit the potential of continuous remeshing, only nodal variables are stored and consequently only linear interpolation are used. In addition, an artificial viscosity has been introduced to stabilize the formation and propagation of shock waves. Finally, explicit time integration of the governing equations enables a highly efficient solution of the discretized problem.

The proposed approach has been validated against typical benchmarks of gas dynamics in the presence of strong shock waves. A very good agreement has been shown in all the tests proving the excellent accuracy and versatility of the proposed method.

1 INTRODUCTION

The solution of the equations of high-speed compressible fluid has a wide range of engineering and physics applications. Due to the presence of shocks, viscous boundary layers and mixing of different constituents the numerical solution of these equations can be very complex.

Different approaches can be used for the solution of hydrodynamics equations: Eulerian [3, 2], Lagrangian [14, 7] or ALE [1, 16]. In this work, a Lagrangian finite element approach is exploited to describe strong evolving material interfaces and shocks propagations.

The differential problem has been solve with a Particle Finite Element Method [11, 5]. As expected with a full Lagrangian approach the mesh can deteriorate in time, consequently, to guarantee a good quality of the mesh, a remeshing technique is proposed. Morevover, an explicit time integration of the governing equations is proposed to solve efficiently the discretized problem. This paper is structured as follows. In section 2 the equations of motion for a compressible fluid are presented. Space and time discretizations are introduced in sections 3.1 and 3.2 respectively. Section 4 is devoted to the mesh update techniques while 5 presents the numerical results.

2 GOVERNING EQUATIONS

A compressible fluid in an evolving domain $\Omega(t)$ is considered. In the Lagrangian description, the particle position $\mathbf{x}$ at time t can be expressed as a function of velocity field $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ through:

$$\mathbf{u} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \tag{1}$$

In this framework, balance equations can be written as:

$$\text{mass conservation:} \qquad \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\nabla\cdot\mathbf{u} \quad \text{in } \Omega(t)\times[0,T] \tag{2}$$

$$\text{momentum conservation:} \qquad \rho\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \nabla\cdot\boldsymbol{\sigma} \quad \text{in } \Omega(t)\times[0,T] \tag{3}$$

$$\text{energy conservation:} \qquad \rho\frac{\mathrm{d}e}{\mathrm{d}t} = \boldsymbol{\sigma} : \nabla\mathbf{u} \quad \text{in } \Omega(t)\times[0,T] \tag{4}$$

where $\rho = \rho(\mathbf{x}, t)$ is the density, $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{x}, t)$ the Cauchy stress tensor and $e = e(\mathbf{x}, t)$ the internal energy per unit of mass. In many compressible gas-dynamics applications, the viscosity in very low and consequently shear terms in the Cauchy stress tensor can be neglected, leading to:

$$\boldsymbol{\sigma} = -p\mathbf{I} \tag{5}$$

where $p = p(\mathbf{x}, t)$ is the pressure field. Following this assumption, (2)-(4) reduce to the standard compressible Euler equations.

An equation of state for polytropic ideal gas is considered, linking pressure to density and internal energy through:

$$p = (\gamma - 1)\rho e \tag{6}$$

where $\gamma > 1$ is a constant adiabatic index.

3 DISCRETIZATIONS

3.1 Space discretization

To discretize in space the equations (2)-(4) a standard Galerkin approach has been followed. Linear continuous finite element are used to discretize both kinematic variables (velocities) and thermodynamic variables (pressure,density and internal energy). Introducing $\mathbf{U}$ the vector of the nodal values of the velocity and $\mathbf{E}$ the nodal values of the internal energy, the semidiscretized form of momentum and energy conservation reads:

$$\mathbf{M}_u \frac{d\mathbf{U}}{dt} = -\mathbf{F}_u \tag{7}$$

$$\mathbf{M}_e \frac{d\mathbf{E}}{dt} = \mathbf{F}_e \tag{8}$$

where $\mathbf{M}_u$ and $\mathbf{M}_e$ are mass matrices while $\mathbf{F}_u$ and $\mathbf{F}_e$ are the vectors of internal forces.

Mass conservation (2) is discretized starting from the strong form [14, 8]:

$$\rho(\mathbf{x}, t) J(\mathbf{x}, t) = \rho(\mathbf{x}, 0) \tag{9}$$

where $J(\mathbf{x}, t) = \det \mathbf{J}(\mathbf{x}, t)$ is determinant of the deformation Jacobian. Introducing the finite element approximation, the discretized problem reads:

$$\mathbf{M}_\rho(t)\mathbf{R} = \mathbf{R}_0 \tag{10}$$

where vector $\mathbf{R}$ contains the unknown nodal density. The equation of state (6) can be then used pointwise in every node of the mesh to directly to find the pressure field:

$$\mathbf{P} = (\gamma - 1)\,\mathbf{R}\mathbf{E} \tag{11}$$

3.2 Time discretization

Introducing a forward Euler integration scheme to discretize the time derivatives, the full discretized system writes:

$$\mathbf{M}_u \mathbf{U}^{n+1} = \mathbf{M}_u \mathbf{U}^n + \Delta t \mathbf{F}_u(\boldsymbol{\sigma}^n, \mathbf{U}^n) \tag{12}$$

$$\mathbf{M}_e \mathbf{E}^{n+1} = \mathbf{M}_e \mathbf{E}^n - \Delta t \mathbf{F}_e(\boldsymbol{\sigma}^n, \mathbf{U}^{n+1/2}) \tag{13}$$

$$\mathbf{M}_\rho^{n+1} \mathbf{R}^{n+1} = \mathbf{R}^0 \tag{14}$$

where $\mathbf{U}^{n+1/2} = \frac{1}{2}(\mathbf{U}^{n+1} + \mathbf{U}^n)$ is used to preserve total energy from time t^n to time t^{n+1} [7, 8]. The pressure unknowns is computed directly by equation of state:

$$\mathbf{P}^{n+1} = (\gamma - 1)\,\mathbf{R}^{n+1}\mathbf{E}^{n+1} \tag{15}$$

The system of equations (12)-(14) is particolarly suited for the fully explicit solution. If mass lumping is introduced velocity and energy can be evaluated node by node without solving linear systems. To respect the CFL condition, an adaptive time step is defined as:

$$\Delta T = \alpha \min_e \left(\frac{h_e}{\sqrt{\gamma_e\, p_e / \rho_e}} \right) \tag{16}$$

where h_e is a characteristic length of an element and α is a safety parameter [15, 8].

3.3 Artificial viscosity

To follow the shocks propagation, a tensor artificial viscosity has been introduced in the equations. A viscous term, activated only in the shock regions, has been added to the Cauchy stress tensor (5):

$$\boldsymbol{\sigma} = -p\mathbf{I} + \boldsymbol{\sigma}_a \tag{17}$$

where $\boldsymbol{\sigma}_a$ is defined as:

$$\boldsymbol{\sigma}_a = \mu_e \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T) \tag{18}$$

and μ_e is a viscous coefficient defined on every element e of the mesh as:

$$\mu_e = \psi_0 \psi_1 \rho h_e (\frac{2}{3} h_e |C_e| + \frac{1}{4} \psi_2 c_e) \tag{19}$$

where c_e is the speed of sound a C_e a measure of the compressibility. The stabilization is activated only in compression region through the compressibility switch ψ_1. The switch ψ_2 is a measures the vorticity-compressibility ratio while ψ_0 represents a smoothness detector. Details on the artificial viscosity can be found in [12].

4 NUMERICAL TECNIQUE

The Lagrangian Particle Finite Element Method is here used to solved the system of equations (12)-(14). The PFEM was orginally developed to solve problems involving free surfaces fluid flows [9, 13, 4] and fluid-structure interaction [10, 6]. The method is here revisited and applied to the simulation of compressible problems.

Due to the Lagrangian nature of the equations, mesh nodes are moved following the fluid velocity leading to possible mesh distortion. An index of the element distortion is used to check whether the mesh should be regenerated or not. When a new mesh should be created aN efficient Delaunay tesselletion technique is used to redefine the nodal connectivity starting from the current node position. The solution scheme is sketched in 1.

Algorithm 1 Solution scheme with remeshing

for $n = 1, ..., N_{steps}$ **do**
  check mesh distortion
  if mesh too distorted **then**
    Generate new mesh with Delaunay tessellation
    Compute $\mathbf{M}_u$, $\mathbf{M}_e$ and $\mathbf{R}^0$
  end if
  compute velocity $\mathbf{U}^{n+1}$ from eq.(12)
  update position $\mathbf{x}^{n+1} = \mathbf{x}^n + \Delta t\, \mathbf{U}^{n+1}$
  compute density $\mathbf{R}^{n+1}$ through eq.(14)
  compute internal energy $\mathbf{E}^{n+1}$ from eq.(13)
  compute pressure $\mathbf{P}^{n+1}$ from eq.(15)
  estimate the new stable Δt from eq. (16)
end for

5 NUMERICAL EXAMPLE

5.1 Sod shock tube

In the simple domain defined in figure 1, two gases are separated by a discontinuity in the middle. Density, pressure and velocity are initialized as reported in table 1. Internal energy is

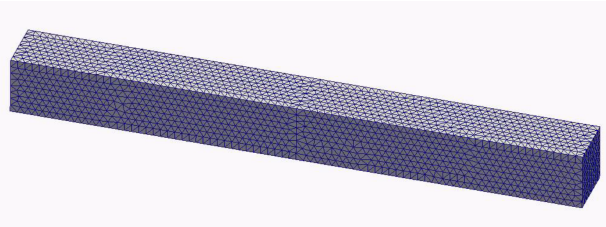


Figure 1: Sod shock tube: geometry and mesh.

directly derived from equation of state (6) and γ is fixed to $5/3$.

	Left	Right
density (ρ)	1.0	0.125
pressure (p)	1.0	0.1
velocity ($\mathbf{u}$)	0.0	0.0

Table 1: Sod shock tube: geometry and mesh.

A time interval $t = [0, 0.2]$ is considered. Results at final time $t = 0.2$ are compared with the analytical solution in Figure 2 showing a good agreement. Contour plot of the final solution are depicted in Figure 3.

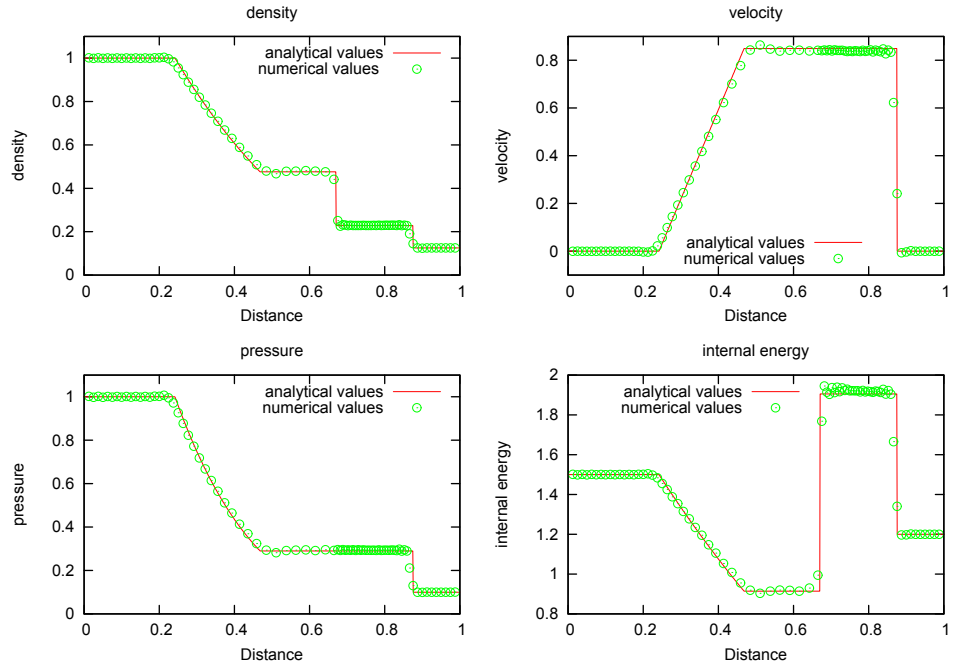


Figure 2: Sod shock tube: results at $t = 0.2$ (horizontal line in the middle of the 3D domain).

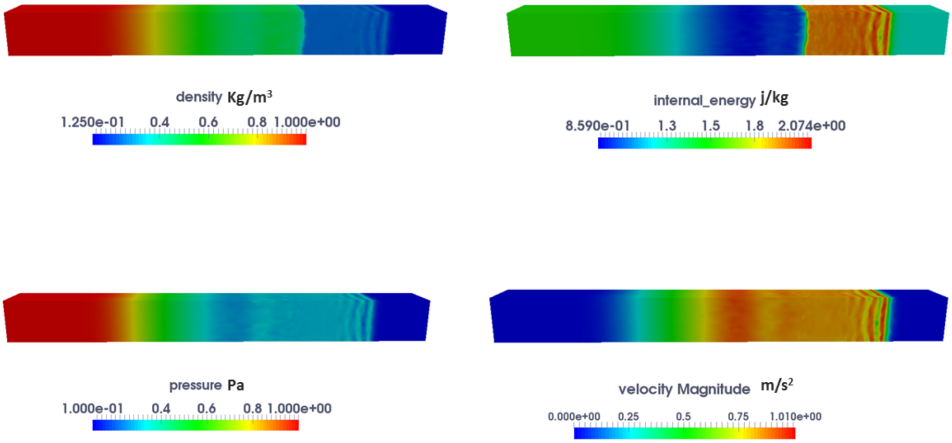


Figure 3: Sod shock tube: contour plot at $t = 0.2$.

5.2 Bubble test

In this test a bubble of helium is immersed in a gas at rest. A shock wave moves from the right size of the box and collides with the bubble. Figure 4 depicts the geometry and the initial conditions of the problem. The mesh is composed by 106.196 nodes and 582.626 elements.

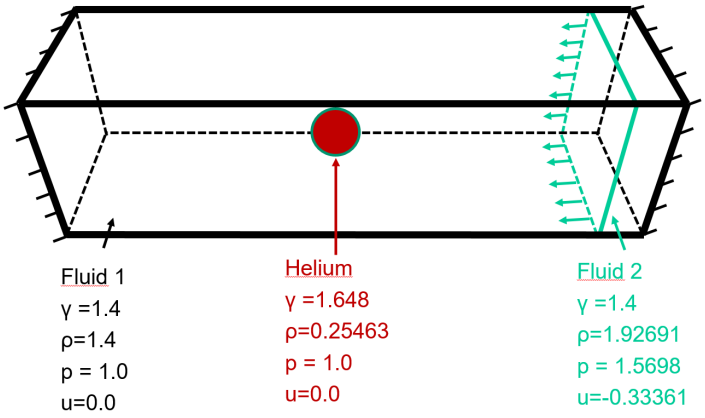


Figure 4: Bubble test: geometry and initial conditions

Figure 5 shows the time evolution of the density and internal energy of a central section of the 3D domain.

6 CONCLUSIONS

In the present work, a Lagrangian finite element method has been introduced for the numerical simulation of compressible fluid flows. Particle Finite Element Method has been used to the numerical solution of the balance equations. An efficient Delaunay tessellation has been used to avoid excessive mesh distortion typical of the Lagrangian flows. An explicit time integration has been proposed to solve rapidly the discretized balance equations. To validate the proposed approach, two benchmarks have been presented.

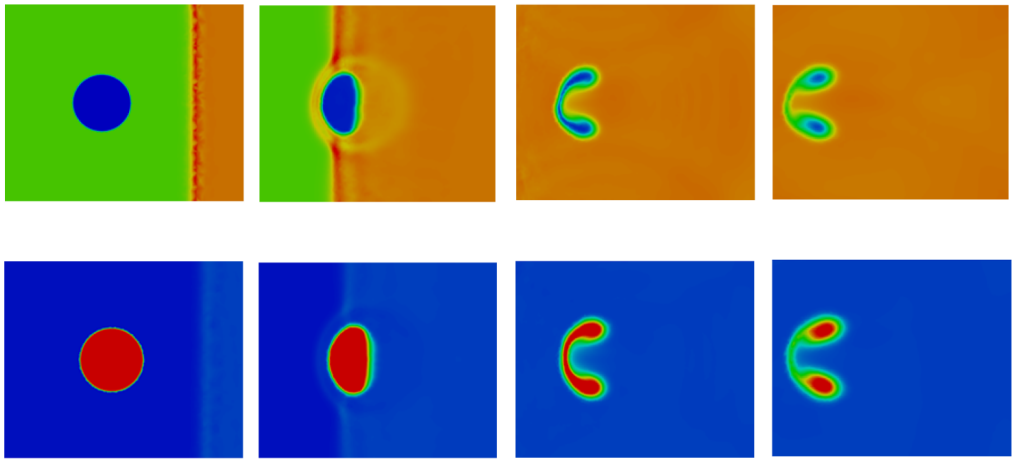


Figure 5: Bubble test. Top: time evolution of the density; bottom: time evolution of the internal energy.

REFERENCES

[1] A. J. Barlow. A compatible finite element multi-material ale hydrodynamics algorithm. *International Journal for Numerical Methods in Fluids*, 56(8):953–964, 2008.

[2] A. Bernard-Champmartin, J.P. Braeunig, and J.M. Ghidaglia. An eulerian finite volume solver for multi-material fluid flows with cylindrical symmetry. *Computers & Fluids*, 83:170 – 176, 2013. Numerical methods for highly compressible multi-material flow problems.

[3] J.-P. Braeunig, B. Desjardins, and J.-M. Ghidaglia. A totally eulerian finite volume solver for multi-material fluid flows. *European Journal of Mechanics - B/Fluids*, 28(4):475 – 485, 2009.

[4] M. Cremonesi, L. Ferrara, A. Frangi, and U. Perego. Simulation of the flow of fresh cement suspensions by a lagrangian finite element approach. *Journal of Non-Newtonian Fluid Mechanics*, 165:1555–1563, 2010.

[5] M. Cremonesi and A. Frangi. A lagrangian finite element method for compressible flow applications. *submitted for the pubblication*, 2016.

[6] M. Cremonesi, A. Frangi, and U. Perego. A lagrangian finite element approach for the analysis of fluidstructure interaction problems. *International Journal of Numerical Methods in Engineering*, 84:610–630, 2010.

[7] V.A. Dobrev, T.E. Ellis, T.V. Kolev, and R.N. Rieben. Curvilinear finite elements for lagrangian hydrodynamics. *International Journal for Numerical Methods in Fluids*, 65(11-12):1295–1310, 2011.

[8] V.A. Dobrev, T.V. Kolev, and R.N. Rieben. High-order curvilinear finite element methods for lagrangian hydrodynamics. *SIAM Journal on Scientific Computing*, 34(5):B606–B641, 2012.

[9] S. R. Idelsohn, E. Oñate, and F. Del Pin. The particle finite element method: a powerful tool to solve incompressible flows with free-surfaces and breaking waves. *International Journal for Numerical Methods in Engineering*, 61:964–989, 2004.

[10] S. R. Idelsohn, E. Oñate, F. Del Pin, and N. Calvo. Fluid-structure interaction using the particle finite element method. *Computer methods in applied mechanics and engineering*, 195:2100–2113, 2006.

[11] K. Kamran, R. Rossi, E. Oñate, and S.R. Idelsohn. A compressible lagrangian framework for the simulation of the underwater implosion of large air bubbles. *Computer Methods in Applied Mechanics and Engineering*, 255(0):210 – 225, 2013.

[12] Tz.V. Kolev and R.N. Rieben. A tensor artificial viscosity using a finite element approach. *Journal of Computational Physics*, 228(22):8336 – 8366, 2009.

[13] E. Oñate, S. R. Idelsohn, F. Del Pin, and R. Aubry. The particle finite element method. an overview. *International Journal for Computational Methods*, 1:267–307, 2004.

[14] G. Scovazzi, M.A. Christon, T.J.R. Hughes, and J.N. Shadid. Stabilized shock hydrodynamics: I. a lagrangian method. *Computer Methods in Applied Mechanics and Engineering*, 196(4-6):923–966, 2007.

[15] G. Scovazzi, E. Love, and M.J. Shashkov. Multi-scale lagrangian shock hydrodynamics on q1/p0 finite elements: Theoretical framework and two-dimensional computations. *Computer Methods in Applied Mechanics and Engineering*, 197(912):1056 – 1079, 2008.

[16] J. Waltz, N.R. Morgan, T.R. Canfield, M.R.J. Charest, L.D. Risinger, and J.G. Wohlbier. A three-dimensional finite element arbitrary Lagrangian-Eulerian method for shock hydrodynamics on unstructured grids. *Computers Fluids*, 92(0):172 – 187, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A LAGRANGIAN PFEM APPROACH TO THE NUMERICAL SIMULATION OF 3D LARGE SCALE LANDSLIDES IMPINGING IN WATER RESERVOIRS

M. Cremonesi[1], **F. Ferri**[1], **and U. Perego**[1]

[1]Department of Civil and Environmental Engineering
Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133, Milano, Italy
e-mail: {massimiliaano.cremonesi,francesco.ferri,umberto.perego}@polimi.it

Keywords: PFEM, landslides, water reservoir, Lagrangian approach.

Abstract. *Landslides are exceptional natural hazards that can generate extensive damage to structures and infrastructures causing a large number of casualties. A particularly critical condition occurs when the landslide impinges in water reservoirs generating high waves. This work proposes a numerical tool to simulate the macroscopic behavior of a propagating landslide. The Particle Finite Element Method (PFEM) is here used and adapted to the specific case of landslide runout. The Lagrangian Navier-Stokes equations of incompressible fluids are used to describe the macroscopic landslide behavior. A rigid-visco-plastic law with a pressure dependent threshold, typical of a non-Newtonian, Bingham-like fluid, is used to characterize the constitutive behavior of the flowing material. Special attention is devoted to the definition of ad-hoc pressure-dependent slip boundary conditions at the interface between the flowing mass and the basal surface to better represent the real landslide-slope interaction. The proposed approach has been validated against numerical benchmarks and small scale experimental tests, showing a good agreement with the physical measurements. Real case scenarios have also been considered. 3D geometries of critical sites, where landslides have occurred, have been reconstructed allowing for the simulation of large scale real landslide runouts. Results are compared with post-event images and measurements, showing the accuracy and the capability of the method.*

1 INTRODUCTION

Catastrophic landslides impinging into water reservoirs may generate impulsive waves whose propagation can cause considerable damages. Instability of a slope can be triggered by natural events (erosion, earthquakes, intense rainfalls) or by human activities (deforestation, construction). Once initiated, a landslide can run along a slope with different velocities and cover long distances, depending also on the slope morphology. The prediction of landslides velocity, runout distance and traveling path is useful for preventing and mitigating the consequences of these events.Consequently, the interest for numerical tools capable to simulate landslides with potentially long run-out distance, including their interaction with water basins, is continuously growing.

Recent developments in the simulation techniques for coupled problems have led to efficient analysis procedures allowing for the accurate reproduction of landslide-reservoir interactions (see for example [1, 2]). The numerical analysis of these events requires capabilities for tracking interfaces and free surfaces undergoing large displacements, and accounting for the mixing of different constituents, for complex constitutive behaviors and for multi-physics processes. In this work, a recently developed Lagrangian finite element approach [3, 4, 5], formulated in the spirit of the Particle Finite Element Method (PFEM), is here reconsidered and adapted to the specific case of landslide-reservoir interaction. The PFEM was originally developed [5, 6, 7] for solving problems involving free surfaces fluid flows and fluid-structure interaction. The Lagrangian nature of the approach allows for a natural treatment of free surfaces undergoing large displacements and of fast evolving interfaces, making the method particularly suitable for the simulation of landslide phenomena and of landslide-water interaction problems, which are dominated by fast propagating waves and interfaces. Due to the excessive distortion of the mesh, typical of Lagrangian approaches for fluids, a continuous remeshing, based on a fast 3D Delaunay triangulation, is implemented. A 3D version of the alpha-shape technique [5] is used to define the position and the evolution of the free-surfaces. New slip boundary conditions of Navier type have been formulated to model the interaction between the sliding material and the basal surface. The PFEM formulation has been modified to incorporate these new conditions.

2 GOVERNING EQUATIONS

The running landslide material is assumed to behave as a rigid-viscoplastic fluid. As a consequence, both landslide and water motions are governed by Navier-Stokes equations (here written in an Arbitrary Lagrangian-Eulerian framework [8]):

$$\begin{aligned} \rho\left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{c}\cdot\nabla_x)\mathbf{u}\right) &= \nabla_x\cdot\boldsymbol{\sigma} + \rho\mathbf{b} \quad \text{in } \Omega_t\times(0,T) \\ \nabla_x\cdot\mathbf{u} &= 0 \quad \text{in } \Omega_t\times(0,T) \end{aligned} \tag{1}$$

where ∇_x is the gradient spatial operator computed with respect to the current configuration, $\boldsymbol{\sigma}$ is the Cauchy stress tensor, ρ is the density and and $\mathbf{b}$ are the external body forces. The convective velocity $\mathbf{c}$ is defined as:

$$\mathbf{c} = \mathbf{u} - \mathbf{r} \tag{2}$$

$\mathbf{u}$ being the velocity of the fluid and $\mathbf{r}$ the mesh velocity. In general, an additional equation governing the evolution of the mesh $\mathbf{r}$ is necessary. The standard Eulerian description of the equation of motion can be recovered when $\mathbf{r} = 0$ (i.e. $\mathbf{c} = \mathbf{u}$). Conversely, the Lagrangian description is obtained with $\mathbf{r} = \mathbf{u}$ (i.e. $\mathbf{c} = 0$). As a consequence, in the Eulerian approach

the mesh is fixed, while in the Lagrangian one the mesh moves with the velocity of the fluid particles.

Equations (1) need be supplemented with proper initial and boundary conditions. The boundary $\partial\Omega_t$ is partitioned into three non-overlapping subsets $\partial\Omega_t = \Gamma_t^D \cup \Gamma_t^N \cup \Gamma_t^S$. On Γ_t^D and Γ_t^N standard Dirichlet and Neumann boundary conditions are imposed, while on Γ_t^S slip boundary conditions are considered. On Γ_t^D Dirichlet boundary conditions are imposed on velocities and on Γ_t^N Neumann boundary conditions are imposed on surface tractions:

$$\begin{aligned} \mathbf{u}(\mathbf{x},t) &= \bar{\mathbf{u}}(\mathbf{x},t) \qquad \text{on } \Gamma_t^D \\ \boldsymbol{\sigma}(\mathbf{x},t)\cdot\mathbf{n} &= \mathbf{h}(\mathbf{x},t) \qquad \text{on } \Gamma_t^N \end{aligned} \tag{3}$$

where $\bar{\mathbf{u}}(\mathbf{x},t)$ and $\mathbf{h}(\mathbf{x},t)$ are assigned functions and $\mathbf{n}$ is the outward normal to the boundary.

3 CONSTITUTIVE LAW

Both the landslide and the reservoir water are modeled as viscous fluids. The Cauchy stress tensor $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{x},t)$ is decomposed into its hydrostatic p and deviatoric $\boldsymbol{\tau}$ components as $\boldsymbol{\sigma} = -p\mathbf{I} + \boldsymbol{\tau}$ where $\mathbf{I}$ the identity tensor. Water is assumed to be a Newtonian isotropic incompressible fluid so that the deviatoric stress $\boldsymbol{\tau}$ can be directly related to the symmetric part of the velocity gradient:

$$\boldsymbol{\tau} = \frac{1}{2}\mu\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right) \tag{4}$$

where μ is the dynamic viscosity.

Unlike in standard Navier-Stokes formulations, the landslide material is assumed to obey a regularized elastic-viscoplastic non-Newtonian, Bingham-like constitutive model. When external actions trigger the landslide motion and the yield limit is exceeded, large viscoplastic deformations take place and the running landslide behaves as a viscoplastic Bingham fluid. In the assumed model the deviatoric stress follows again the relation (4), but the dynamic viscosity μ is replaced by the apparent viscosity $\widetilde{\mu}$ defined as:

$$\widetilde{\mu} = \mu + \frac{p\,\tan(\varphi)}{|\dot{\gamma}|}\left(1 - e^{-n|\dot{\gamma}|}\right) \tag{5}$$

where φ is the friction angle and $|\dot{\gamma}|$ represents a norm of the symmetric part of the velocity gradient. The exponential term in (5) has only a regularization purpose [2, 9], and does not have a constitutive interpretation. The constitutive law expressed by equation (5) can be interpreted as a viscoplastic behavior with a Drucker-Prager type yield criterion. A detailed description of the derivation of this model can be found in [10].

4 SLIP CONDITIONS

Along the boundary Γ_t^S, Navier type slip boundary conditions with a pressure dependent threshold are enforced:

$$\mathbf{u}_{slip} = -\beta\frac{\mathbf{t}}{\|\mathbf{t}\|}\langle\|\mathbf{t}\| - p\tan\varphi_{basal}\rangle \tag{6}$$

where $\mathbf{u}_{slip}$ is the velocity at the interface and $\mathbf{t}$ the traction force defined as (see also Figure 1 for a graphical representation):

$$\mathbf{t} = (\mathbf{I} - \mathbf{n}\otimes\mathbf{n})\,(\boldsymbol{\sigma}\mathbf{n}) \tag{7}$$

being $\beta = h_{slip}/\mu$ a parameter, having the dimension of a length over a viscosity, characterizing the basal interface, h_{slip} is the slip length for an ideal Couette flow with no threshold ($\varphi_{basal} = 0$) and φ_{basal} is the friction angle of the basal interface.

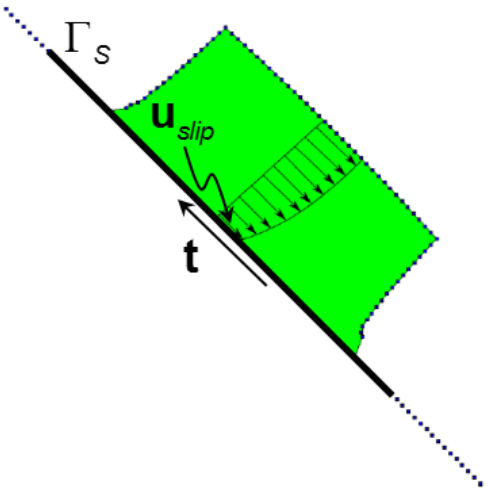


Figure 1: Slip velocity profile and basal tangential traction.

5 DISCRETIZATION AND NUMERICAL TECHNIQUE

Slip boundary conditions are difficult to enforce in a fully Lagrangian framework, in which nodes on the basal surface move according to their own velocity, but at the same time, in view of the PFEM remeshing strategy, have to define the position of the boundary. To overcome this difficulty, all nodes in the mesh are treated as Lagrangian (i.e. $\mathbf{c} = \mathbf{0}$ in eq. 1) except those on Γ^S (where slip conditions are imposed), which are treated as Eulerian (i.e. $\mathbf{c} = \mathbf{u}$) and therefore remain fixed in their initial position. This special treatment gives rise to a mixed Lagrangian-Eulerian formulation, where some nodes are Lagrangian and others are Eulerian.

A classical Galerkin Finite Element procedure is used to discretize the problem in space with linear interpolation functions for both pressure and velocity. The following semidiscretized form is obtained:

$$\mathbf{M}\dot{\mathbf{U}} + (\mathbf{K} + \mathbf{K}_{slip} + \mathbf{K}_c)\mathbf{U} + \mathbf{D}^T\mathbf{P} = \mathbf{B} \tag{8}$$

$$\mathbf{D}\mathbf{U} = \mathbf{0} \tag{9}$$

where $\mathbf{U}$ and $\mathbf{P}$ contain the nodal values of velocity and pressure respectively, $\mathbf{M}$ is the mass matrix, $\mathbf{K}$ is the matrix of viscoplastic coefficients, $\mathbf{D}$ is the discretization of the divergence operator, $\mathbf{B}$ is the vector of body forces and boundary tractions (see [3] for the definition of these matrices). $\mathbf{K}_c$ represents the discretization of the convective term on the slip boundary and $\mathbf{K}_{slip}$ is the discretization of the integral slip term. The contribution to matrices $\mathbf{K}_c$ and $\mathbf{K}_{slip}$ by all elements with no nodes on Γ^S is zero. A detailed description of these matrices can be found [10].

Introducing now a backward Euler integration scheme for the time integration, the final fully

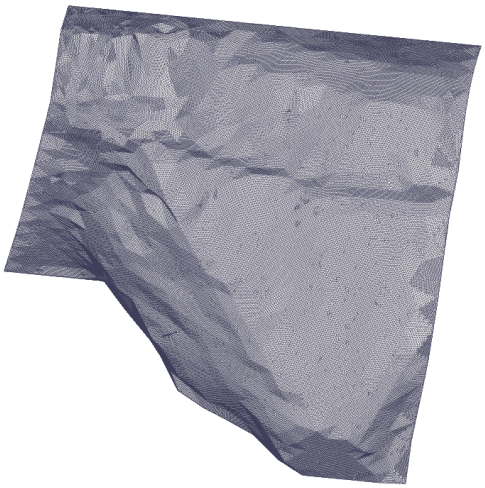


Figure 2: Cougar Hill. Finite element mesh used in the simulation.

discretized nonlinear problem writes:

$$\left(\frac{\mathbf{M}}{\Delta t}+\mathbf{K}+\mathbf{K}_{slip}+\mathbf{K}_c\right)\mathbf{U}^{n+1}+\mathbf{D}^T\mathbf{P}^{n+1}=\mathbf{B}^{n+1}+\frac{\mathbf{M}}{\Delta t}\mathbf{U}^n \tag{10}$$

$$\mathbf{D}\mathbf{U}^{n+1}=\mathbf{0} \tag{11}$$

The Lagrangian PFEM is here used to solve the differential problem (10-11). In the spirit of the Particle Finite Element Method, to avoid excessive mesh distortion due to the Lagrangian nature of the equations, the domain is frequently remeshed. An index of the element distortion is used to check whether the mesh should be regenerated or not. When a new mesh has to be created, a Delaunay tessellation technique is used to redefine the element connectivities starting from the current nodal positions. Moreover, an "alpha shape" technique is introduced to identify the free-surfaces and the interacting surfaces between water and landslide. Details on the numerical procedure can be found in [2, 3, 4, 5].

6 NUMERICAL EXAMPLES

6.1 Cougar Hill Landslide

The real event considered in this test case took place on Rocky Mountain in west Canada in 1992 [11]. Approximately 200 000 m^3 of debris started to flow from a 100 m high dump; the material stopped to flow after about 700 meters. The dump slope angle was estimated to be $37°$-$38°$. According to [11], the debris material was mainly sandy gravel and the foundation was composed by sand and gravel colluvium. The material parameters can be found in Table 1. Density and friction angles are provided in [11] while viscosity and slip length are estimated through back analysis.

Numerical results on the same example as well as some interesting discussions can be found in [12, 1, 13]. The 3D CAD model of the terrain (kindly provided by prof. M. Pastor) is plotted in figure 2. For the numerical analysis, a finite element mesh of 106046 nodes has been used with a fixed time step of $5 \cdot 10^{-3}$ seconds. The real time simulated in the analysis is 100 s.

Results are shown in Figure 3, where contour plots of the velocity at different time instants are plotted, and are in good agreement with results of other simulations available in the literature. Moreover, in Figure 4 the final deposit obtained with the present approach is compared with the result of the post-failure geotechnical investigations proposed in [11].

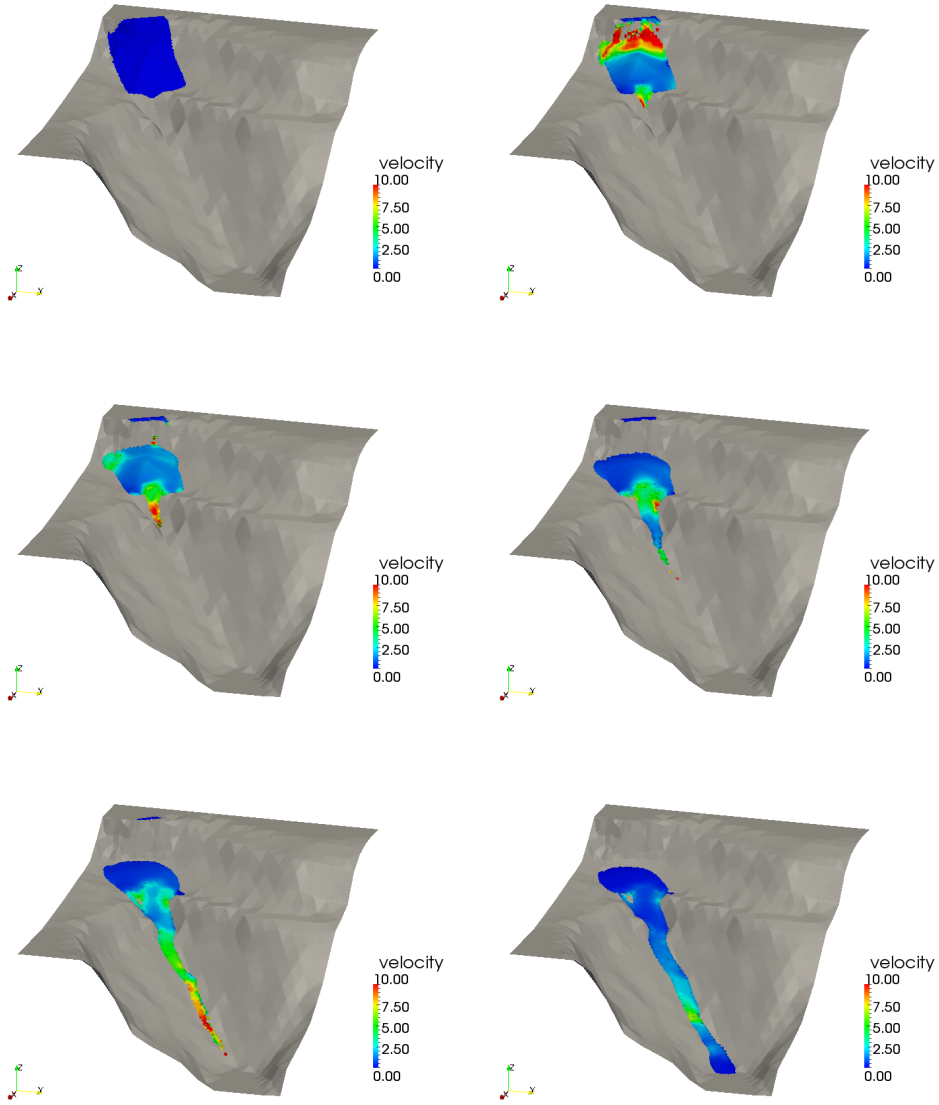


Figure 3: Cougar Hill. Velocity contour plots at t=0, 5, 20, 30, 70, 110 seconds.

ρ	$1.9\ 10^3\ Kg/m^3$
ϕ	$32°$
ϕ_{basal}	$28°$
μ_0	$0.01\ Pa\ s$
h_{slip}	$4\ m$

Table 1: Cougar Hill. Material parameters.

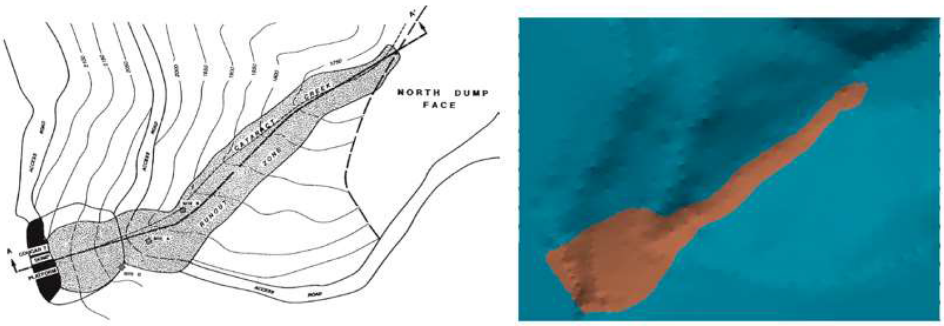


Figure 4: Cougar Hill. Comparison of the final deposit.

7 Vajont Landslide

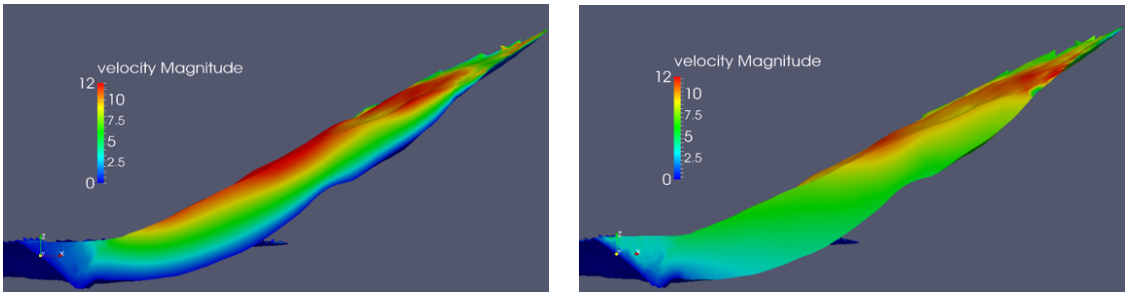
The Vajont valley is located in the western part of Friuli Venezia Giulia region, in Italy. The topographic boundaries of the Vajont valley are the S. Osvaldo pass on the east, the Tóc mount on the south, the SADE dam on the west and the mount Borga on the north. The valley is oriented east-west and was formed as the result of the erosion of an ancient glacier. The two main waterways are the Vajont stream which gives the name to the valley and its major tributary, the stream Mesazzo (see figure 7).



Figure 5: Aerial view of the Vajont and Piave valleis.

At about 10 pm of 9^{th} October 1963, approximately 270 millions cubic meters of material slid from the mount Toc into the dam reservoir, generating a huge wave that overtopped the dam destroying the town of Longarone in the Piave valley, downstream the dam. The landslide

produced a wave that has been estimated to have reached a height of about 105 meters above the the dam crest. Due to the irregular terrain topography, the wave did not reach the same elevation around the reservoir, as could be determined observing the contour of destroyed vegetation. In the Piave valley, the effects were much more destructive: the city of Longarone was almost completely destroyed, only a portion of the city at the north was not affected; the same happened for the surrounding villages of Villanova and Fa. Over 2000 people were killed and this is considered to have been one of the most disastrous landslides occurred in the XX century.



(a) $h_{slip} = 5m$ (b) $h_{slip} = 500m$

Figure 6: Vajont Slide. Effect of the slip boundary conditions.

The previously described numerical technique is here used for the simulation of the complete sequence of events: landslide runout, interaction with the water reservoir and generation of the overtopping wave. Other numerical simulations of this tragedy have been reported in [14, 15]. Table 2 summarizes the parameters used in the simulation [14].

	terrain	water
ρ	$2.4\ 10^3\ Kg/m^3$	$10^3\ Kg/m^3$
ϕ	$26°$	-
ϕ_{basal}	$5°$	-
μ_0	$0.1\ Pa\ s$	$0.001\ Pa\ s$
h_{slip}	$500\ m$	-

Table 2: Vajont Landslide. Material parameters.

For the numerical analysis, a finite element mesh of 78137 nodes has been used with a fixed time step of 10^{-2} seconds. The effect of the slip boundary conditions can be appreciated in Figure 6, where two velocity profiles are plotted for varying slip length. As expected, it can be observed that a larger value of the slip length leads to a more homogeneous velocity profile, which is consistent with the almost rigid motion of the sliding mass observed in reality.

Figure 8 shows snapshots of the numerical simulation at different time instants while figure 7 shows a comparison between numerical results and post-event surveys of the maximum elevation reached by the wave on the right bank. A satisfactory agreement can be noted, with discrepancies mainly due to the coarse adopted discretization. The estimated overtopping height is of 100 meters, which closely matches the observed one.

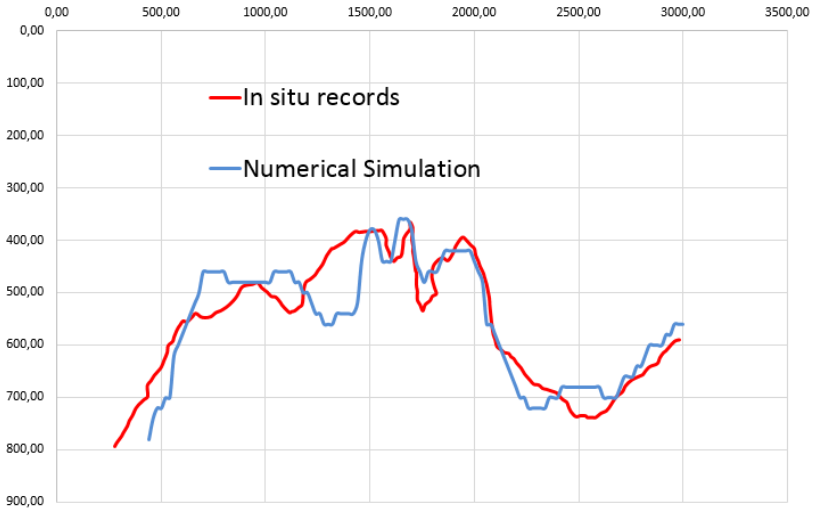


Figure 7: Vajont Slide. Maximum height of the wave reached in the right bank.

8 CONCLUSIONS

In the present work, a Lagrangian finite element method has been presented for the numerical simulation of landslide-water interaction. The landslide has been modeled as a viscoplastic non-Newtonian frictional fluid while water as a standard Newtonian fluid. Slip boundary conditions have been introduced between the landslide and the slope to better represent the real behavior of the basal interface. To validate the proposed approach, the proposed approach has been applied to two real landslides. The second example in particular concerned the simulation of a complex event, consisting of a landslide impinging in a water reservoir and generating a wave overtopping an existing concrete dam. Good accuracy has been obtained in the simulations.

REFERENCES

[1] Quecedo, M., Pastor, M. and Herreros, M.I. Numerical modelling of impulse wave generated by fast landslides. *International Journal for Numerical Methods in Engineering* (2004) **59**:1633–1656.

[2] Cremonesi, M., Frangi, A. and Perego, U. A Lagrangian finite element approach for the simulation of water-waves induced by landslides. *Computers and Structures* (2011) **89**:1086–1093.

[3] Cremonesi, Frangi, A. and Perego, U. A Lagrangian finite element approach for the analysis of fluidstructure interaction problems. *International Journal for Numerical Methods in Engineering* (2010) **84**:610–630.

[4] Cremonesi, M., Ferrara, L., Frangi, A. and Perego, U. Simulation of the flow of fresh cement suspensions by a Lagrangian finite element approach. *Journal of Non-Newtonian Fluid Mechanics* (2010) **165**:1555–1563.

[5] Oñate, E., Idelsohn, S.R., del Pin, F. and Aubry, R. The Particle Finite Element Method. An Overview. *International Journal Computational Method* (2004) **1**:267–307.

[6] Idelsohn, S.R., Oñate, E. and del Pin, F., The Particle Finite Element Method: A Powerful tool to Solve Incompresible Flows with Free-Surfaces and Breaking Waves. *International Journal for Numerical Methods in Engineering* (2004) **61**:964–989.

[7] Idelsohn, S.R., Oñate, E., del Pin, F. and Calvo, N., Fluid Structure Interaction Using the Particle Finite Element Method. *Computer Methods in Applied Mechanics and Engineering* (2006) **195**:2100–2123.

[8] Donea, J. and Huerta, A., Finite Element Methods for Flow Problems. *Wiley* (2005).

[9] Papanastasiou, T.C., Flows of materials with yield *Journal of Rheology* (1987) **31**:385–404.

[10] Cremonesi, M., Ferri, F., and Perego, U. A basal slip model for Lagrangian finite element simulations of 3D landslides. *submitted for publication*.

[11] Dawson, R F, Morgenstern, N R and Stokes, W Liquefaction flowslides in Rocky Mountain coal mine waste dumps. *Canadian Geotechnical Journal*. (1998) **35**:328–343.

[12] Pastor, M. and Blanc, T. and Haddad, B. and Petrone, S. and Sanchez Morles, M. and Drempetic, V. and Issler, D. and Crosta, G. B. and Cascini, L. and Sorbino, G. and Cuomo, S. Application of a SPH depth-integrated model to landslide run-out analysis. *Landslides*. (2014) **11**:793–812.

[13] Herrores, M. I. and Fernández Merodo, J. a. and Quecedo, M. and Mira, P. and Pastor, M. and GonzÁlez, E. Modelling tailings dams and mine waste dumps failures. *Géotechnique* (2002) **52**:579–591.

[14] Crosta, G.B., Imposimato, S. and Roddeman, D., Landslide Spreading, Impulse Water Waves and Modelling of the Vajont Rockslide. *Rock Mechanics and Rock Engineering* (2015):1-24.

[15] Vacondio, R., Mignosa,P. and Pagani, S. 3D SPH numerical simulation of the wave generated by the Vajont rockslide. *Advances in Water Resources* (2013) **59**:146 - 156.

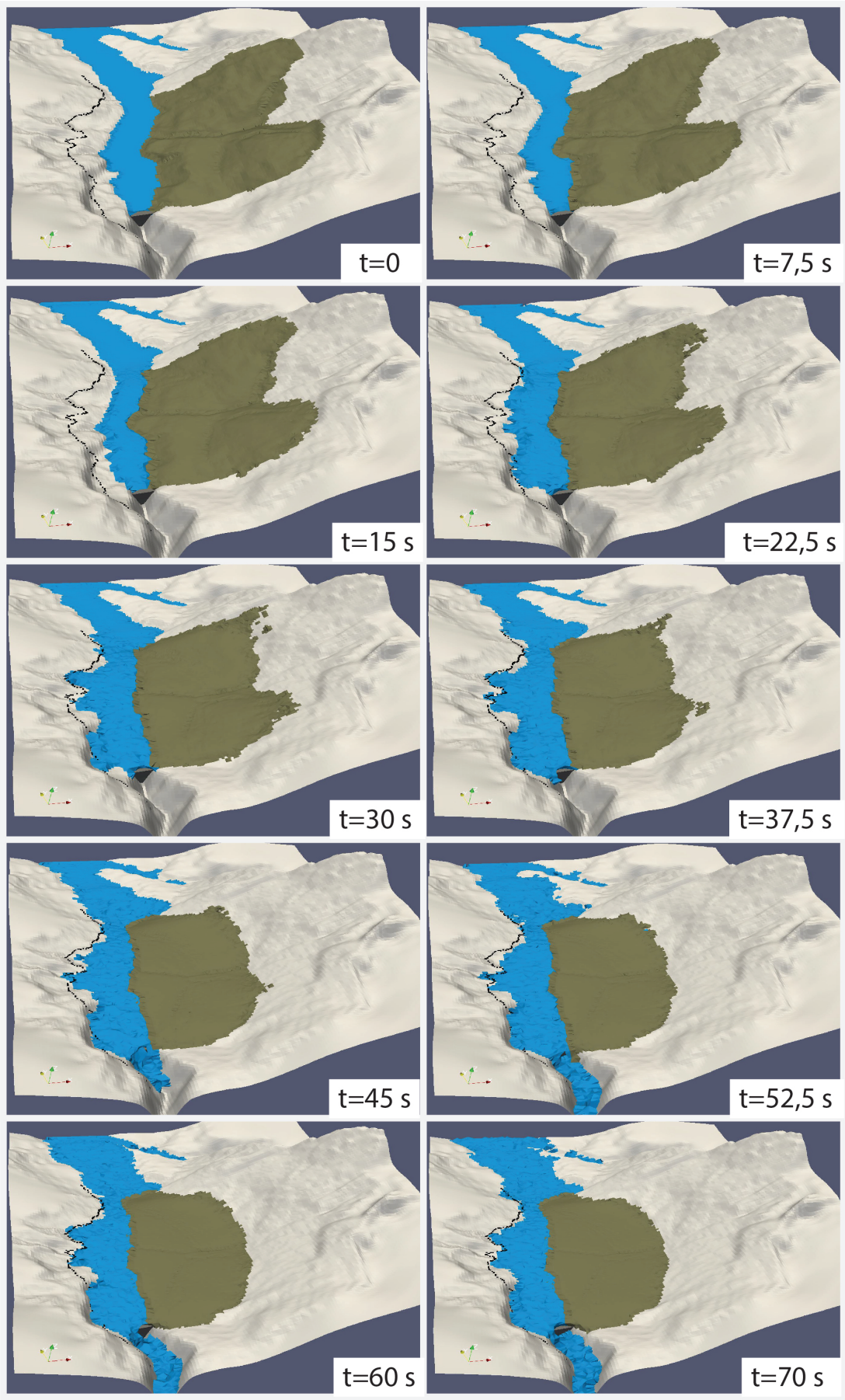


Figure 8: Vajont Slide. Snapshots of the simulation at different time instants. Black dashed contour denotes the observed level reached by the real wave.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A NUMERICAL INVESTIGATION OF FLOW DYNAMICS OVER A TRAPEZOIDAL SMOOTH OPEN CHANNEL

Giacomo Viccione

Department of Civil Engineering, University of Salerno, 84084, Fisciano, Italy
e-mail: gviccion@unisa.it

Keywords: Computational Fluid Dynamics, Flow dynamics, Free Surface Flows, Smoothed Particle Hydrodynamics, Volume of Fluid, HEC-RAS.

Abstract. *The paper presents a numerical investigation of the flow dynamics in a laboratory flume. The adopted geometry consists of a U shaped trapezoidal smooth open channel with a fixed slope and unerodible bed. The branches, 3m of length each, are linked with a joint 1m long, realizing two 90 degrees bends. The system is fed upstream with a water discharge under critical conditions while a Cipolletti weir is set downstream to control flow profiles. Steady flow characteristics are deduced by means of three different softwares: a pure Lagrangian developed by the author, based on the Weakly Compressible Smoothed Particle hydrodynamics (WCSPH) technique, Flow3D® and HEC-RAS®. Depending on the specific boundary conditions being given, velocity profiles and water interfaces at certain cross sections are deducted by using the first two codes. Results are discussed then spatially averaged in terms of mean velocities and water depths respectively to make a comparison with the ones obtained with HEC-RAS, showing a satisfactory agreement.*

1 INTRODUCTION

Experimental investigations of free surface flows, either laboratory or full-scale based, are aimed at describing complex fluid dynamics. They represent a viable tool for the design of structures interacting with liquids as well as for the verification under variable operating conditions. Needless to say, the related literature is quite consistent. Just to mention some meaningful examples, related studies concern the analysis of the interaction between fluids and solid [1,2] or porous interfaces [3-5], the assessment of the solid transport, either of bed or suspended or both types [6-8], sea waves [9,10], granular flow movements [11,12],

Although the effectiveness of experimental studies has been demonstrated by several investigations, the search for numerical solutions [13] is considered as a valuable alternative, as well as a supporting option [12], because it is cost-effective, easily adaptable to the variation of operating condition, more and more convenient for the endless increase of computer power. Numerical methods, either Lagrangian [14], Eulerian [15] or mixed [16] have been extensively applied over the last century as practical tools to give response to engineering and scientific problems. Among Lagrangian types, an emerging method, known as Smoothed Particle Hydrodynamics [17, 18], eventually coupled with others to exploit related advantages as in [19,20], has been applied since the last two decades, see for instance [21, 22].

The aim of this paper is to check the feasibility of three numerical models to describe the flow dynamics in a complex laboratory flume. Over the past three decades, two-dimensional and three-dimensional algorithms have been widely used to describe free surface flows, see for instance [23].

A full three-dimensional investigation is carried by mean of a Lagrangian code, based on the Weakly Compressible Smoothed Particle hydrodynamics (WCSPH) technique [24, 25] and the Eulerian-based software Flow3D® [26, 27]. Related results are compared in terms of velocity distribution and water profiles at certain cross sections. Average fields are then compared with the ones obtained using the one-dimensional computer program HEC-RAS [28, 29]. The algorithm based on the WCSPH technique and Flow3D® basics [30,31] are already described in [32], therefore they are not discussed here. Related applications are given in [33-36]. HEC-RAS (River Analysis System) is a freeware software for modeling open channel flows with the presence of bridges, culverts, weirs and other structures. Developed by the Hydrologic Engineering Center (HEC) of the United States, it is commonly used for engineering design, flood hazard mapping and flood insurance studies. Governing equations are one-dimensional, i.e. there is no hydraulic dependence on the cross section shape or channel bends.

2 MATERIALS AND METHODS

The trapezoidal open channel flow shown in the next Fig.1 is adopted in the present investigation. It is actually used for experimental investigations @ the Environmental and Maritime Hydraulics Laboratory (LIDAM) of the University of Salerno. As the flume is made of stainless steel, the roughness is neglected when carrying three dimensional studies. Manning's roughness is set as $n=0.012 s/m^{1/3}$ [37] when using HEC-RAS instead. The branches, 3m of length each, are linked with a joint 1m long and of trapezoidal shape as well, realizing two 90 degrees bends. The slope of each branch is kept fixed at 1%. The system is fed upstream with a water flow rate Q under critical conditions while a Cipolletti weir of height h measured from the bottom is set downstream to control flow profiles. Steady conditions can be reached in the reality thank to the presence of a recirculating system, consisting of a centrifugal pump which connect the outlet and inlet tanks. Here, steady state conditions are reached after a transitory which is of the order of tens of seconds. During such initial phase the inflow volume takes the space behind the weir: the higher is the latter, the longer is the transitory.

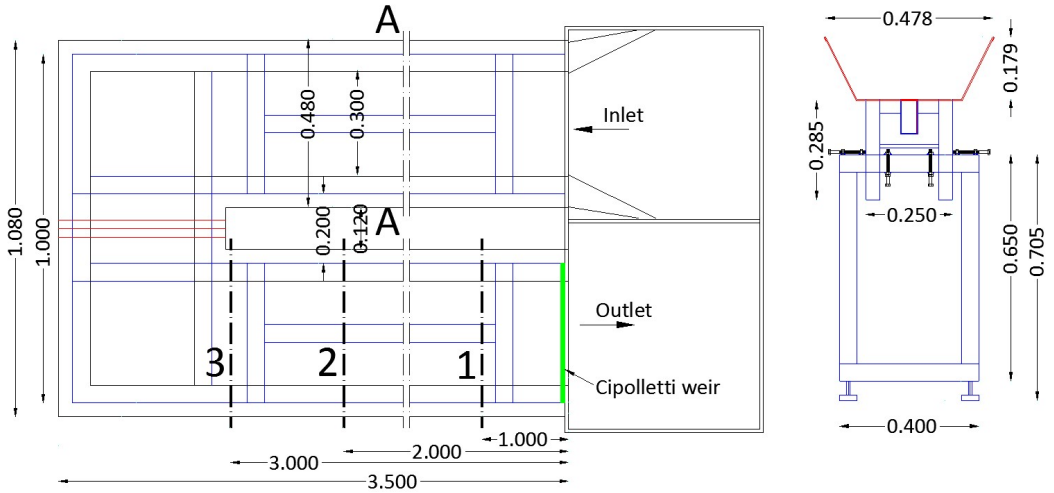


Figure 1: The open channel flow adopted in the present investigation. Left side: top view. Right side: vertical cross section A-A. 3D velocity field is extrapolated from vertical cross sections 1, 2 and 3.

Numerical simulations have been carried out taking the portion of the system between the Inlet and the Outlet zones, i.e. the flume itself. Related geometry was first created in a CAD environment as a collection of three-dimensional solids (Fig. 2, left side), then exported as a STereoLithography (stl) file, a standard compatible with Flow3D®. It basically contains a list of triangular surfaces that approximate the CAD model. The same file was finally converted in a text (xyz) file using TextPad®, obtaining the vertex coordinates of boundary particles (Fig. 2, right side), to be given as input to the WCSPH algorithm. The mesh size of the computing domain was set equal to 0,005m. The same value was adopted as reference for the edges of triangles listed in the stl file.

In the present study we consider the combination of two water discharges (Q_1=0.0085m^3/s, Q_2=0.0170m^3/s) set upstream and four weir's height (d_1=0.025m, d_2=0.05m, d_3=0.070m, d_4=0.010m) for a total of height scenarios. The upstream condition can be straightforwardly fixed in Flow3D® while in WCSPH it is set by fixing the velocity of the inflow particles.

3 RESULTS

This section yields the numerical results obtained with the WCSPH algorithm and the Flow3D® solver. To be concisely, a graphical comparison (Fig. 3) is given only for the simulations with Q_1=0.0085m^3/s and d_3=0.070m, in terms of velocity distribution over the cross sections 1,2 and 3 depicted in Fig. 1, left side. A general view of the dynamics is sketched in Fig. 4.

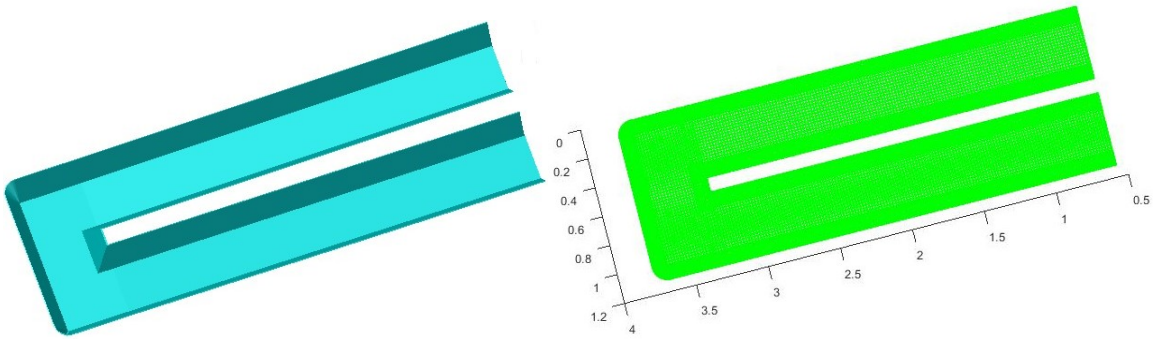
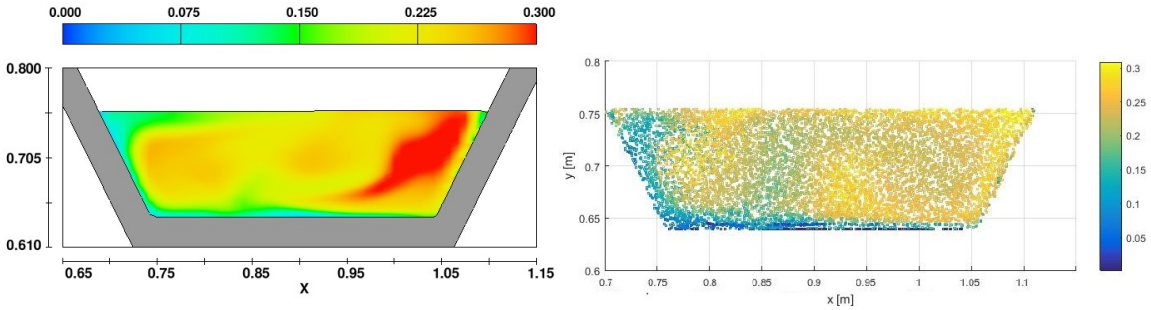
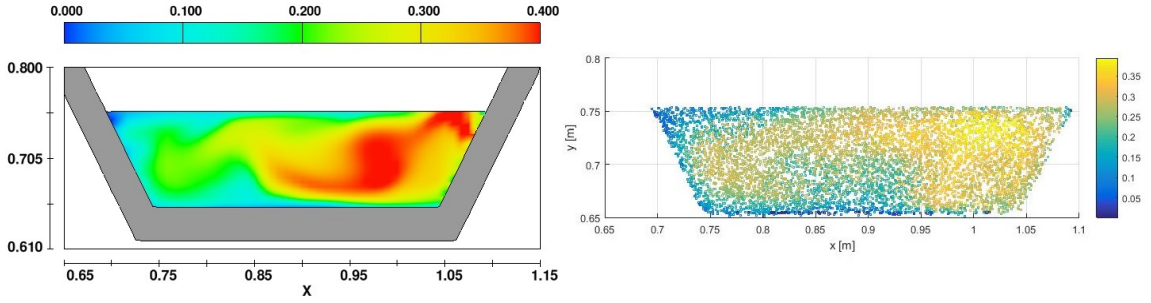


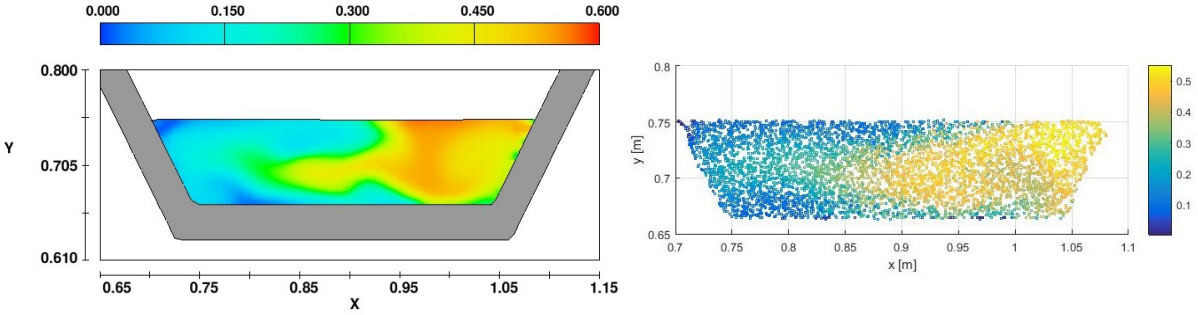
Figure 2: The flume was first created in a CAD environment (left side) then exported as STereoLithography file for Flow3D simulations and finally converted as boundary particles for the WCSPH algorithm.



Cross Section 1



Cross Section 2



Cross Section 3

Figure 3: The velocity distribution obtained with Flow3D® (left side) and WCSPH (right side) at vertical cross section 1, 2 and 3.

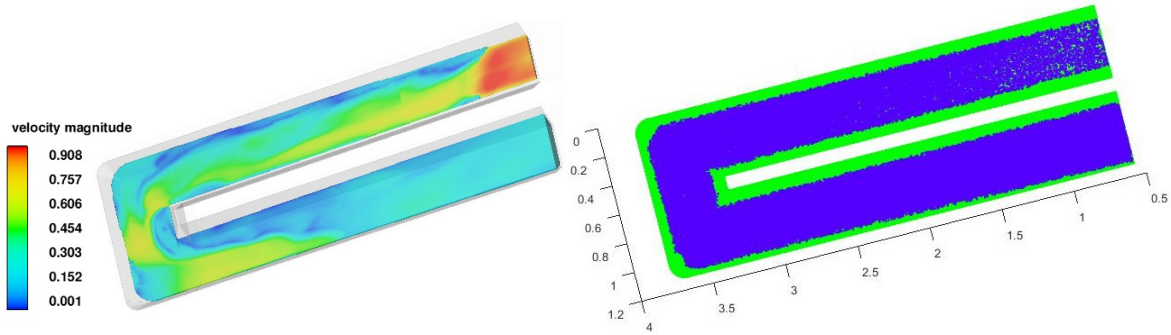


Figure 4: The flow dynamics for Q_1=0,0085m^3/s and h_3=0,070m. It is shown on the left side the velocity distribution obtained with Flow3D® while on the right side the particle disposition obtained with WCSPH.

From Fig. 3 and Fig. 4, left side, it can be observed that the trapezoidal joint influences the flow dynamics.

In particular, the velocity distribution at vertical cross sections is not symmetric respect to the vertical midplane of the branches. Higher velocity values are located on the right hand side of cross sections 1, 2 and 3 because the flow is constrained by the presence of the joint's corners. The phenomenon could be qualitatively compared to some extend to the liquid discharge issuing from a sharp orifice. Symmetry is partially recovered at Section 1 (i.e. the closest one to the weir) as it is the farthest section from what it can be thought as a disturbing region. The comparison of the velocity distribution and water table (the local free surface flow location) is satisfying. Both the numerical approaches basically reproduce the same flow dynamics.

As for the remaining simulations, the next summary Table 1, yield the velocities V_i and water surface elevations h_i, i=1,2,3 at cross sections (CS) 1,2 and 3, obtained by extracting the corresponding output values from HEC-RAS and by spatially averaging the local velocity distribution and the local free surface flow with reference to Flow3D and WCSPH simulations. Comparison among results is satisfying. It can be noted that the weir's height d_j, j=,1,2,3,4 effectively controls water profiles. Anyway, there is not a linear dependence of V_i and WS_i with d_j. This is basically due to the trapezoidal shape of the flume.

Q_1=8.5l/s	CS 1		CS 2		CS 3	
d_1=2.5cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.43**	**0.72**	**0.75**	**0.71**	**0.75**	**0.70**
Flow3D	**0.42**	**0.72**	**0.74**	**0.71**	**0.74**	**0.71**
HEC-RAS	***0.44***	***0.71***	***0.75***	***0.70***	***0.75***	***0.71***

Q_2=17.0l/s	CS 1		CS 2		CS 3	
d_1=2.5cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.56**	**0.74**	**0.67**	**0.76**	**0.62**	**0.75**
Flow3D	**0.57**	**0.74**	**0.68**	**0.76**	**0.61**	**0.76**
HEC-RAS	***0.58***	***0.73***	***0.70***	***0.74***	***0.63***	***0.75***

Q_1=8.5l/s	CS 1		CS 2		CS 3	
d_2=5.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.29**	**0.74**	**0.36**	**0.74**	**0.43**	**0.74**
Flow3D	**0.28**	**0.74**	**0.37**	**0.74**	**0.43**	**0.74**
HEC-RAS	***0.29***	***0.73***	***0.37***	***0.73***	***0.44***	***0.73***

Q_2=17.0l/s	CS 1		CS 2		CS 3	
d_2=5.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.40**	**0.76**	**0.46**	**0.76**	**0.54**	**0.76**
Flow3D	**0.40**	**0.76**	**0.48**	**0.77**	**0.54**	**0.77**
HEC-RAS	***0.41***	***0.76***	***0.48***	***0.76***	***0.55***	***0.76***

Q_1=8.5l/s	CS 1		CS 2		CS 3	
d_3=7.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.22**	**0.76**	**0.27**	**0.76**	**0.31**	**0.75**
Flow3D	**0.23**	**0.76**	**0.28**	**0.76**	**0.33**	**0.75**
HEC-RAS	***0.23***	***0.75***	***0.28***	***0.75***	***0.32***	***0.75***

Q_2=17.0l/s	CS 1		CS 2		CS 3	
d_3=7.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.34**	**0.78**	**0.39**	**0.79**	**0.44**	**0.78**
Flow3D	**0.34**	**0.79**	**0.39**	**0.78**	**0.45**	**0.79**
HEC-RAS	***0.35***	***0.78***	***0.40***	***0.78***	***0.45***	***0.78***

Q_1=8.5l/s	CS 1		CS 2		CS 3	
d_4=10.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.17**	**0.79**	**0.19**	**0.77**	**0.22**	**0.79**
Flow3D	**0.16**	**0.79**	**0.19**	**0.77**	**0.21**	**0.79**
HEC-RAS	***0.17***	***0.78***	***0.20***	***0.78***	***0.22***	***0.78***

Q_2=17.0l/s	CS 1		CS 2		CS 3	
d_4=10.0cm	V_1	h_1	V_2	h_2	V_3	h_3
SPH	**0.27**	**0.82**	**0.30**	**0.82**	**0.33**	**0.81**
Flow3D	**0.28**	**0.82**	**0.30**	**0.81**	**0.33**	**0.82**
HEC-RAS	***0.28***	***0.81***	***0.31***	***0.81***	***0.34***	***0.81***

Table 1: Mean velocities [m/s] and water surface elevations [m] obtained from Flow3D and WCSPH results. A comparison is made with corresponding values (in italic) obtained with HEC-RAS, showing a satisfactory agreement.

4 CONCLUSIONS

The capability of three numerical approaches to reproduce the flow dynamics over a U shaped trapezoidal smooth open channel with a fixed slope and unerodible bed is here discussed. Three-dimensional results obtained by running Flow3D and WCSPH revealed that

the velocity distribution at vertical cross sections is not symmetric respect to the vertical mid-plane of channel branches because it is influenced by the presence of the channel joint. Mean velocities and water levels at three cross sections were compared with corresponding values obtained with HEC-RAS, showing a satisfactory agreement.

REFERENCES

[1] G. Viccione, M. Genovese, F. Rossi, D. Guida, T.L.L. Lenza, Physical Modelling of Laboratory Debris Flows by Using the Sodium Carboxymethylcellulose (Na-CMC), *WSEAS Transactions on Fluid Mechanics*, 2015, **10**, 164-174.

[2] T. Mai, Z.Z. Hu, D. Greaves, A. Raby, Investigation of hydroelasticity: Wave impact on a truncated vertical wall, *Proceedings of the International Offshore and Polar Engineering Conference*, 647-654, 2015.

[3] P. Contestabile, F. Aristodemo, D. Vicinanza, P. Ciavola, Laboratory study on a beach drainage system, *Coastal Engineering*, **66**, 50-64, 2012.

[4] P. Ciavola, D. Vicinanza, F. Aristodemo, P. Contestabile, Large-scale morphodynamic experiments on a beach drainage system, *Journal of Hydraulic Research*, 2011, **49**(4), 523-528.

[5] K. Harada, F. Imamura, T. Hiraishi, Experimental Study on the Effect in Reducing Tsunami by the Coastal Permeable Structures, *Proceedings of the International Offshore and Polar Engineering Conference*, 652-658, 2002.

[6] D. Termini, River-bed erosion due to changing boundary conditions: Performance of a protective measure, *IAHS-AISH Proceedings and Reports*, **364**, 112-117, 2014,

[7] D. Termini, Bed scouring downstream of hydraulic structures under steady flow conditions: Experimental analysis of space and time scales and implications for mathematical modeling, *Catena*, **84**(3), 125-135, 2014.

[8] M. Guerrero, The investigation of sediment processes in rivers by means of the Acoustic Doppler Profiler, *IAHS-AISH Proceedings and Reports*, **364**, 368-373, 2014.

[9] D. Vicinanza, F. Dentale, D. Salerno, M. Buccino, Structural response of seawave Slot-Cone Generator (SSG) from Random Wave CFD simulations, *Proceedings of the International Offshore and Polar Engineering*, 985-991, 2015.

[10] L.F.C. Lopez, D. Salerno, F. Dentale, A. Capobianco, M. Buccino, Experimental campaign on the overtopping of the seawall Malecòn Tradicional, *Proceedings of the International Offshore and Polar Engineering*, 1499-1505, 2015.

[11] L. Sarno, M.N. Papa, R. Martino, Dam-break flows of dry granular material on gentle slopes, *International Conference on Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment*, 503-512, 2011.

[12] S. Evangelista, G. De Marinis, C. Di Cristo, A. Leopardi, Dam-break dry granular flows: Experimental and numerical analysis, WSEAS Transactions on Environment and Development, **10**(1), 382-392, 2014.

[13] R.W. Yeung, Numerical methods in free-surface flows, *Annual Review of Fluid Mechanics*, **14**, 395-442, 1982.

[14] G. Viccione, V. Bovolin, E. Pugliese Carratelli, Simulating Flows with SPH: Recent Developments and Applications, Book ch. in: *Intech Hydrodynamics - Optimizing Methods and Tools*, 69-84, 2011.

[15] F. Dentale, G. Donnarumma, E. Pugliese Carratelli, Numerical wave interaction with tetrapods breakwater, *International journal of naval architecture and ocean engineering*, **6**, 1-13, 2014.

[16] M. Souli, A. Ouahsine, L. Lewin, ALE formulation for fluid–structure interaction problems, *Computer Methods in Applied Mechanics and Engineering*, **190**(5–7), 659–675, 2000.

[17] J.J. Monaghan, Smoothed Particle Hydrodynamics, *Reports on Progress in Physics*, **68**(8), 1703–1759, 2005.

[18] D. Violeau, B.D. Rogers, Smoothed particle hydrodynamics (SPH) for free-surface flows: Past, present and future, Journal of Hydraulic Research, 54(1), 1-26, 2016.

[19] D.M. Robb, S.J. Gaskin, J.-C. Marongiu, SPH-DEM model for free-surface flows containing solids applied to river ice jams, *Journal of Hydraulic Research*, 54 (1), 27-40, 2016.

[20] C. Altomare, J.M. Domínguez, A.J.C. Crespo, (...), I. Caceres, M. Gómez-Gesteira, Hybridization of the Wave Propagation Model SWASH and the Meshfree Particle Method SPH for Real Coastal Applications, *Coastal Engineering Journal*, **57**(4), 1550024, 2015.

[21] M. Gómez-Gesteira, B.D. Rogers, R.A. Dalrymple, A.J.C. Crespo State-of-the-art of classical SPH for free-surface flows, *J. Hydraul. Res.*, **48** Extra Issue, 6–27, 2010.

[22] G. Viccione, V. Bovolin, E.P ugliese Carratelli, Defining and optimizing algorithms for neighbouring particle identification in SPH fluid simulations, *International Journal for Numerical Methods in Fluids*, **58**, 625-638, 2008.

[23] M.B. Kocyigit, R.A. Falconer, B. Lin, Three-dimensional numerical modelling of free surface flows with non-hydrostatic pressure*, International Journal for Numerical Methods in Fluids*, **40**(9) , 1145-1162, 2002.

[24] A. Amicarelli, R. Albano, D. Mirauda, (...), A. Sole, R. Guandalini, A Smoothed Particle Hydrodynamics model for 3D solid body transport in free surface flows, *Computers and Fluids*, **116**, 205-228, 2015.

[25] D. Le Touzé, A. Colagrossi, G. Colicchio, M. Greco, A critical investigation of smoothed particle hydrodynamics applied to problems with free-surfaces, *International Journal for Numerical Methods in Fluids*, **7**(7), 660-691, 2013.

[26] F. Dentale, G. Donnarumma, E. Pugliese Carratelli, Simulation of flow within armour blocks in a breakwater, *J. Coastal Res*., **30**(3), 528-536, 2014.

[27] P. Ghadimi, A. Dashtimanesh, M. Farsi, S. Najafi, Investigation of free surface flow generated by a planing flat plate using smoothed particle hydrodynamics method and FLOW3D simulations, *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, **22**(2), 125-135, 2013.

[28] G.W. Brunner, HEC-RAS, River Analysis System, User's Manual, version 3.0. US Army Corps of Engineers, Davis, California, USA , 2001.

[29] G. Di Baldassarre, A. Montanari, Uncertainty in river discharge observations: A quantitative analysis, Hydrology and Earth System Sciences, **13**(6), 913-921, 2009

[30] Violeau D., Fluid Mechanics and the SPH Method: Theory and Applications, Oxford University Press, 2012.

[31] Flow Sciences Inc. Flow-3D User Manual, release 9.4, Santa Fe, NM, USA, 2009.

[32] G. Viccione, E. Pugliese Carratelli, V. Bovolin, A numerical investigation of liquid impact on planar surfaces, VII European Congress on Computational Methods in Applied Sciences and Engineering, Crete Island, Greece, 5–10 June 2016, under review.

[33] E. Pugliese Carratelli, G. Viccione, V. Bovolin, Free surface flow impact on a vertical wall: a numerical assessment, *Theoretical and Computational Fluid Dynamics*, accepted for publication, 2016.

[34] G. Viccione, V. Bovolin, E. Pugliese Carratelli, Simulating fluid-structure interaction with SPH. In: *International Conference of Numerical Analysis and Applied Mathematics*, Kos, Greece, Vol.1479, 209-212, 2012.

[35] G. Viccione, V. Bovolin, Simulating Triggering and Evolution of Debris-Flows with Smoothed Particle Hydrodynamics (SPH), *Italian Journal of Engineering Geology and Environment*, 523-532, 2011.

[36] L. Cavallaro, F. Dentale, G. Donnarumma, E. Foti, R.E. Musumeci, E. Pugliese Carratelli, Rubble Mound reakwater Overtopping: Estimation of the reliability of a 3D Numerical Simulation, In: *ICCE 2012, International Conference on Coastal Engineering*, Santander, Spain, 2012.

[37] V.T. Chow, Open-Channel Hydraulics, McGraw-Hill, 1959.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A NUMERICAL INVESTIGATION OF LIQUID IMPACT ON PLANAR SURFACES

Giacomo Viccione[1], Vittorio Bovolin[1], and Eugenio Pugliese Carratelli[2]

[1] Department of Civil Engineering, University of Salerno, 84084, Fisciano, Italy
e-mail: {gviccion,v.bovolin}@unisa.it

[2] InterUniversity Centre for the Prediction and Prevention of Major Hazards (C.U.G.RI.), 84084, Fisciano, Italy
epc@unisa.it

Keywords: Fluid Structure Interaction, Computational Fluid Dynamics, Liquid Impact, Free Surface Flows, Smoothed Particle Hydrodynamics, Volume of Fluid.

Abstract. *The impact of a free surface liquid flow on planar surfaces is a common occurrence and a challenging issue in many research fields. There are many situations in science and technology where this field of study finds related applications: the falling objects on a liquid surface, sloshing dynamics, flow run-up and overtopping, the action of train waves on maritime structures. Under certain circumstances, the impact process may result into high, spatially localized pressure peaks, thus inducing dangerous solicitations. The present work focuses on some relevant computational aspects of the fluid impact on inclined planar surfaces, making use of the Weakly Compressible Smoothed Particle Hydrodynamics (WCSPH) Lagrangian technique. With reference to the early stages of the impact process, pressure distribution is described as function of the incident wave's features and the angle of incidence of the solid surface assumed. Results are then discussed and compared with the corresponding ones obtained via Eulerian software.*

1 INTRODUCTION

Assessing the interaction of fluids and structures (FSI) is a topic of great interest in nature and science. The literature is huge, therefore no attempt is made here to review it. Some general, yet worth examples are reported in [1-4]. Among so many applications, the propagation and impact of steep sea waves on coastal structures [5-7] or the run-up and overtopping on defense structures [8, 9] have probably been studied in more detail starting from the pioneering work of by Kirkgöz [6], therefore they can provide some deeper insight.

Over the last few decades the prediction of impact pressures have been made by some developed theories. Peregrine and co-workers [10] proposed the "Pressure-impulse" approximation where the pressure is considered within a significant time window. The impulsive pressure field is therefore used for the derivation of the velocity field as the impact takes place. Korobkin proposed the "acoustic approximation" in [11], where the pressure distribution is evaluated analytically for a cylindrical jet impacting on a rigid plane and the "Asymptotic" assumption in [12] where the liquid-wall interaction is analyzed with the method of matched asymptotic expansions, through which properties such as compressibility are taken into account. Recently a semi-analytical model based on the Wagner theory has recently proposed in [13].

Current literature provides many contributions about full scale, laboratory or empirical investigations as well. More than a century went by since the publication of the original work of Stevenson [14] in which conditions which affect the force of impacting waves are discussed. In [15], an extensive campaign of pressure measurements on a vertical wall were performed at the Deltares laboratory of Delft (Holland) under wave impact. In [16] a large number of field measurements were carried with the aim of detecting most violent wave impacts on the Admiralty breakwater, Alderney (United Kingdom). Bullock and co-workers [17] measured a large number of impacts, showing that in some – rare – circumstances, local pressures are comparable with corresponding ones obtained from the water hammer model [18,19]. Blackmore and Hewson [20] recorded measurements of full-scale wave impact pressures on seawalls over a period of about four years. They derived an empirical expression for wave impact pressure that takes into account the concentration of entrapped air.

Lagrangian [21], Eulerian [22] or mixed [23] numerical methods have been extensively developed and applied as well. In [24] the impact process resulting from the interaction of breaking waves with a vertical wall was numerically solved by means of a Finite Difference (FD) scheme, based on the Volume of Fluid (VOF) method [25]. Among Lagrangian types, an emerging method, known as Smoothed Particle Hydrodynamics [26], has been applied in the recent past, see for instance [27, 28].

The aim of this paper is the study of the interaction between an approaching steep wave and a planar surface by means of two numerical approaches: a Lagrangian one, based on Weakly Compressible Smoothed Particle Hydrodynamics (WCSPH) technique and an Eulerian one, making use of the Flow3D ® commercial package.

2 THE WCSPH MODEL

The Weakly Compressible Smoothed Particle Hydrodynamics technique (WCSPH) is based on integral interpolants using kernels that approximate the delta function. See [21, 26, 29, 30] for up-to-date reviews and related applications. SPH is a mesh-free Lagrangian method based on computing particles discretizing the evolving domain. Hydrodynamic properties on a location, either fixed or moving with a particle, depend on the local particle neighbourhood where the kernel W is defined (Fig.1).

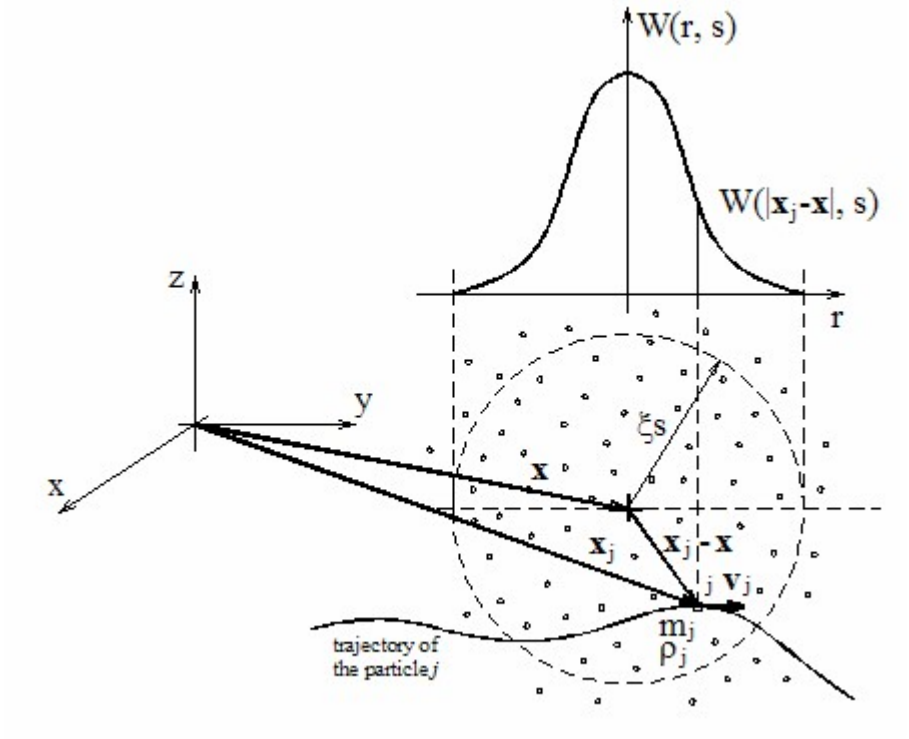


Figure 1: Hydrodynamic computation at the location **x** on the basis of neighbouring particles within a circular (2D) or spherical (3D) compact support where the interpolating kernel is defined. A neighbouring particle j, with mass m_j, density ρ_j and velocity $\mathbf{v}_j$ is sketched along with its trajectory.

Despite its inception in 1977 [31,32], we found hydrodynamic applications of SPH only since mid-1990s [33] and specific investigations of FSI since 2000s [34,35]. Recently, SPH has been considered as a valuable method for solving real-life problems in coastal engineering as in [36] due its intrinsic capability to deal with large deformation and complex geometries. Altomare et al., [37] carried advanced investigations of wave interaction with armour block sea breakwater.

In this work, mass and momentum equations

$$-\frac{1}{\rho}\frac{\mathrm{D}\,\rho}{\mathrm{D}t} = \nabla\cdot\underline{v} \tag{1}$$

$$\rho\frac{D\underline{v}}{\mathrm{D}t} = -\nabla p + \mu\nabla^2\underline{v} + \underline{f} \tag{2}$$

are discretized as in [18] yielding the numerical scheme

$$\begin{aligned}\frac{\mathrm{D}\log(v_i)}{\mathrm{D}t} &= \sum_{j=1}^{N_i}\left(\underline{v}_i - \underline{v}_j\right)\cdot\nabla_i W_{ij}\cdot\mathrm{d}\Omega_j + \\ &+ \xi\, h\, c_0\sum_{j=1}^{N_i}\psi_{ij}\cdot\nabla_i W_{ij}\cdot\mathrm{d}\Omega_j\end{aligned} \tag{3}$$

$$\frac{\mathrm{D}\underline{v}_\mathrm{i}}{\mathrm{Dt}} = -\sum_{\mathrm{j=1}}^{\mathrm{N_i}} \left(\frac{p_\mathrm{i}}{\rho_\mathrm{i}} + \frac{p_\mathrm{j}}{\rho_\mathrm{j}} + 19.6\nu \left(\frac{\underline{v}_\mathrm{ij} \cdot \underline{r}_\mathrm{ij}}{r_\mathrm{ab}{}^2 + \varepsilon \bar{h}_\mathrm{ij}^2} \right) \right) \nabla_\mathrm{i} W_\mathrm{ij} \mathrm{d}\Omega_\mathrm{j} + \underline{f}_\mathrm{i} \tag{4}$$

where i denotes the generic moving particle, j refers to one of its N_i neighbours (see Fig. 1), v_i is the specific volume $1/\rho_i$, defined as

$$\frac{\mathrm{D}\log(\mathrm{v_i})}{\mathrm{D}t} = -\frac{1}{\rho_\mathrm{i}} \frac{\mathrm{D}\,\rho_\mathrm{i}}{\mathrm{D}t} = \nabla \cdot \underline{v}_\mathrm{i} \tag{5}$$

ρ is the density, p is the pressure, $\underline{v}$ is the velocity vector, f is the external force, W is the above mentioned kernel or weighting function, defined over a compact support of radius r=2h, being $h=1.3d_0$ the smoothing length proportional to the initial interparticle distance d_0 = 0.005m, $\nu = 10^{-6}$ m^2/s is the kinematic viscosity of water at 20°C, c_0 is the speed of sound in the case of no compression.

Ruling equations (3,4), constrained with (5) and coupled with the Tait equation of state

$$p_\mathrm{i} = \frac{c_0^2 \rho_0}{\gamma} \left[\left(\frac{\rho_\mathrm{i}}{\rho_0} \right)^\gamma - 1 \right] \tag{6}$$

being $\gamma = 7$, are then solved in time with a second order predictor – corrector step scheme, already tested in a variety of situations, see for instance [38,39]

3 THE FLOW3D SOLVER

Flow3D® solver [40] is a CFD commercial software based onto a Finite Volume formulation of the Navier Stokes equations (eqs. 1-2) in a Eulerian framework. Free surfaces and interfaces are solved with the volume of fraction (VOF) method and the Fractional Area/Volume Obstacle Representation (FAVOR). Velocity and pressure fields are coupled by using the time-advanced velocities in the continuity equations and time-advanced pressures in the momentum equations. The model has been widely validated over the years particularly in connection with wave impact, see for instance [41, 42].

Water compressibility is treated by the acoustic approximation, which links the density variation to the pressure increase by the following $c^2 = \partial p / \partial \rho$. All the tests shown in the following have been carried out by assuming a fluid with a reference density ρ_0 = 1000 kg/m^3 and a reference sound speed c_0 = 1400m/s, as in the SPH model. The chosen values, representing realistic liquids, satisfy the acoustic condition $|\delta\rho/\rho| < 0.1$ which stands for low Mach numbers as it is of order 10^{-2} or less.

4 THE NUMERICAL SET-UP

An open channel water flow with an initial velocity v_0 and liquid height h_0 = 0.50m is assumed to suddenly impact against a planar surface (Fig. 2). The angle of the impacting surface, measured counterclockwise with respect to the vertical direction, ranges between -45° and +45° (Fig. 3), for a total of 19 geometry configurations. Therefore, the incremental angle between two consecutive configurations is equal to 5°. For computational purposes, a 4.00m long volume of fluid was assumed to be close to the solid interface.

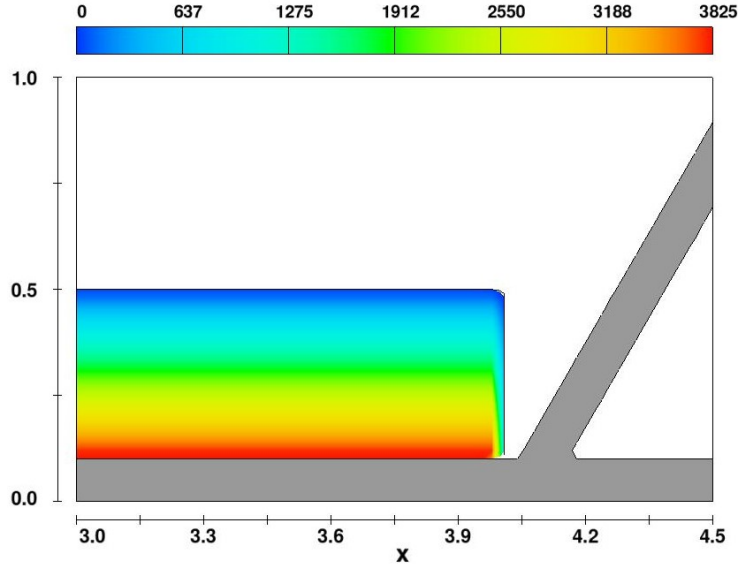


Figure 2: Initial conditions set in WCSPH and Flow3D (picture taken from Flow3D).
Colour bar refers to the pressure field which distribution is assumed to be initially hydrostatic.

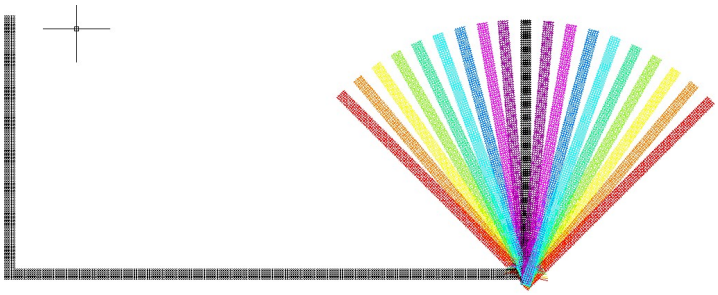


Figure 3: The inclined surface, the wave is impacting with, has an angle with respect the vertical direction ranging between -45° and +45°. Nineteen configurations are therefore assumed in the present work (picture represents solid boundaries given in WCSPH)

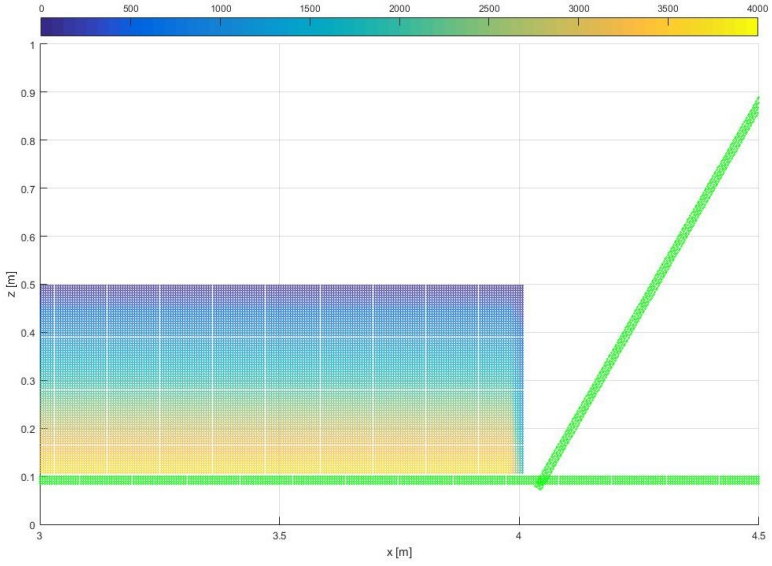
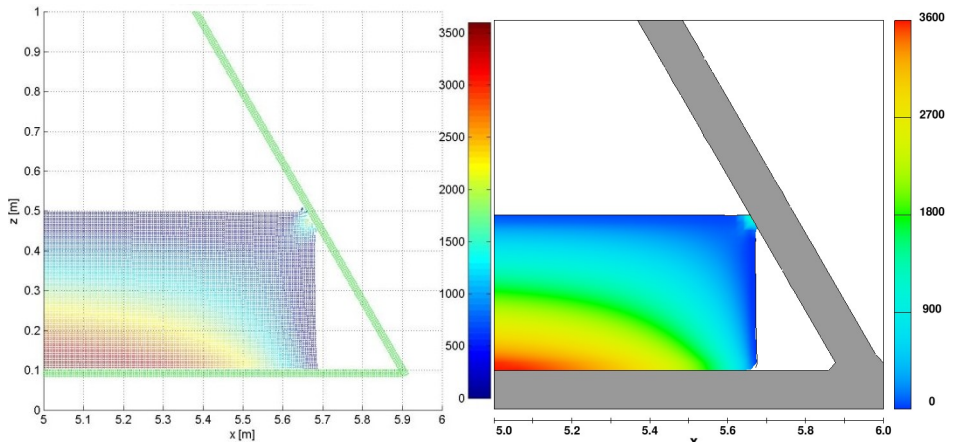


Figure 4: The fluid domain represented by particles in WCSPH. Solid walls are depicted by green particles. Colour bar refers to the pressure field which distribution is assumed to be initially hydrostatic. Conditions are equivalent as those sketched in Figure 2.

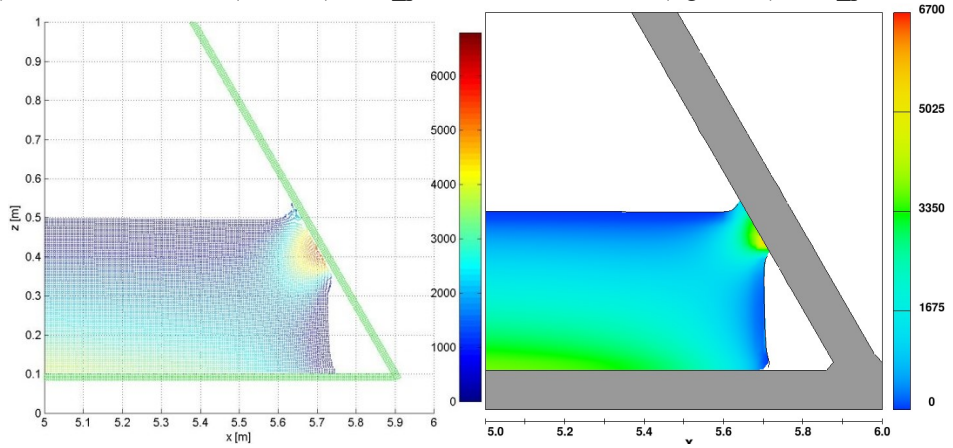
The computation domain is split into square 0.01m wide grid elements in Flow3D environment. Total number of cells is of order of 10^4, depending on how large is the simulating space between left and right walls (see Fig. 3). The fluid domain is discretized into $8 \cdot 10^4$ particles, with an initial interparticle distance $d_0 = 0.005$m (see Fig. 4, previous page).

5 RESULTS

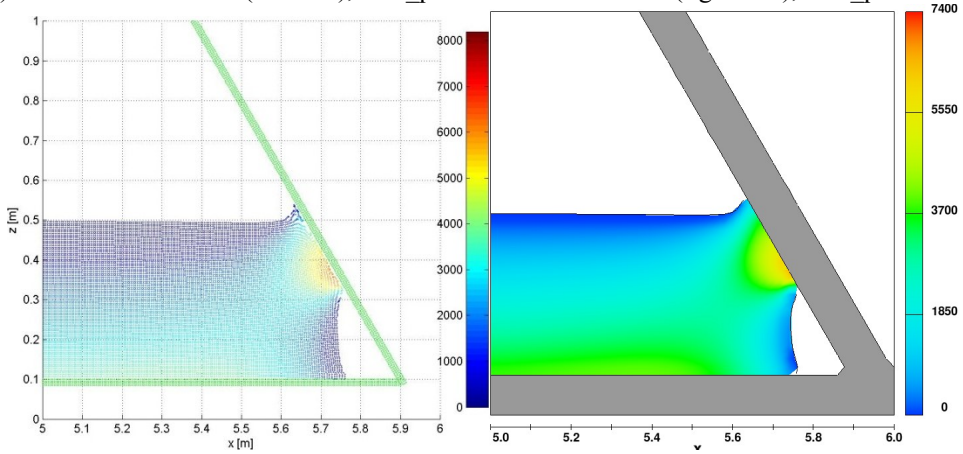
The impact process is here analyzed for the nineteen configurations above introduced (Fig. 2).



a) t=0.02sec. WCSPH (left side), max_p=3524Pa and Flow3D (right side), max_p=3578Pa.



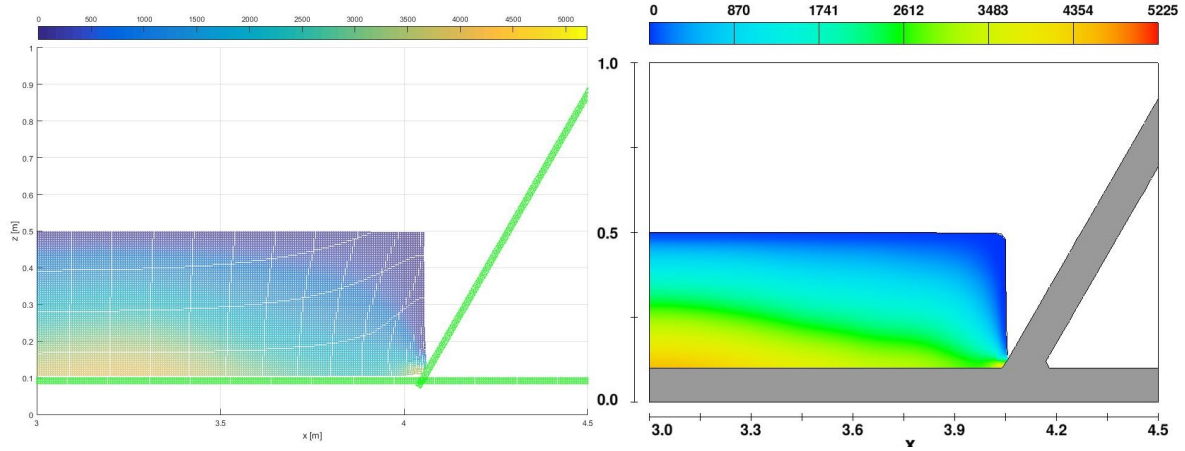
b) t=0.035sec. WCSPH (left side), max_p=6824Pa and Flow3D (right side), max_p=6659Pa.



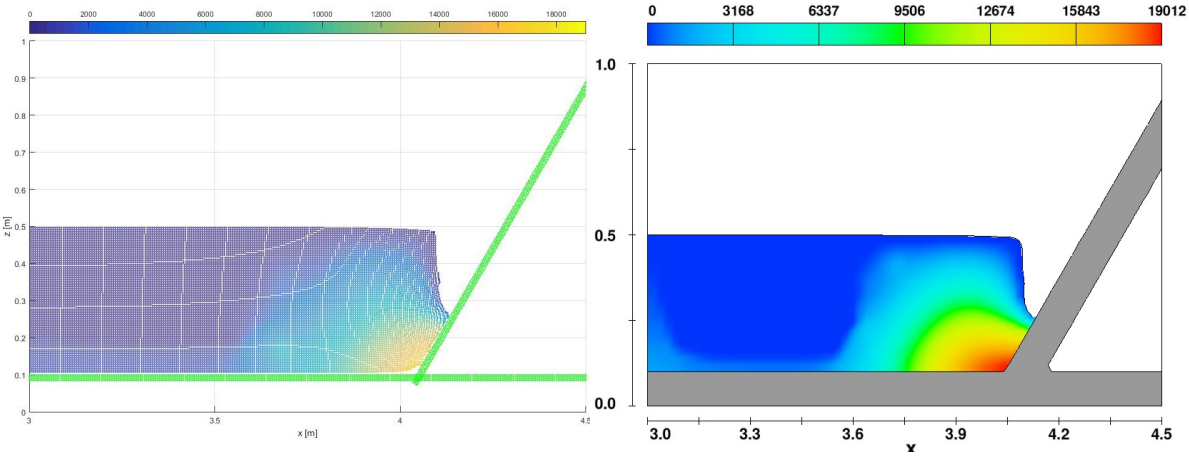
c) t=0.050sec. WCSPH (left side), max_p=8145Pa and F3D (right side), max_p=7482Pa.

Figure 5: Comparison of pressure fields for the initial velocity v0 = 2m/s. inclined wall angles α=+45°

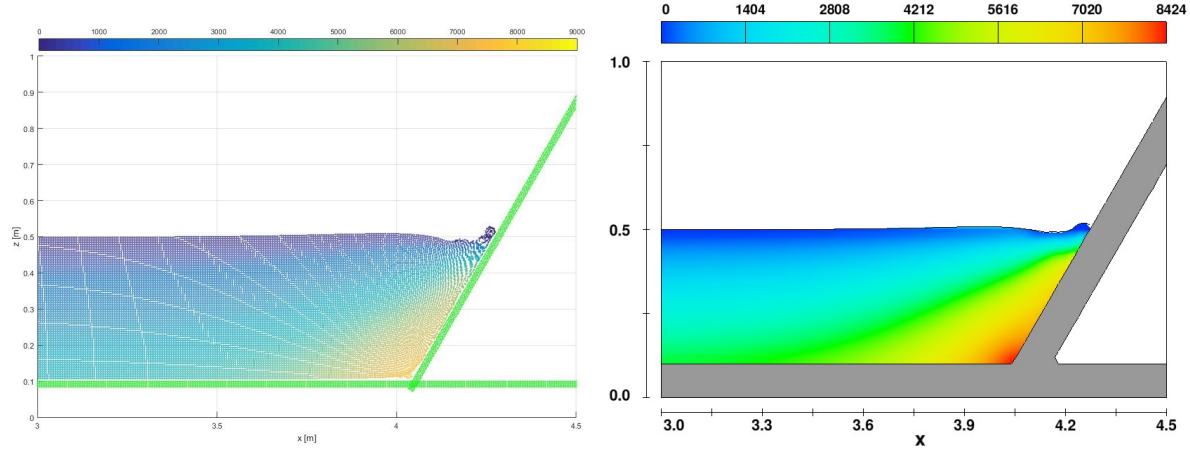
For each configuration, we consider an initial fluid mass advancing at velocities v_0=2m/s, 4m/s or 6m/s. For the sake of clarity and to be concisely, only results with v_0=2m/s and inclined wall angles α=+45° (Fig. 5) and -45° (Fig. 6) are sketched for some significant time instants. The summary Table 1 yields maximum pressures attained at the wall for each simulation being carried out with Flow3D and WCSPH.



a) t=0.02sec. WCSPH (left side), max_p=5158Pa and Flow3D (right side), max_p=5225Pa.



b) t=0.04sec. WCSPH (left side), max_p=18746Pa and Flow3D (right side), max_p=19012Pa.



c) t=0.080sec. WCSPH (left side), max_p=8924Pa and F3D (right side), max_p=8424Pa.

Figure 6: Comparison of pressure fields for the initial velocity v_0 = 2m/s. inclined wall angles α=-45°.

v_0=2m/s	+45°	+40°	+35°	+30°	+25°	+20°	+15°	+10°	+5°	0°
SPH	**8145**	**9057**	**9528**	**10279**	**10175**	**10879**	**12076**	**13249**	**11905**	**14212**
Flow3D	**7482**	**7635**	**8386**	**9714**	**10988**	**10460**	**11303**	**13007**	**13628**	**12200**

v_0=2m/s	-5°	-10°	-15°	-20°	-25°	-30°	-35°	-40°	-45°
SPH	**14020**	**13527**	**15167**	**14506**	**15177**	**16419**	**16813**	**17812**	**18746**
Flow3D	**14304**	**14223**	**16504**	**15446**	**18062**	**15860**	**17802**	**19455**	**19012**

v_0=4m/s	+45°	+40°	+35°	+30°	+25°	+20°	+15°	+10°	+5°	0°
SPH	**34127**	**36623**	**36874**	**37934**	**41188**	**44703**	**42685**	**47103**	**51634**	**57794**
Flow3D	**35092**	**38570**	**43283**	**40688**	**42289**	**45736**	**52629**	**58442**	**63862**	**63974**

v_0=4m/s	-5°	-10°	-15°	-20°	-25°	-30°	-35°	-40°	-45°
SPH	**49490**	**61920**	**57162**	**60870**	**60410**	**63688**	**67250**	**72083**	**71240**
Flow3D	**59970**	**68335**	**71138**	**74727**	**73549**	**72211**	**88151**	**81375**	**86762**

v_0=6m/s	+45°	+40°	+35°	+30°	+25°	+20°	+15°	+10°	+5°	0°
SPH	**116461**	**135041**	**129161**	**135434**	**152353**	**159675**	**151696**	**159429**	**170968**	**191092**
Flow3D	**82098**	**92074**	**101800**	**97465**	**114190**	**111149**	**121455**	**145804**	**149615**	**156696**

v_0=6m/s	-5°	-10°	-15°	-20°	-25°	-30°	-35°	-40°	-45°
SPH	**165619**	**177112**	**200596**	**182267**	**204412**	**216512**	**222823**	**242878**	**231928**
Flow3D	**158871**	**177986**	**171353**	**186598**	**191367**	**209781**	**199594**	**227296**	**214957**

Table 1: Detected maximum pressure values as function of the impacting velocity v_0 and surface's inclination α.

As can be seen from the above Table 1, the Eulerian and the Lagrangian approaches yield maximum pressure values in a good agreement. In addition, from the following Fig. 7 it is possible to derive that, under the operating limits of the present investigation, the maximum pressure is almost linear with the surface's angle, keeping fixed the impacting velocity v_0. This is mainly due to the characteristics of the impacting area: for positive surface's angles the contact region is located in the upper part of the liquid domain. As a consequence the liquid, once in contact with the solid surface, is free to spread both sides along it. In contrast, for negative surface's angles the contact region is located in the lower part of the liquid domain. In this case the interacting liquid is constrained to move only ahead along the inclined surface as it opposes to the remaining approaching mass.

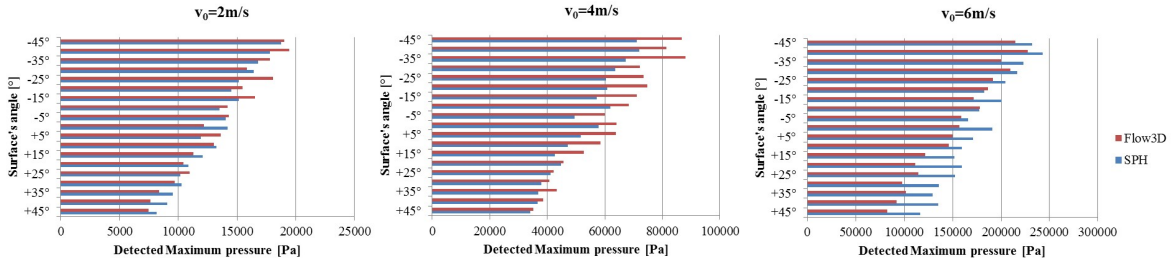


Figure 7: The Maximum detected pressures.

6 CONCLUSIONS

The interaction between an approaching liquid mass and an inclined surface was investigated by means of a Lagrangian and an Eulerian numerical approach. The obtained reflected waves and maximum pressure are in good agreement. A linear dependency of the maximum pressure from the surface's angle was deduced. Therefore, the lower is the surface's angle the worse are normal solicitations at the contacting area as the impact takes place.

REFERENCES

[1] T. Belytschko, Fluid Structure Interaction, *Comput. Struct.*, **12**(4), 459-469, 1980.

[2] P. Guruswamy, A review of numerical fluids/structures interface methods for computations using high-fidelity equations, *Comput. Struct*., **80**(1), 31-41, 2002.

[3] D.H. Peregrine, Water wave impact on walls, *Annu. rev. fluid mech.*, **35**, 23–43, 2003.

[4] B. Peseux, L. Gornet, B. Donguy, Hydrodynamic impact: Numerical and experimental investigations, *J. Fluid. Struct*., **21**, 277–303, 2005.

[5] O.M. Faltinsen, M. Landrini, M. Greco, Slamming in marine applications, *J. Eng. Math.*, **48**(3–4), 187–217, 2004.

[6] M.S. Kirkgöz, Breaking wave impact on vertical and sloping coastal structures, *Ocean Eng*., **22**, 35–48, 1995.

[7] M.J. Cooker, D.H. Peregrine, Computations of violent motion due to waves breaking against a wall, In: *Proc. 22nd Intl. Conf. Coastal Eng*., Delft, Holland 164-176, 1990.

[8] X. Chen, B. Hofland, C. Altomare, T. Suzuki, W. Uijttewaal, Forces on a vertical wall on a dike crest due to overtopping flow, *Coast. Eng*., **95**, 94-104, 2015.

[9] G. Cuomo, W. Allsop, T. Bruce, J. Pearson, Breaking wave loads at vertical seawalls and breakwaters, Coast. Eng, **57**(4), 424-439 2010.

[10] D.J. Wood, D.H. Peregrine, T. Bruce, Study of wave impact against a wall with pressure-impulse theory. Part 1: trapped air. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **126**, 182–190, 2000.

[11] A.A. Korobkin, Global characteristics of jet impact, *J. Fluid Mech.*, **307**, 63-84, 1996.

[12] A.A. Korobkin, Asymptotic theory of liquid–solid impact, *Philosophical Transactions: Mathematical, Physical, and Engineering Sciences*, **35**, 507–522, 1997.

[13] T.I. Khabakhpasheva, A.A. Korobkin, Elastic wedge impact onto a liquid surface: Wagner's solution and approximate models, *Journal of Fluids and Structures*, **36**, 32–49, 2013.

[14] T. Stevenson, The design and construction of harbours, 2nd edition, Edinburgh, UK, 1874.

[15] B. Hofland, M. Kaminski, G. Wolters, Large scale wave impacts on a vertical wall, *Coastal Engineering Proceedings*, Vol. 1, 1-15, 2011.

[16] D.H. Peregrine, H. Bredmose, G. Bullock, C. Obhrai, Müller G., Wolters G. Violent water wave impact on a wall, *Proceedings* 14th Aha Huliko Winter Workshop, Honolulu, Hawaii, 2005.

[17] G. Bullock, C. Obhrai, G. Müller, G. Wolters, D.H. Peregrine, H. Bredmose, Field and laboratory measurements of wave impacts, *Proceedings of the 3rd Coastal Structures Conference*, Portland, Oregon. ASCE, 602–617, 2004.

[18] E. Pugliese Carratelli, G. Viccione, V. Bovolin, Free surface flow impact on a vertical wall: a numerical assessment, *Theoretical and Computational Fluid Dynamics*, accepted for publication, 2016

[19] G. Viccione, V. Bovolin, E. Pugliese Carratelli, Simulating fluid-structure interaction with SPH. In: *International Conference of Numerical Analysis and Applied Mathematics*, Kos, Greece, Vol.**1479**, 209-212, 2012.

[20] P.A. Blackmore, P.J. Hewson, Experiments on full-scale wave impact pressures, *Coastal Engineering*, **8**(4), 331–346, 1984.

[21] G. Viccione, V. Bovolin, E. Pugliese Carratelli, Simulating Flows with SPH: Recent Developments and Applications, Book ch. in: *Intech Hydrodynamics - Optimizing Methods and Tools*, 69-84, 2011.

[22] F. Dentale, G. Donnarumma, E. Pugliese Carratelli, Numerical wave interaction with tetrapods breakwater, *International journal of naval architecture and ocean engineering*, **6**, 1-13, 2014.

[23] M. Souli, A. Ouahsine, L. Lewin, ALE formulation for fluid–structure interaction problems, *Computer Methods in Applied Mechanics and Engineering*, **190**(5–7), 659–675, 2000.

[24] N.T. Wu, H. Oumeraci, H.W. Partenscky, Numerical Modelling of Breaking Wave Impacts on a Vertical Wall, *Coastal Engineering Proceedings*, Vol. **24**, 1672–1686, 1994.

[25] B.D. Nichols, C.W. Hirt, R.S. Hotchkiss, SOLA-VOF - A Solution Algorithm for Transient Fluid Flow with Multiple Free Boundaries, Report LA-8355, Los Alamos Scientific Laboratory, University of California, 1980.

[26] J.J. Monaghan, Smoothed Particle Hydrodynamics, *Reports on Progress in Physics*, **68**(8), 1703–1759, 2005.

[27] A.J.C. Crespo, M. Gómez-Gesteira, R.A. Darlymple, 3D SPH Simulation of large waves mitigation with a dike, *J. Hydraul. Res.*, **45**(5), 631–642, 2007.

[28] S. De Chowdhury, S.A. Sannasiraj, Numerical Simulation of 2D Wave Impact Pressure Using SPH Schemes. *The Twenty-third International Offshore and Polar Engineering Conference*, Anchorage, Alaska, 2013.

[29] M. Gómez-Gesteira, B.D. Rogers, R.A. Dalrymple, A.J.C. Crespo State-of-the-art of classical SPH for free-surface flows, *J. Hydraul. Res.*, **48** Extra Issue, 6–27, 2010.

[30] Violeau D., Fluid Mechanics and the SPH Method: Theory and Applications, Oxford University Press, 2012.

[31] R.A. Gingold, J.J. Monaghan, Smoothed Particle Hydrodynamics: Theory and Application to Nonspherical Stars, *Mon. Not. R. Astr. Soc*., **181**, 375–389, 1977.

[32] L. Lucy, A Numerical Approach to the Testing of Fusion Process, *J. Astronom.*, **82**, 1013–1024, 1977.

[33] J.J. Monaghan, Simulating Free Surface Flows with SPH, *J. Comput. Phys*., **110**, 399–406, 1994.

[34] R.A. Dalrymple, O. Knio, D.T. Cox, M. Gómez-Gesteira, S. Zou. Using a Lagrangian Particle Method for Deck Overtopping, *Proc. Waves 2001*, ASCE, 1082–1091, 2002.

[35] M. Gómez-Gesteira, R.A. Dalrymple, Using a 3D SPH Method for Wave Impact on a Tall Structure, *J. Wtrwy. Port, Coast. Ocean Engg*., **130**(2), 63–69, 2004.

[36] C. Altomare, A.J.C. Crespo, J.M. Domínguez, M. Gómez-Gesteira, T. Suzuki, T. Verwaest, Applicability of Smoothed Particle Hydrodynamics for estimation of sea wave impact on coastal structures, *Coast. Eng*., **95**, 94-104, 2015.

[37] F. Aristodemo, D.D. Meringolo, P.H.L. Groenenboom, A. Lo Schiavo, P. Veltri, M. Veltri, Assessment of dynamic pressures at vertical and perforated breakwaters through diffusive SPH schemes, *Math. Probl. Eng.*, 10 pages, 2015.

[38] G. Viccione, V. Bovolin, E.P ugliese Carratelli, Defining and optimizing algorithms for neighbouring particle identification in SPH fluid simulations, *International Journal for Numerical Methods in Fluids*, **58**, 625-638, 2008.

[39] G. Viccione, V. Bovolin, Simulating Triggering and Evolution of Debris-Flows with Smoothed Particle Hydrodynamics (SPH), *Italian Journal of Engineering Geology and Environment*, 523-532, 2011.

[40] Flow Sciences Inc. Flow-3D User Manual, release 9.4, Santa Fe, NM, USA, 2009.

[41] L. Cavallaro, F. Dentale, G. Donnarumma, E. Foti, R.E. Musumeci, E. Pugliese Carratelli, Rubble Mound reakwater Overtopping: Estimation of the reliability of a 3D Numerical Simulation, In: *ICCE 2012, International Conference on Coastal Engineering*, Santander, Spain, 2012.

[42] F. Dentale, G. Donnarumma, E. Pugliese Carratelli, Simulation of flow within armour blocks in a breakwater, *J. Coastal Res*., **30**(3), 528-536, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

AERODYNAMICS OF AN UNSTEADY PARTICLE FLOW IN THE CHUTE

O. A. Averkova[1], I. N. Logachev[1], and K. I. Logachev[1],

[1] Belgorod State Technological University named after V.G. Shukhov,
308012 Belgorod, Russia, Kostukova str., 46
e-mail: {logachevin, olga _19572004, kilogachev}@mail.ru

Keywords: Granular Materials, Bulk Material Transfer, Air Suction, Local Exhaust Ventilation.

Abstract. *The conducted studies are aimed at improving the theoretical basis for design of ventilated suction hoods of bulk material transfer groups, i.e. the most typical and common sources of dust emissions in reprocessing of dust-forming materials.Unsteady processes occur in the chute at the equipment start-up or short-time loading of bulk material. Assess the force action exerted by the flow on air for two cases: at an instantaneous change and a gradual one (smooth) in the material flow. `It was theoretically demonstrated that at instant chuting of a heated bulk material an increase in the induction pressure is strongly related with a variation in the transferred material flow, and due to a considerable inertness of the intercomponent heat exchange temperature variations are far behind of variations in the transferred material quantity. A gradual variation in the flow rate of particles at the beginning and at the end of a material transfer is associated with excessive (peak) pressures against steady-state values. This is explained by the maximum induction head at a certain material flow rate. There are no pressure surges if the material flow rate does not reach the necessary value.*

.

1 INTRODUCTION

The conducted studies are aimed at improving the theoretical basis for design of ventilated suction hoods of bulk material transfer groups, i.e. the most typical and common sources of dust emissions in reprocessing of dust-forming materials [1-8, 12-14].

Unsteady processes occur in the chute at the equipment start-up or short-time loading of bulk material. Assess the force action exerted by the flow on air for two cases: at an instantaneous change and a gradual one (smooth) in the material flow.

2 SUDDEN CHANGE IN THE MATERIAL FLOW

Examine the change in the forces of the induction and thermal pressures and the starting and stop times of the material feeding.

Observe the change in the induction pressure by the example of the pressure distribution along the length of the vertical pipe in case of its irregular load with bulk material of moderate temperature, thereby preventing from the heat and mass exchange. Imagine that the lower end of the pipe is closed to the air passage, i.e. $v_2 = 0$, but open to the passage of material.

With these simplifications [1]

$$\frac{\partial v_2}{\partial t} = -\frac{\partial P}{\partial x} \cdot \frac{1}{\rho_2} + \kappa_m \beta_1 \psi \frac{v_1^2}{2}. \tag{1}$$

Examine a “hard” loading of material (instant inlet of material). Then, at an arbitrary point of time $t > 0$, the change in the material flow along the length will be of a step-type shape.

$$G_1 = G_{\max} \quad \text{at} \quad x < x_0 \ ,$$
$$G_1 = 0 \quad \text{at} \quad x > x_0 \, .$$

The kink will be moved down. Let us assume that its movement speed is

$$\overline{v}_1 = \frac{1}{l}\int_0^l v_1 dx \approx 0,5(v_{1н} + v_{1k}) \, .$$

The analytical expression of the change in the material flow is as follows

$$G_1 = G_{\max} \cdot f(x - v_1 t) \, , \tag{2}$$

where

$$f(x - v_1 t) = 0 \quad \text{at} \quad x - v_1 t > 0 \, ; \tag{3}$$

$$f(x - v_1 t) = 1 \quad \text{at} \quad x - v_1 t < 0 \, . \tag{4}$$

For simplicity of notation, the symbol of averaging (a line above v_1) is omitted here and hereinafter.

Express the function $f(x - v_1 t)$ by means of the infinite Fourier series [9]

$$f(x - v_1 t) = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{\cos \pi n - 1}{n\pi} \sin \frac{\pi}{l_\infty} (x - v_1 t) \, , \tag{5}$$

where

$$l_\infty = \lim_{t \to \infty} (x - v_1 t) \, . \tag{6}$$

Then, the volume concentration of material is

$$\beta_1 = \beta_{1\infty}\left[\frac{1}{2}+\sum_{n=1}^{\infty}\frac{\cos n\pi - 1}{n\pi}\sin n\frac{\pi}{l_\infty}(x - v_1 t)\right], \tag{7}$$

where

$$\beta_{1\infty} = \frac{G_{1\max}}{\rho_1 S_{ж} v_1} \tag{8}$$

or, solving the sine function of difference of two angles,

$$\beta_1 = \beta_{1\infty}\left[\frac{1}{2}+\sum_{n=1}^{\infty}\frac{\cos n\pi - 1}{n\pi}\sin n\frac{\pi}{l_\infty}x\cos n\frac{\pi}{l_\infty}tv_1 - \sum_{n=1}^{\infty}\frac{\cos n\pi - 1}{n\pi}\cos n\frac{\pi}{l_\infty}x\sin n\frac{\pi}{l_\infty}v_1 t)\right] \tag{9}$$

Write mass transfer equation [1]

$$\frac{\partial \rho_2}{\partial t}+\frac{\partial \rho_2 v_2}{\partial x}=0. \tag{10}$$

Assuming that that process is adiabatic (with factor χ)

$$\frac{P}{P_0}=\left(\frac{\rho_2}{\rho_0}\right)^{\chi}, \tag{11}$$

simple transformations made, taking into account that the density of the medium does not change significantly, equality (10) can be reduced to

$$\frac{\partial v_2}{\partial x}=-\frac{\rho_0}{\bar{\rho}_2 P_0 \chi}\frac{\partial P}{\partial t}, \tag{12}$$

where P_0, ρ_0 are the pressure and density of air in the pipe before inputting the material.

Taking account of correlation (12), after differentiation of equality (1) with respect to x, we obtain an inhomogeneous equation of acoustics:

$$\frac{\partial^2 P}{\partial t^2}=v_a^2\frac{\partial^2 P}{\partial x^2}-v_a^2\frac{\partial}{\partial x}\kappa_m\beta_1\psi\frac{v_1^2}{2}\bar{\rho}_2, \tag{13}$$

where $v_a=\sqrt{\chi P_0/\rho_0}$ is the propagation speed of elastic disturbances (speed of sound), m/sec.

Assume that the force of the dynamic interaction is constant

$$\kappa_m\psi\frac{v_1^2}{2}\bar{\rho}_2 = const. \tag{14}$$

Under this assumption, a linear pressure distribution along the pipe would exist for stationary conditions

$$P_{cm}=P_0+\kappa_m\beta_{1\infty}\psi\frac{v_1^2}{2}\bar{\rho}_2 x \tag{15}$$

or

$$P_\infty = P_l\frac{x}{l}, \tag{16}$$

where P_l is the excessive pressure at the end of the pipe

$$P_l = \kappa_m \beta_{1\infty} \psi \frac{v_1^2}{2} \bar{\rho}_2 l, \tag{17}$$

P_{cm}, P_∞ are the absolute and excessive pressures, respectively, in a stable process, Pa.

Taking into account the assumptions made and designations, the last term of the right-hand side of equation (13) can be written as

$$v_a^2 \frac{\partial}{\partial x} \kappa_m \beta_1 \psi \frac{v_1^2}{2} \bar{\rho}_2 = v_a^2 \frac{P_l}{l l_\infty} \left[\sum_{n=1}^{\infty} (\cos n\pi - 1) \cdot \cos n\pi \frac{x}{l_\infty} \cos n\pi \frac{v_1 t}{l_\infty} + \right.$$
$$\left. + \sum_{n=1}^{\infty} (\cos n\pi - 1) \sin n\pi \frac{x}{l_\infty} x \cdot \sin n\pi \frac{v_1 t}{l_\infty} \right]. \tag{18}$$

Formulate the initial and boundary conditions. Pressure P in equation (13) is to mean the excessive pressure. As dead air was in the pipe before the input of material, and the absolute pressure in it was P_0:

$$P\big|_{t=0} = 0; \tag{19}$$

$$v_2\big|_{t=0} = 0. \tag{20}$$

Integrate equation (12) to obtain

$$v_2 = v_2\big|_{x=0} - \frac{\rho_0}{P_0 \chi} \int_0^x \frac{\partial P}{\partial t} dx, \tag{21}$$

where from, taking into account (20), obtain the second initial condition

$$\left. \frac{\partial P}{\partial t} \right|_{t=0} = 0. \tag{22}$$

The boundary condition for the open end of the pipe (input) will be

$$P(0,t) = 0. \tag{23}$$

For the lower end, taking in account the air tightness of the pipe bottom,

$$v_2(l,t) = 0. \tag{24}$$

Integrate the initial equation (1) to obtain

$$v_2 = v_2(x,0) + \int_0^t \left(-\frac{\partial P}{\partial x} \cdot \frac{1}{\rho_2} + \kappa_m \beta_1 \psi \frac{v_1^2}{2} \right) dt. \tag{25}$$

Taking into account (24), obtain the second boundary condition:

$$\left. \frac{\partial P}{\partial x} \right|_{x=l} = \left. \kappa_m \beta_1 \psi \bar{\rho}_2 \frac{v_1^2}{2} \right|_{x=l}, \tag{26}$$

which can be written as follows, having in mind correlation (7), after some transformations

$$P\big|_{x=l} = 2 \frac{P_l}{l} \sum_{n=1}^{\infty} \frac{1 - \cos n\pi}{n\pi} \cos^2 n\pi \frac{l}{2 l_\infty} \sin n\pi \frac{v_1 t}{l_\infty} + \frac{P_l}{l} \sum_{n=1}^{\infty} \frac{\cos n\pi - 1}{n\pi} \sin n\pi \frac{l}{l_\infty} \cos n\pi \frac{v_1 t}{l_\infty}. \tag{27}$$

Thus, we have to solve the following inhomogeneous differential partial equation with second-order partial derivatives [10] with the initial conditions (19) and (22) and the boundary conditions (23) and (27)

$$\frac{\partial^2 P}{\partial t^2} = v_a^2 \frac{\partial P}{\partial x^2} - v_a^2 \frac{P_l}{l l_\infty} \sum_{n=1}^{\infty} (\cos n\pi - 1) \cos n\pi \frac{x}{l_\infty} \cos n\pi \frac{v_1 t}{l_\infty} - v_a^2 \frac{P_l}{l l_\infty} \sum_{n=1}^{\infty} (\cos n\pi - 1) \sin n\pi \frac{x}{l_\infty} \sin n\pi \frac{v_1 t}{l_\infty} \quad (28)$$

The solution will be calculated as a sum of functions

$$P = u + \xi, \quad (29)$$

where u is the solution of equation (28) only with the boundary conditions; ξ is the solution of this equation without a constant term with the following initial and boundary conditions:

$$\xi\big|_{t=0} = -u\big|_{t=0}; \quad (30)$$

$$\left.\frac{\partial \xi}{\partial t}\right|_{t=0} = -\left.\frac{\partial u}{\partial t}\right|_{t=0}; \quad (31)$$

$$\xi\big|_{x=0} = 0; \quad (32)$$

$$\left.\frac{\partial \xi}{\partial x}\right|_{x=l} = 0. \quad (33)$$

The task to determine u can, in turn, be divided into two sub-tasks:

a) the solution of the equation

$$\frac{\partial^2 u}{\partial t^2} = v_a^2 \frac{\partial^2 u}{\partial x^2} - v_a^2 \frac{P_l}{l \cdot l_\infty} \sum_{n=1}^{\infty} (\cos n\pi - 1) \cos n\pi \frac{x}{l_\infty} \cos n\pi \frac{v_1 t}{l_\infty} \quad (34)$$

with

$$u\big|_{x=0} = 0; \quad \left.\frac{\partial u}{\partial x}\right|_{x=l} = \frac{P_l}{l} \sum_{n=1}^{\infty} \frac{\cos n\pi - 1}{n\pi} \sin n\pi \frac{l}{l_\infty} \cos n\pi \frac{v_1 t}{l_\infty}; \quad (35)$$

b) the solution of the equation

$$\frac{\partial^2 u}{\partial t^2} = v_a^2 \frac{\partial^2 u}{\partial x^2} - v_a^2 \frac{P_l}{l \cdot l_\infty} \sum_{n=1}^{\infty} (\cos n\pi - 1) \sin n\pi \frac{x}{l_\infty} \sin n\pi \frac{v_1 t}{l_\infty} \quad (36)$$

with

$$u\big|_{x=0} = 0 \quad ; \quad \left.\frac{\partial u}{\partial x}\right|_{x=l} = \frac{P_l}{l} \sum_{n=1}^{\infty} \frac{\cos n\pi - 1}{n\pi} \left(1 + \cos n\pi \frac{l}{l_\infty}\right) \sin n\pi \frac{v_1 t}{l_\infty}. \quad (37)$$

Subtask a). Try the solution of equation (36) as follows

$$u_a = \sum_{n=1}^{\infty} X_n(x) \cos n\pi \frac{v_1 t}{l_\infty}. \quad (38)$$

Inserting this solution into the initial equation after obvious reductions, we obtain

$$\frac{d^2 X_n(x)}{dt^2} + X_n(x) \cdot \left(\frac{n\pi M_v}{l_\infty}\right)^2 = \frac{P_l}{l \cdot l_\infty} (\cos n\pi - 1) \cdot \cos n\pi \frac{x}{l_\infty}, \quad (39)$$

where

$$M_v = v_1 / v_a \,. \tag{40}$$

Integrate this equation, with

$$X_n(0) = 0\,,$$

$$\left.\frac{dX_n(x)}{dx}\right|_{x=l} = \frac{P_l}{l}\frac{\cos n\pi - 1}{n\pi}\cdot \sin n\pi \frac{l}{l_\infty} \tag{41}$$

based on the conditions (35) we obtain

$$X_n(x) = \Lambda\left[\Phi \sin \pi n \frac{x}{l_\infty} M_v - \cos n\pi \frac{x}{l_\infty} M_v + \cos n\pi \frac{x}{l_\infty}\right] \tag{42}$$

and the solution of Subtask a) is

$$u_a = \sum_{n=1}^{\infty} \Lambda\left[\Phi \sin \pi n \frac{x}{l_\infty} M_v - \cos n\pi \frac{x}{l_\infty} M_v + \cos n\pi \frac{x}{l_\infty}\right] \cdot \cos \pi n \frac{v_1 t}{l_\infty}\,, \tag{43}$$

where for simplicity of notation we assume

$$\Lambda = \frac{P_l \cdot l_\infty}{(n\pi)^2 l} \cdot \frac{\cos n\pi - 1}{M_v^2 - 1}\,, \tag{44}$$

$$\Phi = \left(M_v \sin \pi n \frac{l}{l_\infty} - \sin \pi n \frac{l}{l_\infty} M_v\right) \Big/ \cos n\pi \frac{l}{l_\infty} M_v \,. \tag{45}$$

Subtask b). Equation (36) is solved in the same way.
Obtain

$$u_\delta = \sum_{n=1}^{\infty} \Lambda\left(\sin \pi n \frac{x}{l_\infty} - \Omega \sin \pi n \frac{x}{l_\infty} M_v\right) \cdot \sin \pi n \frac{v_1 t}{l_\infty}\,, \tag{46}$$

where

$$\Omega = \frac{M_v^2 - 1}{M_v} \cdot \frac{1 + \cos n\pi \dfrac{l}{l_\infty}}{\cos n\pi \dfrac{l}{l_\infty} M_v} + \frac{1}{M_v}\,. \tag{47}$$

We now calculate the function ξ, which is the solution of equation

$$\frac{\partial^2 \xi}{\partial t^2} = v_a^2 \frac{\partial^2 \xi}{\partial x^2}\,, \tag{48}$$

with the boundary conditions (32) and (33) and with the following initial conditions based on equalities (43) and (46)

$$\xi\big|_{t=0} = -(u_a + u_\delta)\big|_{t=o} = -\sum_{n=1}^{\infty} \Lambda\left[\Phi_n \sin \pi n \frac{x}{l_\infty} M_v - \cos n\pi \frac{x}{l_\infty} M_v + \cos n\pi \frac{x}{l_\infty}\right] = f(x)\,; \tag{49}$$

$$\left.\frac{\partial \xi}{\partial t}\right|_{t=0} = -\left(\frac{\partial u_a}{\partial t} + \frac{\partial u_\delta}{\partial t}\right)\Big|_{t=0} = -\sum_{n=1}^{\infty} \Lambda n\pi \frac{v_1}{l_\infty}\left(\sin \pi n \frac{x}{l_\infty} - \Omega \sin \pi n \frac{x}{l_\infty} M_v\right) = F(x)\,. \tag{50}$$

We solve it using the Fourier method – by expansion of the function in series in the orthogonal function system $\sin\frac{2n+1}{2l}\pi x$ [11].

The solution is

$$\xi = \sum_{n=0}^{\infty}\left[a_к \cos\frac{2n+1}{2l}\pi v_a t + \delta_к \sin\frac{2n+1}{2l}\pi v_a t\right]\sin\frac{2n+1}{2l}\pi x, \tag{51}$$

where

$$a_к = \frac{2}{l}\int_0^l f(x)\sin\frac{2n+1}{2l}\pi x dx; \tag{52}$$

$$\delta_к = \frac{4}{(2n+1)\pi v_a}\int_0^l F(x)\sin\frac{2n+1}{2l}\pi x dx. \tag{53}$$

Inserting into equality (29) the obtained functions ξ, u_a, u_δ, we obtain the desired solution. For a small falling velocity of the material ($v_1/v_2 << 1$), the solution can be reduced to a simpler form after a number of simplifications

$$P = P_l\frac{l_\infty}{l}\sum_{n=1}^{\infty}\frac{1-\cos n\pi}{(n\pi)^2}\cos n\pi\frac{x-v_1 t}{l_\infty} + \frac{1}{2}P_l\frac{l_\infty}{l}\left(\frac{v_1 t}{l_\infty}-\frac{1}{2}\right)+\frac{1}{2}P_l\frac{x}{l}. \tag{54}$$

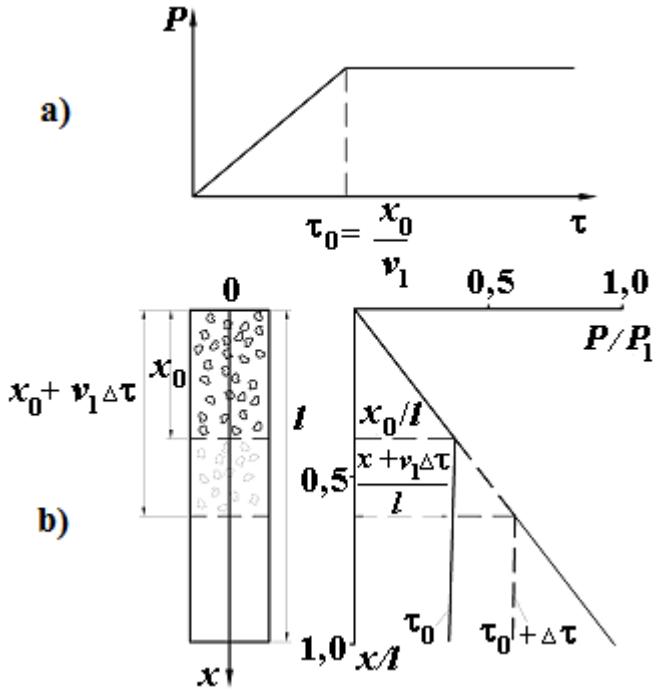


Figure 1: Change in the induction pressure over time (a) and along the tube (b) in case of an instant loading of bulk material

As can be seen from the graph (Figure 1) plotted based on this equation, the pressure in an arbitrary section x_0 increases in this section up to the maximum according to $t_0 \equiv \tau_0 = x_0/v_1$ s, i.d., as soon as the first particles of the material reach the section under consideration.

Along the entire pipe length, the pressure reaches its maximum value as soon as the pipe is filled with the falling material.

Thus, a change in the induction pressure is "rigidly" connected with the change in the material flow. The steady mode of dynamic interaction between the material and the air occurs almost simultaneously with a constant flow of material in all sections of the pipe.

In contrast to the dynamic interaction, temperature changes "fall behind" the fluctuations significantly in the mode of the material transfer.

To see this, consider the same task having slightly simplified it. Assume that the air thermal conductivity is high supposing that the same temperature is instantly set in all the sections of the pipe. Thus, the temperature will depend on time only. With the previous assumptions, the heat exchange equation can be as follows

$$\frac{d}{d\tau}\beta_2\rho_2 c_2 t_2 = \alpha\kappa_s\beta_{1l}(t_1 - t_2) - 4\frac{k}{D}(t_2 - t_0), \tag{55}$$

where k is a coefficient of heat exchange with surrounding air, Вт/(м^2 · °К); β_{1l} is the volume concentration of material in the chute.

Taking into account the step-type change in the flow rate, a solution is obtained for two intervals with $0 < \tau < \frac{l}{v_1}$, with $\tau > \frac{l}{v_1}$:

in the first interval

$$\beta_{1l} = \frac{G_1\tau}{\rho_1 S_{ж} l}\ ; \tag{56}$$

in the second interval

$$\beta_{1l} = \frac{G_1}{\rho_1 S_{ж} v_1}\ . \tag{57}$$

Integrate equality (55) with the initial condition

$$t_2(x,0) = t_0, \tag{58}$$

to obtain:

a) with $0 < \tau < \frac{l}{v_1}$

$$t_2 = t_1 - (t_1 - t_0)\left\{\exp\left[-\left(Б\tau + \frac{Б\tau^2}{2}\right)\right] + \frac{Б}{\sqrt{B}}\exp\left(-\frac{y}{2}\right)\int_{Б/\sqrt{B}}^{y} erf\left(\frac{z^2}{2}\right)dz\right\}; \tag{59}$$

b) with $\tau > l/v_1$

$$t_2 = t_{2\infty} - (t_{2\infty} - t_{2н})\exp\left[-\left(B\frac{l}{v_1} - Б\right)\left(\tau - \frac{l}{v_1}\right)\right], \tag{60}$$

where $t_{2н}$ is the air temperature in the pipe with $\tau = l/v_1$ is defined by equality (59); $t_{2\infty}$ is the temperature in the pipe with $\tau \to \infty$

$$t_{2\infty} = \frac{B\frac{l}{v_1}t_1 + Бt_0}{B\frac{l}{v_1} + Б}; \tag{61}$$

B, *Б* are parameters introduced for simplicity of notation and equal to

$$B = \alpha\kappa_S G_1 / (\rho_2 c_2 \rho_1 S_{ж} l); \tag{62}$$

$$Б = 4\kappa / (D\rho_2 c_2);$$

$$y = \sqrt{B}\tau + Б/\sqrt{B}\,. \tag{63}$$

For a heat sealed pipe (*Б* = 0) we have:

a) with $0 < \tau < l/v_1$

$$t_2 = t_1 - (t_1 - t_0)\exp(-B\tau^2/2); \tag{64}$$

b) with $\tau > l/v_1$

$$t_2 = t_1 - (t_1 - t_0)\exp\left[-B\frac{l}{v_1}\left(\tau - \frac{l}{2v_1}\right)\right]. \tag{65}$$

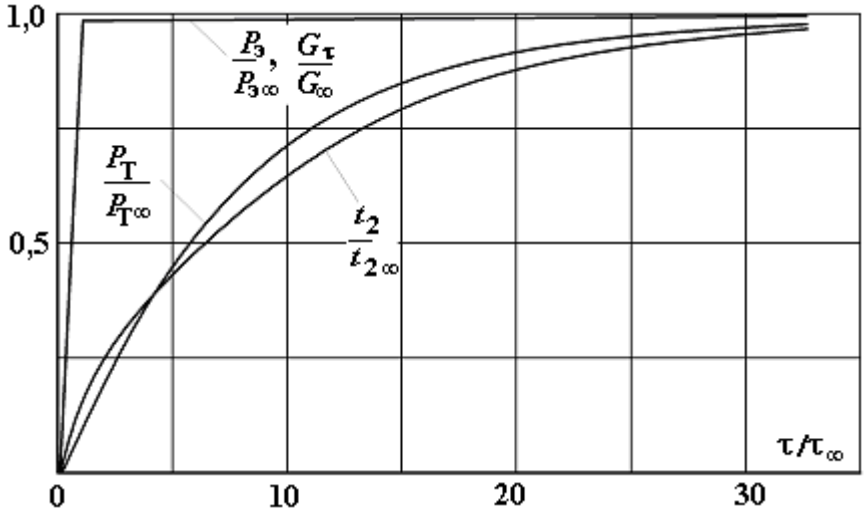


Figure 2: Temporal variation in temperature, thermal and induction pressures, mass of particles in the chute ($P_{Э\infty}$, $P_{T\infty}$ are the induction and thermal pressure with $\tau \to \infty$; $G\tau$, $G\infty$ are the masses of particles in the chute at the moments in time τ and $\tau_\infty = l/\overline{v}_1$)

Figure2 shows the temperature curves plotted according to these formulas. It also shows the change in the thermal and induction pressures. As the curves show, the thermal pressure has a considerable "inertia" if compared with the induction one.

3 SMOOTH CHANGE IN THE MATERIAL FLOW

Given that the induction pressure is "rigidly" connected with the material flow, changes in the dynamic interaction can be assessed in the conditions of a changing flow of particles by means of the correlations obtained in the study of stationary flows. We can be used for a pressure at the end of the pipe, the lower end of which is closed to the air passage [1]:

$$P_Э = \psi_0 \kappa_m \beta e^{\frac{-1{,}8\sqrt{\beta\cdot 10^3}}{d_Э\cdot 10^3}} \cdot l\frac{v_{1k}^2}{2}\rho_2\frac{1-n^3}{3(1-n)}, \tag{66}$$

where β is an averaged volume concentration of particles in the chute, which varies over time due to changes in flow.

As seen from equation (65), the induction pressure has a maximum at the volume concentration (β_{max}) defined by the equality

$$2-\frac{1,8}{d_э\cdot 10^3}\sqrt{\beta_{max}\cdot 10^3}=0. \qquad (67)$$

Thus, if it appears that the concentration of particles varies widely from 0 to $\beta_\infty > \beta_{max}$, pressure surges occur during an unsteady-state process. This is clearly seen in the curves in Figure 3. Here is a case of a bulk material transfer where its flow rate varies from 0 to the steady-state (constant) value $G_{1\infty}$, then a stationary process ($G_1 = G_{1\infty}$) continues for a while and, finally, the flow rate decreases from $G_{1\infty}$ down to 0.

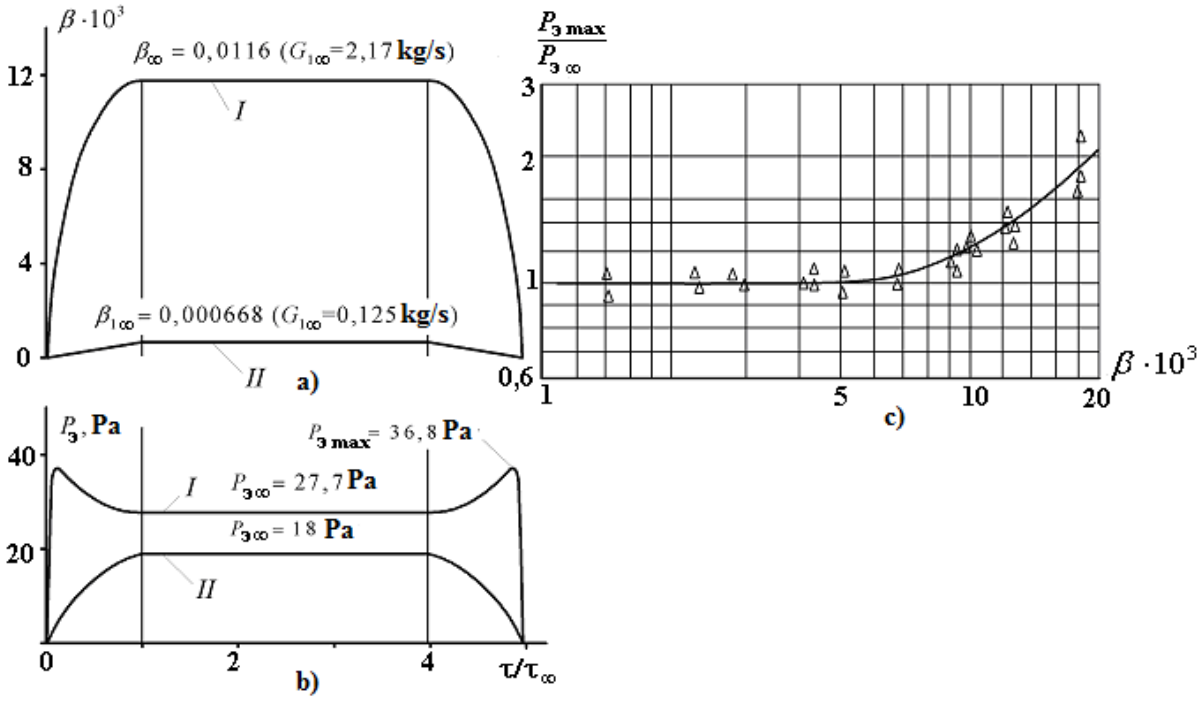


Figure 3: Change in the volume concentration and the induction pressure in the chute with a slow change in the material flow (Δ is experimental data for conditions of granite bulking d_e = 1.88 mm in the chute at Θ = 75°, H = 3.3 m, S_{ch} = 0.0169 m^2)

However, the pressure surge may be absent where the steady-state flow is so small that the volume concentration of particles in the chute is $\beta_\infty < \beta_{max}$.

The maximum value of the induction pressure according to (67) is

$$P_{э_{max}}=\psi_0\kappa_m 10^{-3}\left(\frac{2\cdot d_э\cdot 10^3}{1,8}\right)^2 e^{-2}\cdot l\frac{v_{1k}^2}{2}\rho_2\frac{1-n^3}{3(1-n)} \qquad (68)$$

or the relative value of the pressure surge is

$$P_{э\max}/P_{э\infty}=10^{-3}\left(\frac{2\cdot 10^3 d_э}{1,8}\right)^2 e^{-2}\Bigg/\left[\beta_\infty\exp\left(-1,8\frac{\sqrt{\beta_\infty\cdot 10^3}}{d_э\cdot 10^3}\right)\right] \text{ with } \beta_\infty\geq\beta_{max}. \qquad (69)$$

In studies of the induction properties of a bulk material flow in inclined chutes at the unit (Figure 4), a pressure surge was often observed when the material feeding from the upper bin both began and stopped. The value of this surge was significant at large material flows. No pressure rise was observed with small flows. A pressure rise is absolutely in line with equations (69) not only in qualitative terms but also in quantitative terms (see Figure 3.c). The observed falling of the thermal pressure behind the induction one, as well as the surge of the induction pressure during the start or stop of the process equipment must be taken into account when calculating the required volumes of aspirated air.

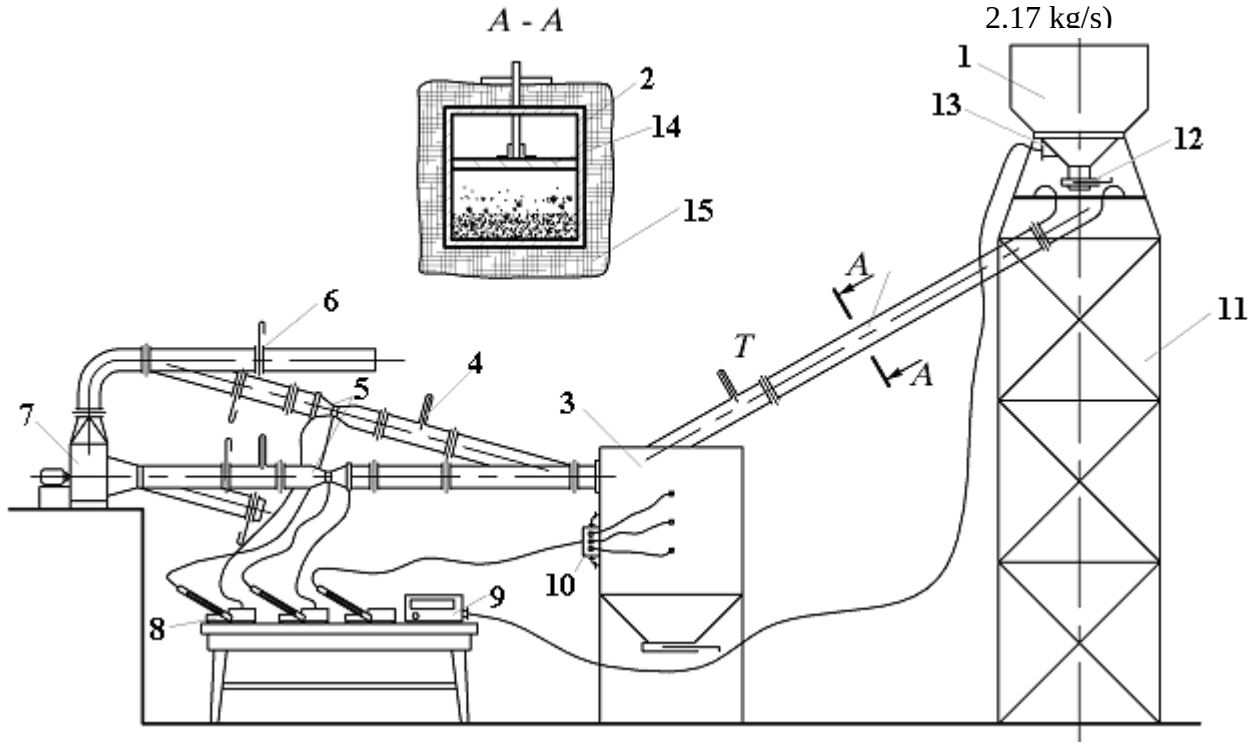


Figure 4: Diagram of the experimental arrangement for the study of injective properties of bulk materials: 1 – upper bin; 2 – chute; 3 – lower bin; 4 – thermometer; 5 – Venturi tube; 6 – damper; 7 – fan; 8 – micropressure gauge; 9 – galvanometer; 10 – blending chamber; 11 – metal frame; 12 – diaphragm;13 – thermocouple; 14 – chute upper wall; 15 – heat insulation layer

4 CONCLUSIONS

- The conducted studies are aimed at improving the theoretical basis for design of ventilated suction hoods of bulk material transfer groups, i.e. the most typical and common sources of dust emissions in reprocessing of dust-forming materials. The basic findings and resulting conclusions are as follows:
- It was theoretically demonstrated that at instant chuting of a heated bulk material an increase in the induction pressure is strongly related with a variation in the transferred material flow (54), and due to a considerable inertness of the intercomponent heat exchange temperature variations are far behind of variations in the transferred material quantity (Figure 2). A gradual variation in the flow rate of particles at the beginning and at the end of a material transfer is associated with excessive (peak) pressures against steady-state values (Figure 3). This is explained by the maximum induction head at a certain

material flow rate (67, 68). There are no pressure surges if the material flow rate does not reach the necessary value.

5 ACKNOWLEDGEMENTS

The reported study was funded by RFBR according to the research project № 14-41-08005 ofi_m и and Russian President Scholarship СП-121.2016.1.

REFERENCES

[1] I.N. Logachev, K.I. Logachev, O.A. Averkova, *Local Exhaust Ventilation: Aerodynamic Processes and Calculations of Dust Emissions*. Boca Raton: CRC Press, 2015.

[2] I.N. Logachev, K.I. Logachev, *Industrial Air Quality And Ventilation: Controlling Dust Emissions*. Boca Raton, 2014.

[3] Ze Qin Liu. *Air Entrainment in Free Falling Bulk Materials. Ph. D. Thesis*. University of Wollngong, 2003. http:/ro.uow.edu.au/theses/1816.

[4] S. Mallick, P. W. Wypych. Evaluation of scaleup procedures using "system" approach for pneumatic conveying of powders. *Particulate Sci. Technol*. **28**, No 1, 41–50, 2010.

[5] D. Hastie, P. Wypych. Researching conveyor belt trajectories. *Engineers Australia*, **81**, 8, 2009.

[6] Li Xiaochuan, Wang Qili, Liu Qi, Hu Yafei. Developments in studies of air entrained by falling bulk materials. *Powder Technology*, **291**, 159-169, 2016.

[7] Chenchen Fang, Jianliang Xu, Hui Zhao, Weifeng Li, Haifeng Liu. Experimental investigation on particle entrainment behaviors near a nozzle in gas–particle coaxial jets. *Powder Technology*, **286**, 55-63, 2015.

[8] Koichiro Ogata, Katsuya Funatsu, Yuji Tomita. Experimental investigation of a free falling powder jet and the air entrainment// *Powder Technology*, **115**, 90-95, 2001

[9] G.M. Fichtengolz. *Calculus course. V.III.* M: Fizmatgiz, 1963.

[10] O.D. Neykov, I.N. Logachev, R.N. Shumilov. *Aspiration of vapor and dust mixtures in dedust-ing of process equipment*. Kiev: Naukova dumka, 1974.

[11] G. Aramanovich, V.I. Levin. *Equations of mathematical physics*. M: Nauka, 1964.

[12] V.G. Shaptala, V.V. Shaptala, A.V. Gavrilenko, Numerical simulation of cement-air mixture in the pneumochamber pump. *Bulletin of BSTU named after V.G. Shukhov*, **2**, 159-161, 2015.

[13] V.V. Shaptala, Modeling and calculation systems of cleaning dust emissions of cement production. *Bulletin of BSTU named after V.G. Shukhov*, **6**, 205-209, 2015.

[14] V.G. Shaptala, V.V. Shaptala, The method of calculating efficiency of centrifugal collectors sticky dust in the production of building materials. *Bulletin of BSTU named after V.G. Shukhov*, **3**, 58-62, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PROJECTION-BASED MODEL REDUCTION FOR FINITE ELEMENT APPROXIMATION OF SHALLOW WATER FLOWS

Matthew Farthing[1], **Alexander Lozovskiy**[2], **and Chris Kees**[1]

[1]Coastal and Hydraulics Laboratory, U.S. Army Engineer Research and Development Center,
3909 Halls Ferry Road, Vicksburg, MS 39180, USA
e-mail: matthew.w.farthing@usace.army.mil

[2] Institute of Scientific Computation, Texas A&M University,
155 Ireland Street, College Station, TX 77840, USA
e-mail: lozovskiy@math.tamu.edu

Keywords: Shallow Water equations, model reduction, POD, GNAT, global basis.

Abstract. *The shallow water equations (SWE) are used to model a wide range of environmental flows from dam breaks and riverine hydrodynamics to hurricane storm surge and atmospheric processes. Despite significant gains in numerical model efficiency stemming from algorithmic and hardware improvements, accurate shallow water modeling can still be very computationally intensive. The resulting computational expense remains as a barrier to the inclusion of fully resolved two-dimensional shallow water models in many applications, particularly when the analysis involves optimal design, parameter inversion, risk assessment, and/or uncertainty quantification.*

Here, we consider projection-based model reduction as a way to alleviate the computational burden associated with high-fidelity shallow-water approximations in ensemble forecast and sampling methodologies. In order to develop a robust approach that can resolve advection-dominated problems with shocks as well as more smoothly varying riverine and estuarine flows, we consider techniques using both Galerkin and Petrov-Galerkin projection on global bases provided by Proper Orthogonal Decomposition (POD). To achieve realistic speedup, we consider alternative methods for the reduction of the non-polynomial nonlinearities that arise in typical finite element formulations. We evaluate the schemes' performance by considering their accuracy and robustness for test problems in one and two space dimensions.

1 INTRODUCTION

The two-dimensional, depth-averaged shallow water equations (SWE) are used throughout science and engineering to model free-surface flows in systems where vertical scales are small with respect to horizontal length scales [22]. Their widespread utility has led to the development of a host of numerical methods and solution techniques for the SWE. A non-exhaustive list includes high-resolution finite volume [18] and discontinuous Galerkin methods [1] as well as characteristic-based [21] and stabilized finite element methods (FEMs) [7].

While ongoing work continues to improve the performance of core numerical methods for the SWE, it is clear that alternative approaches are needed to achieve the speedups necessary to make use of the SWE common in real-time simulation or sampling-intensive methods for risk assessment or uncertainty quantification [9]. Here, we consider global model reduction for the SWE via projection. Specifically, we consider Galerkin-based projection via Proper Orthogonal Decomposition (POD) [3] as well as Petrov-Galerkin approximations based on the Gauss-Newton with Approximate Tensors (GNAT) technique introduced by [11, 12]. In both cases, we use a series of high-fidelity training simulations (or snapshots) in an offline stage to build a basis of empirical modes using a Singular Value Decomposition (SVD) for the solution in order to capture flow dynamics using (hopefully) many fewer degrees of freedom than the original, high-fidelity simulation. Calculations using this low-dimensional or reduced basis can then, in principle, be performed much faster during a time or resource sensitive online stage [3].

Due to the non-polynomial nonlinearities that arise in standard roughness parameterizations and stabilized FEM approximations, further reduction of the nonlinearities in the SWE are required to break dependence on the fine-scale dimension [2]. We evaluate two types of so-called hyper-reduction for our reduced models: Discrete Empirical Interpolation (DEIM) [5, 13] and gappy POD [14, 23].

In the following, we lay out these reduction techniques for a semi-implicit, stabilized FEM approximation of the SWE. We pay particular attention to conditions when the schemes are *consistent*. That is, we consider conditions under which the models will restore the original fine-scale solution if the entire empirical basis obtained from the snapshot collection is used [11]. We end with numerical experiments comparing performance of the various methods for a one-dimensional Burgers equation problem as well as a two-dimensional dam-break.

2 HIGH-FIDELITY FORMULATION FOR THE SHALLOW WATER EQUATIONS

2.1 Continuous formulation

We begin with the fine scale formulation. The standard two-dimensional shallow water equations consist of a mass conservation equation and two scalar momentum conservation equations that can be written in conservative form as

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{F} - \mathbf{D}\nabla \mathbf{u}) + \mathbf{r} = 0, \tag{1}$$

where

$$\mathbf{u} \equiv \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix}$$

Here, $h \geqslant 0$ is the total depth of the water column and the velocity components in the x and y directions are u and v, respectively. We assume (1) holds over a domain $\Omega \subset \mathbb{R}^2$ with outward

normal $\mathbf{n}$ and spatially varying bathymetry represented by a function $b(x, y)$ that is strictly negative.

The flux matrix $\mathbf{F}(\mathbf{u})$ and diffusion tensor in (1) have the form

$$\mathbf{F} = \begin{pmatrix} u_2 & u_3 \\ \frac{u_2^2}{u_1} + \frac{g u_1^2}{2} & \frac{u_2 u_3}{u_1} \\ \frac{u_2 u_3}{u_1} & \frac{u_3^2}{u_1} + \frac{g u_1^2}{2} \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \nu & 0 \\ 0 & 0 & \nu \end{pmatrix}$$

with turbulent viscosity ν, and gravitational acceleration g. The source term $\mathbf{r}(\mathbf{u})$ contains the topography gradient and the Manning's bed friction term with coefficient n_{mn}

$$\mathbf{r} = gh \begin{pmatrix} 0 \\ \frac{\partial b}{\partial x} \\ \frac{\partial b}{\partial y} \end{pmatrix} + g n_{mn}^2 \frac{\sqrt{u_2^2 + u_3^2}}{u_1^{\frac{7}{3}}} \begin{pmatrix} 0 \\ u_2 \\ u_3 \end{pmatrix}.$$

Note that the fluid surface elevation is defined as $\eta = u_1 + b$, and the discharge is $(hu, hv)^T$. Everywhere below, we assume that b has as much continuity as required.

We assume that initial conditions $h^0(x, y), u^0(x, y)$ and $v^0(x, y)$ are specified for h, u, and v, respectively. For simplicity, we do not assume any Dirichlet boundary conditions for (1). Instead, we impose that $\mathbf{D}\nabla\mathbf{u} \cdot \mathbf{n} = 0$ on $\partial\Omega$ and, possibly, that $u_2 n_x + u_3 n_y = 0$ on portions of $\partial\Omega$ where no discharge is expected.

2.2 Stabilized Finite Element scheme

The classical finite element scheme for equation (1) with the considered boundary conditions relies on the following exact weak formulation. If $\mathbf{w}$ denotes an arbitrary test function from a reasonable test space, then

$$\int_\Omega \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{w} - \int_\Omega \mathbf{F} \cdot \nabla \mathbf{w} + \int_\Omega \mathbf{D}\nabla\mathbf{u} \cdot \nabla\mathbf{w} + \int_\Omega \mathbf{r} \cdot \mathbf{w} + \int_{\partial\Omega} (\mathbf{F} \cdot \mathbf{n}) \cdot \mathbf{w} = 0 \tag{2}$$

In the above, the first component of $\mathbf{F} \cdot \mathbf{n}$ in the boundary integral term is zero whenever the boundary condition $u_2 n_x + u_3 n_y = 0$ is imposed.

Let ϕ_j denote the j-th scalar shape function from a standard space of piecewise linear functions on a conformal mesh of Ω with N_n nodes. We can then approximate the components u_i of our solution in (2) via

$$u_i \approx u_i^h = \sum_{j=1}^{N_n} \phi_j v_i^j, \tag{3}$$

where $1 \leqslant i \leqslant 3$ refers to the unknown of (1) and $1 \leqslant j \leqslant N_n$. We label the trial space for the full unknown $\mathbf{u}^h$ as $\mathbf{V}_h$. Setting $\mathbf{w}^h$ to all of the available basis functions $\mathbf{w}_j^h$ from $\mathbf{W}_h = \mathbf{V}_h$ in (2), together with (3), yields a time-dependent nonlinear system of $N = 3N_n$ ordinary differential equations (semi-discrete scheme):

$$\begin{aligned} &\int_\Omega \frac{\partial \mathbf{u}^h}{\partial t} \cdot \mathbf{w}_j^h - \int_\Omega \mathbf{F}(\mathbf{u}^h) \cdot \nabla \mathbf{w}_j^h + \int_\Omega \mathbf{D}\nabla\mathbf{u}^h \cdot \nabla\mathbf{w}_j^h + \\ &\int_\Omega \mathbf{r}(\mathbf{u}^h) \cdot \mathbf{w}_j^h + \int_{\partial\Omega} \big(\mathbf{F}(\mathbf{u}^h) \cdot \mathbf{n}\big) \cdot \mathbf{w}_j^h = 0, \ j = 1, \ldots, 3N_n \end{aligned} \tag{4}$$

Regardless of the time discretization employed for (4), the system in general suffers from convection- and shock-type instabilities, especially for small viscosity ν or insufficient mesh resolution [15, 16], and calls for additional stabilization terms. Letting $a(\cdot,\cdot)$ denote the standard Galerkin approximation for the stationary part of (4), a generalization of classical SUPG stabilization via Hughes variational multiscale paradigm [16] can be written as

$$\int_\Omega \frac{\partial \mathbf{u}^h}{\partial t}\cdot \mathbf{w}^h + a(\mathbf{u}^h,\mathbf{w}^h) + \sum_e \int_{\Omega_e} \mathscr{M}_{\mathbf{u}^h}\mathbf{w}^h \cdot \boldsymbol{\tau}\mathbf{R}_h(\mathbf{u}^h) = 0, \tag{5}$$

where the vector operator $\mathscr{M}_{\mathbf{u}^h}$ is

$$\mathscr{M}_{\mathbf{u}^h}\mathbf{w}^h = -\sum_i \left(\mathbf{A}_{h,i}(\mathbf{u}^h)^T \frac{\partial \mathbf{w}^h}{\partial x_i}\right) + \frac{\partial \mathbf{r}}{\partial \mathbf{u}}^T(\mathbf{u}^h)\mathbf{w}^h \tag{6}$$

and $\mathbf{A}_{h,i}(\mathbf{u}^h)$ for $i = 1, 2$ are advection matrices

$$\mathbf{A}_{h,i}(\mathbf{u}^h) = \frac{\partial \mathbf{F}_i}{\partial \mathbf{u}}(\mathbf{u}^h) \tag{7}$$

and $\mathbf{F}_i$ denotes the i-th column of the matrix $\mathbf{F}$. Here $\boldsymbol{\tau}$ is a matrix of intrinsic time scales that can be specified using an algebraic subgrid scale (ASGS) approach [20]. $\mathbf{R}_h(\mathbf{u}^h)$ denotes the residual of the continuous formulation (1) (so-called strong residual), $x_1 = x$, $x_2 = y$, and $\mathbf{w}^h$ represents a discrete test function taken from $\mathbf{W}_h$.

For problems with strong nonlinearities and sharp fronts, we resort to adding a shock-capturing term having a form of an artificial viscosity [8]:

$$\sum_e \int_{\Omega_e} c_{sh}\frac{h_e}{2}\frac{\left\|\mathbf{R}_h(\mathbf{u}^h)\right\|}{\|\nabla \mathbf{u}^h\| + \delta}\nabla \mathbf{u}^h \cdot \nabla \mathbf{w}^h \tag{8}$$

where h_e denotes the characteristic width of an element Ω_e and c_{sh} is a positive scaling constant usually taken to be less than 1 and all the norms are standard 2-norms. Note that $\mathbf{R}_h$ does not contain diffusive terms since we are using piecewise linear shape functions, and small term δ is added to separate the denominator from zero.

Remark 1. *In the following, we consider stabilization only through the nonlinear shock-capturing term* (8)*, which in our experience is the dominant term for controlling instabilities due to sharp fronts and shocks. While simplifying the implementation, the reduction schemes presented here apply to the to the fine-scale scheme from [20] with only minor differences.*

The final form of our finite element scheme for the high-fidelity simulations is

$$\begin{aligned}&\int_\Omega \frac{\partial \mathbf{u}^h}{\partial t}\cdot \mathbf{w}^h - \int_\Omega \mathbf{F}(\mathbf{u}^h)\cdot\nabla \mathbf{w}^h + \int_\Omega \mathbf{D}\nabla \mathbf{u}^h \cdot \nabla \mathbf{w}^h + \int_\Omega \mathbf{r}(\mathbf{u}^h)\cdot \mathbf{w}^h + \\ &\int_{\partial\Omega}\left(\mathbf{F}(\mathbf{u}^h)\cdot \mathbf{n}\right)\cdot \mathbf{w}^h + \sum_e \int_{\Omega_e} c_{sh}\frac{h_e}{2}\frac{\left\|\mathbf{R}(\mathbf{u}^h)\right\|}{\|\nabla \mathbf{u}^h\| + \delta}\nabla \mathbf{u}^h \cdot \nabla \mathbf{w}^h = 0.\end{aligned} \tag{9}$$

The semi-discrete approximation in equation (9), expressed as a system of ordinary differential equations, can be viewed as

$$\mathbf{M}\dot{\mathbf{v}} + \mathbf{L}\mathbf{v} + \mathbf{f}(\mathbf{v}) = 0 \tag{10}$$

with solution $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T)^T$ consisting of the degrees of freedom for the approximate solution $\mathbf{u}^h$ and $\mathbf{f} = (\mathbf{f}_1^T, \mathbf{f}_2^T, \mathbf{f}_3^T)^T$. The term $\mathbf{f}$ comprises all the nonlinear terms of equation (9), including those with nonlinear stabilization. The vector components $\mathbf{f}_2$ and $\mathbf{f}_3$ also involve boundary integrals. Matrix $\mathbf{L}$ of the linear term is based on the diffusion and topography terms, and in the case of outflow boundaries may also contain the first component of $\int_\Omega (\mathbf{F}\cdot\mathbf{n})\cdot\mathbf{w}$. $\mathbf{M}$ is a standard mass matrix.

2.3 Time discretization

For marching in time from $t_0 = 0$ with a variable time step, Δt_k, we construct the fully discrete scheme using the first-order Euler approximation (higher order backward difference schemes can be used to increase accuracy in time). In order to avoid solving a large nonlinear algebraic system at each time step, a semi-implicit variant of the fully implicit scheme is proposed in such a way that extrapolation is introduced inside all nonlinear terms in (9). This lagging does not influence the accuracy of the scheme, but makes it only conditionally stable [17]. Everywhere below, any discrete in time function $\mathbf{f}$ at time level k is denoted as $\mathbf{f}^k$.

The resulting fine-scale linear system to be solved has the schematic form

$$\left(\underbrace{\frac{1}{\Delta t_k}\mathbf{M} + \mathbf{L}}_{\mathbf{A}} + \underbrace{\begin{pmatrix} \mathbf{N}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{N}_{21} & \mathbf{N}_{22} & \mathbf{N}_{23} \\ \mathbf{N}_{31} & \mathbf{N}_{32} & \mathbf{N}_{33} \end{pmatrix}^{k-1,k-2}}_{\mathbf{N}} \right) \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{pmatrix}^k = \frac{1}{\Delta t_k}\mathbf{M} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{pmatrix}^{k-1} . \tag{11}$$

Block $\mathbf{N}_{ij}$ corresponds to test function component w_i and solution component u_j. Note that the shock-capturing and the Manning friction terms only contribute to the diagonal terms of $\mathbf{N}$. Every block $\mathbf{N}_{ij}$ depends on the solution and has to be reassembled after every time step. Just for the first time step, we set $c_{sh} = 0$, as no information is given regarding $\mathbf{u}^{-1}$ in the mass part of the shock-capturing term, and so $\mathbf{N}$ depends on $\mathbf{v}^0$ only. In the case of strong Dirichlet boundary conditions, we include an additional vector $\mathbf{v}_d$ containing the boundary values to the right hand side and modify $\mathbf{A}$ and $\mathbf{N}_{ij}$, $i, j = 1, 2, 3$ accordingly.

3 GLOBAL BASIS REDUCED ORDER MODEL FRAMEWORK

The fine-scale approximation above allows calculation of a numerical solution that improves with mesh refinement. We assume that computational resources in the offline stage are sufficient to allow the temporal and spatial resolution necessary to resolve all relevant solution features. We then focus on the online stage where computational resources and/or time-constraints dictate that a much faster (i.e., reduced) model with possibly relaxed accuracy is needed. We first present the basic methodology for the reduction relying on a global empirical basis representing the coherent structures of the fine-scale processes. We use a generic algebraic formulation typical of semi-implicit temporal approximations without making specific reference to the SWE.

3.1 Proper Orthogonal Decomposition of the state variable for semi-implicit schemes

POD is probably the best-known projection-based model reduction scheme. It has been applied to a host of general, large-scale parameterized dynamical systems. Fundamentally, it is based on approximating the high-dimensional solution to the dynamical system in a low-dimensional subspace that is chosen by considering modes that capture a targeted portion of the system's energy [4, 10]. To obtain a dynamical system of low dimension, POD proceeds by

projecting the original high-dimensional equations on the solution subspace. Since it uses the same subspace for both the solution and projection steps, POD can be considered as a (Bubnov) Galerkin method [12].

Consider the following semi-implicit Euler discretization with variable time step Δt_i aimed at connecting the state variable at time level i with time levels $i-1$, $i-2$ that represents (among others) our current SWE scheme:

$$\mathbf{R}^i \equiv (\mathbf{A}^i + \mathbf{N}(\mathbf{v}^{i-2}, \mathbf{v}^{i-1}))\mathbf{v}^i - \Delta t_i^{-1}\mathbf{M}\mathbf{v}^{i-1} - \mathbf{d}^i = 0,$$

Here $\mathbf{R}$ denotes the residual of the fully discrete approximation. Matrix $\mathbf{A}^i$ comprises matrices from all the linear terms containing the solution at time level i and depends only Δt^i, while $\mathbf{N}$ consists of the matrices constructed through extrapolating the solution at previous time levels $i-2, i-1$ to the time level i in the nonlinear terms. Term $\mathbf{d}^i$ accounts for the external source and the boundary conditions data at level i and is state-independent. Again, such extrapolation is a common technique behind semi-implicit schemes that retains (theoretically) the same order of accuracy as for the fully implicit approximation, though in general it is not unconditionally stable.

To initiate the time integration, we assume that $\mathbf{v}^0$ is given at time level 0 via projection of the initial condition and that $\mathbf{N}(\mathbf{v}^{-1}, \mathbf{v}^0)$ is replaced with $\mathbf{N}_0(\mathbf{v}^0)$ for the first time step only. Therefore, we have the following fine-scale, fully discrete problem statement within the semi-implicit paradigm:

$$\begin{cases} \text{initial data } \mathbf{v}^0 \text{ given,} \\ (\mathbf{A}^0 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{v}^1 - \Delta t_1^{-1}\mathbf{M}\mathbf{v}^0 - \mathbf{d}^1 = 0, \\ (\mathbf{A}^i + \mathbf{N}(\mathbf{v}^{i-2}, \mathbf{v}^{i-1}))\mathbf{v}^i - \Delta t_i^{-1}\mathbf{M}\mathbf{v}^{i-1} - \mathbf{d}^i = 0, \quad i \geqslant 2 \end{cases} \tag{12}$$

We define $\mathbf{S} = (\mathbf{v}^1, \mathbf{v}^2, ..., \mathbf{v}^M)$ to be an $N \times M$ matrix of solution snapshots, $M \leqslant N$, obtained from the high precision offline runs from $t = 0$ to $t = T$ for the state variable $\mathbf{v}$. Note, that we do not include the initial data into the snapshot set. In practice, the set $\mathbf{S}$ may consist of snapshots at intermediate values of t from $[0, T]$, for each corresponding input parameter or forcing function. Also, unlike [12], we store solution values rather than increments, which will be shown to be adequate for the semi-implicit scheme above.

A thin SVD is used for the matrix of snapshots to achieve the sought basis:

$$\mathbf{S} = \mathbf{U}\text{diag}(\sigma_1, ..., \sigma_M)\mathbf{Q}^T, \tag{13}$$

with σ_i denoting the singular values in decreasing order, $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_M \geq 0$. Both $\mathbf{U}$ and $\mathbf{Q}$ consist of orthonormal vectors that satisfy

$$\mathbf{S}^T\mathbf{S}\mathbf{q}^j = (\sigma_j)^2\mathbf{q}^j, \quad \mathbf{S}\mathbf{S}^T\mathbf{u}_j = (\sigma_j)^2\mathbf{u}_j.$$

The columns $\mathbf{u}_j$ of the matrix $\mathbf{U} \in \mathbb{R}^{N\times M}$ provide the desired basis vectors, also called the *empirical modes* for the solution $\mathbf{v}$. They are ordered correspondingly with σ_j.

The use of the reduced-order approach is justified when we connect the reduced-order discrete solution $\mathbf{z}$ of size $m \leqslant M$ to the high-fidelity discrete solution via the approximation

$$\mathbf{v} \approx \mathbf{U}_r\mathbf{z}, \tag{14}$$

where $\mathbf{U}_r$ denotes the submatrix of $\mathbf{U}$ consisting only of the first m columns of $\mathbf{U}$ and $m \ll N$. Plugging this approximation into the original semi-discrete system makes it over-determined in general. Indeed, denote the approximate residual at time level i as

$$\mathbf{R}_r(\mathbf{z}^i) \equiv (\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i - \Delta t_i^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{d}^i. \tag{15}$$

Unlike the original high-dimensional model, this residual is not necessarily zero, due to non-exactness of (14). Thus some form of projection is required to reduce the number of the resulting equations, i.e., multiplication from the left by a full rank matrix of size $m \times N$. We consider two alternative approaches, namely a standard Galerkin (POD) projection as well as a Petrov-Galerkin projection [6, 12].

3.1.1 Galerkin projection

Standard Galerkin projection uses left multiplication by some constant (time and state-independent) matrix $\mathbf{V}^T$. The benefits of this approach stem directly from the fact that $\mathbf{V}^T$ is constant and allows for pre-assembly of various components of the reduced-order model that do not require update at every time step. The drawback is in that such projection is not based on an optimality criterion. Usually, a Bubnov-Galerkin approach is used in which $\mathbf{V} = \mathbf{U}_r$, but we consider arbitrary full-rank $\mathbf{V}$ to allow some flexibility. The reduced-order model relying on the Galerkin projection then takes the form

$$\mathbf{V}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i - \Delta t_i^{-1}\mathbf{V}^T\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{V}^T\mathbf{d}^i = 0, \tag{16}$$

The above system will have a unique solution $\mathbf{z}^i$ at every time step i, provided the matrix $\mathbf{V}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r^T$ always remains non-singular. When $\mathbf{V} = \mathbf{U}_r$, this is equivalent to $\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})$ being non-singular. With initial conditions taken into account, we obtain

$$\begin{cases} \text{initial data } \mathbf{v}^0 \text{ given,} \\ \mathbf{V}^T(\mathbf{A}^i + \mathbf{N}_0(\mathbf{v}^0))\mathbf{U}_r\mathbf{z}^1 - \Delta t_1^{-1}\mathbf{V}^T\mathbf{M}\mathbf{v}^0 - \mathbf{V}^T\mathbf{d}^1 = 0, \\ \mathbf{V}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i - \Delta t_i^{-1}\mathbf{V}^T\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{V}^T\mathbf{d}^i = 0, \quad i \geqslant 2 \end{cases} \tag{17}$$

In the last equation above, the notation $\mathbf{U}_r\mathbf{z}^0$ is simply used for the initial data $\mathbf{v}^0$ when $i = 2$. In the numerical experiments below, we will refer to (17) as the nonlinear POD (NPOD) approximation.

3.1.2 Petrov-Galerkin projection

We next consider a Petrov-Galerkin projection following the GNAT approach from [12]. Unlike the Galerkin (NPOD) approach above, it is constructed directly from the minimization considerations for the approximate residual at every time step. Considering (15), we pose the following problem at step i:

$$\mathbf{z}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \|\mathbf{R}_r(\mathbf{z})\|$$

This problem is equivalent to a standard least-squares problem obtained through the left multiplication of equation (15) by the transpose of $(\mathbf{A} + \mathbf{N})\mathbf{U}$ and setting the result to zero:

$$\mathbf{U}_r^T(\mathbf{A}^{i,T} + \mathbf{N}^T(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i = \\ \mathbf{U}_r^T(\mathbf{A}^{i,T} + \mathbf{N}^T(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))(\Delta t^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} + \mathbf{d}^i)$$

The above linear system for $\mathbf{z}^i$ is typically solved through a Moore-Penrose pseudoinverse of the rectangular matrix $(\mathbf{A}+\mathbf{N})\mathbf{U}_r$.

Analogous to the Galerkin reduced-order model, we construct the following reduced-order initial value problem:

$$\begin{cases} \text{initial data } \mathbf{v}^0 \text{ given,} \\ \mathbf{z}^1 = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| (\mathbf{A}^1 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{U}_r\mathbf{z} - \Delta t_1^{-1}\mathbf{M}\mathbf{v}^0 - \mathbf{d}^1 \right\|, \\ \mathbf{z}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| (\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z} - \Delta t_i^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{d}^i \right\|, \quad i \geqslant 2 \end{cases} \tag{18}$$

We will refer to (18) as the nonlinear Petrov-Galerkin (PG-NPOD) approximation below.

3.1.3 Consistency

As mentioned above, consistency is a desirable property for projection-based reduced-order models, since a consistent reduced order model will restore the fine-scale solution precisely when the entire empirical basis is used (that is, there is no truncation, meaning $\mathbf{U}_r = \mathbf{U}$). It can be shown that both the Galerkin (17) and Petrov-Galerkin (18) models are consistent under suitable assumptions on the time step selection and invertibility of the iteration matrices.

3.1.4 Multiple variables

For systems of PDEs, involving more than one variable (such as, for example, the shallow water system), it is a common strategy to collect separate snapshots sets for each of the solution components, to have more control over reduction errors corresponding to certain variables. For situations involving n variables, we construct n snapshot matrices $\mathbf{S}_j$, $1 \leqslant j \leqslant n$, corresponding to the same time instances, and apply the SVD procedure to each of them separately. The calculated bases $\mathbf{U}_j$ are then grouped block-wise along the diagonal for the global basis $\mathbf{U}$:

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{U}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{U}_n \end{pmatrix} \tag{19}$$

The truncation can then be done for each variable independently. In this case, when the whole bases $\mathbf{U}_j$ are used for all of the variables in the reduced-order models a consistency result can be shown for Galerkin projection via a constant, full-rank (not necessarily block-diagonal) matrix $\mathbf{V}^T$ and Petrov-Galerkin projection. This is due to the fact that every snapshot for the i-th variable belongs to range$(\mathbf{U}_i)$ for all M time levels.

3.2 Hyper-reduction for Galerkin projection

For reduced models based on Galerkin projection, precomputation of the constant matrix factor $\mathbf{V}^T\mathbf{M}\mathbf{U}_r$ can be done only once before marching in time and therein be used for cheap multiplication by $\mathbf{z}^{i-1}$ in equation (17) at every time step. Precomputation (with scalar multiplication) can also be done for the term $\mathbf{V}^T\mathbf{A}^i\mathbf{U}_r$. Unfortunately, the nonlinear piece $\mathbf{V}^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r$ requires recomputation in the original high-dimensional space and subsequent projection back to the reduced space at every time step. Such computational burden, involving operations with

complexity that scales linearly with the order N of the high-dimensional space, diminishes the execution speedup achieved by state reduction. To overcome this performance issue, further complexity reduction — this time applied to the terms calling for update at every step — is required.

To achieve this, we will consider hyper-reduction [2, 6]. The hyper-reduction approaches we consider rely on the same paradigm of expansion through a basis as we use for the solution state itself. We begin by constructing an additional snapshot matrix $\mathbf{T}$ for the nonlinear portion of the discrete residual during the offline stage. In the semi-implicit scheme (17), this corresponds to the action of the nonlinear part of the left-hand-side (LHS) matrix on the state vector $\mathbf{N}\mathbf{U}_r\mathbf{z}^i$. The aim of collecting these snapshots is to capture the range of the entire nonlinear portion with a corresponding global, empirical (SVD) basis denoted as $\mathbf{W}$. To be specific, during the run of the state-reduced model with fixed basis $\mathbf{U}_r$, the following snapshot set is accumulated during M time steps:

$$\mathbf{T}^G = \left(\mathbf{N}_0(\mathbf{v}^0)\mathbf{U}_r\mathbf{z}^1, \mathbf{N}(\mathbf{v}^0, \mathbf{U}_r\mathbf{z}^1)\mathbf{U}_r\mathbf{z}^2, \ldots, \mathbf{N}(\mathbf{U}_r\mathbf{z}^{M-2}, \mathbf{U}_r\mathbf{z}^{M-1})\mathbf{U}_r\mathbf{z}^M\right) \tag{20}$$

The empirical basis $\mathbf{W}$ is then found via SVD as $\mathbf{T}^G = \mathbf{W}\mathrm{diag}(\bar{\sigma}_1, \bar{\sigma}_2, ..., \bar{\sigma}_M)\bar{\mathbf{Q}}^T$. With this basis, we then attempt to approximate $\mathbf{N}\mathbf{U}_r\mathbf{z}^i$ as

$$\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})\mathbf{U}_r\mathbf{z}^i \approx \mathbf{W}_r\mathbf{c}^i$$

via a time-dependent coefficient vector $\mathbf{c}^i$, where $\mathbf{W}_r$ is a column-reduced submatrix from $\mathbf{W}$. The best possible approximation of such type in the standard 2-norm is a projection $\mathbf{c}^i = \mathbf{W}_r^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})\mathbf{U}_r\mathbf{z}^i$. This procedure, however, has no computational benefit, since it employs the computation of the entire N-dimensional vector. Two ways to avoid this pitfall are presented below.

As mentioned above, the two methods we consider for choosing $\mathbf{c}^i$ are DEIM [13] and gappy POD [14, 23]. Both rely on approximating nonlinearities at a collection of l_w indices. DEIM takes a collocation approach and uses the approximation

$$\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})\mathbf{U}_r\mathbf{z}^i \approx \mathbf{W}_r(\mathbf{P}^T\mathbf{W}_r)^{-1}\underbrace{\mathbf{P}^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})}_{\text{cheap evaluation}}\mathbf{U}_r\mathbf{z}^i.$$

where $\mathbf{P}$ is a sampling (or permutation) matrix. Gappy POD instead uses a least squares reconstruction to approximate the nonlinear term as

$$\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})\mathbf{U}_r\mathbf{z}^i \approx \mathbf{W}_r(\mathbf{P}^T\mathbf{W}_r)^{+}\underbrace{\mathbf{P}^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})}_{\text{cheap evaluation}}\mathbf{U}_r\mathbf{z}^i$$

Here $(\mathbf{P}^T\mathbf{W}_r)^+$ is the Moore-Penrose pseudoinverse. DEIM requires that the number of sampling indices $l_w = n_w$, while gappy POD requires $l_w \geqslant n_w$. In the case of $l_w = n_w$, the approaches are identical assuming the same set of indices are used [20].

Defining the matrix $\mathbf{G} = \mathbf{V}^T\mathbf{W}_r\left(\mathbf{P}^T\mathbf{W}_r\right)^{-1}$ or $\mathbf{G} = \mathbf{V}^T\mathbf{W}_r\left(\mathbf{P}^T\mathbf{W}_r\right)^{+}$ in the case of DEIM and gappy POD, respectively, we can write the hyper-reduced Galerkin model as

$$\begin{cases} \text{initial data } \mathbf{v}^0 \text{ given,} \\ (\mathbf{V}^T\mathbf{A}^1\mathbf{U}_r + \mathbf{G}\mathbf{P}^T\mathbf{N}_0(\mathbf{v}^0)\mathbf{U}_r)\bar{\mathbf{z}}^1 - \Delta t_1^{-1}\mathbf{V}^T\mathbf{M}\mathbf{v}^0 - \mathbf{V}^T\mathbf{d}^1 = 0, \\ (\mathbf{V}^T\mathbf{A}^i\mathbf{U}_r + \mathbf{G}\mathbf{P}^T\mathbf{N}(\mathbf{U}_r\bar{\mathbf{z}}^{i-2}, \mathbf{U}_r\bar{\mathbf{z}}^{i-1})\mathbf{U}_r)\bar{\mathbf{z}}^i - \\ \qquad\qquad - \Delta t_i^{-1}\mathbf{V}^T\mathbf{M}\mathbf{U}_r\bar{\mathbf{z}}^{i-1} - \mathbf{V}^T\mathbf{d}^i = 0, \quad i \geqslant 2 \end{cases} \tag{21}$$

Note that we add an overbar to the variable $\mathbf{z}$ to distinguish the solution of the hyper-reduced-order model (21) from that of the state-only reduced-order model in equation (17). We will refer to equation (21) with DEIM as POD-DEIM below and POD-GPOD when the hyper-reduction is done with gappy POD.

Remark 2. *It can be shown that equation* (21) *with* (20) *using either DEIM or gappy POD is consistent with respect to* (17) *given appropriate assumptions on the time step selection and invertibility of the matrices* $\mathbf{V}^T\mathbf{A}^i\mathbf{U}_r + \mathbf{G}\mathbf{P}^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})\mathbf{U}_r$. *That is, if the untruncated basis* $\mathbf{W}$ *is used on place of* $\mathbf{W}_r$ *in matrix* $\mathbf{G}$ *in equation* (21)*, then* $\bar{\mathbf{z}}^i = \mathbf{z}^i$.

A fine-scale analog of (20) for the collection of nonlinearity snapshots can be used:

$$\mathbf{T}^G_{fs} = \left(\mathbf{N}_0(\mathbf{v}^0)\mathbf{v}^1, \mathbf{N}(\mathbf{v}^0, \mathbf{v}^1)\mathbf{v}^2, \ldots, \mathbf{N}(\mathbf{v}^{M-2}, \mathbf{v}^{M-1})\mathbf{v}^M\right) \tag{22}$$

Although it loses the consistency property and introduces irreducible error even if the untruncated basis $\mathbf{W}_{fs}$ (obtained from an SVD applied to $\mathbf{T}^G_{fs}$) is used in the hyper-reduced model, this strategy is computationally more economical because it does not require simulation of the state-reduced (NPOD) model (17) with fixed $\mathbf{U}_r$ during the offline stage.

3.3 Hyper-reduction for the Petrov-Galerkin projection model

In this section, we follow the same hyper-reduction approach settling on the expansion of the term being reduced through its empirical basis $\mathbf{W}$ and evaluating only a subset of its rows specified by the matrix $\mathbf{P}$. The hyper-reduction applied to the Petrov-Galerkin (PG-NPOD) projection (18), however, differs from the Galerkin hyper-reduced model discussed above in two ways. First, if we aim to drop dependence of complexity of the solution on the fine-scale dimension, we are limited in our choice for the correct term to reduce. Second, the sufficient conditions for consistency of reduction are more strict than for the Galerkin-based model and in practice may be infeasible or at least impracticable. We consider these two issues below.

3.3.1 Proper choice of the reduction term

Assume we are aiming at performing hyper-reduction for the PG-NPOD model in equation (18). Whether we solve this minimization problem via QR-decomposition or transforming to a system of normal equations, the only way to reduce the computational complexity is by dropping the size of the total LHS matrix. Consider, for example, QR-decomposition of the LHS matrix in (18). If the reduction is performed on the vector $\mathbf{N}\mathbf{U}_r\mathbf{z}^i$ only, the high dimension N in the complexity of the QR-factorization process still emanates from the matrix $\mathbf{A}^i$:

$$(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i \approx (\mathbf{A}^i + \mathbf{W}_r(\mathbf{P}^T\mathbf{W}_r)^+\mathbf{P}^T\mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i$$

Therefore, the following reduction is proposed:

$$(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i \approx \mathbf{W}_r(\mathbf{P}^T\mathbf{W}_r)^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z}^i.$$

Then, since $\mathbf{W}_r$ consists of orthonormal columns, the minimization problem

$$\mathbf{z}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\|\mathbf{W}_r(\mathbf{P}^T\mathbf{W}_r)^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z} - \Delta t_i^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{d}^i\right\|$$

is equivalent to the problem

$$\mathbf{z}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\|(\mathbf{P}^T\mathbf{W}_r)^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z} - \Delta t_i^{-1}\mathbf{W}_r^T\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{W}_r^T\mathbf{d}^i\right\|,$$

in which the high-dimensional cost is removed, since the matrix $\mathbf{W}_r^T\mathbf{M}\mathbf{U}_r$ is precomputed before the time-integration and evaluation of $\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))$ scales with the number of sample rows of matrix $\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1})$.

Therefore, we are bound to record snapshots that are representative of the action of the *whole* LHS matrix on the reduced state to gain consistency. Note that for the above least-squares system to have unique solution, we require that $n_w \geqslant m$, i.e., the number of columns of $\mathbf{W}_r$ should be larger than or equal to the number of columns of $\mathbf{U}_r$. For Galerkin-projection with hyper-reduction, (21), no such restriction was required.

3.3.2 Proper choice of a snapshot set

Determining the consistency for a hyper-reduced version of (18) requires comparing the state-reduced (PG-NPOD) solution

$$\mathbf{z}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| (\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z} - \Delta t_i^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{d}^i \right\|$$

and

$$\bar{\mathbf{z}}_i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| \mathbf{W}(\mathbf{P}^T\mathbf{W})^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r\mathbf{z} - \Delta t_i^{-1}\mathbf{M}\mathbf{U}_r\mathbf{z}^{i-1} - \mathbf{d}^i \right\|.$$

For these solutions to be the same, we ask for the matrix of the first problem to be the same as for the second one:

$$\mathbf{W}(\mathbf{P}^T\mathbf{W})^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r = (\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r.$$

These conditions are fulfilled if we require that $(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r \in \operatorname{range}(\mathbf{W})$. Therefore, while collecting snapshots during the run with M time steps of the PG-NPOD model (18), we include columns of $(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r$ at every time step into the snapshot matrix:

$$\mathbf{T}^{PG,c} = \left((\mathbf{A}^1 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{U}_r, (\mathbf{A}^2 + \mathbf{N}(\mathbf{v}^0, \mathbf{U}_r\mathbf{z}^1))\mathbf{U}_r, \ldots, (\mathbf{A}^M + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{M-2}, \mathbf{U}_r\mathbf{z}^{M-1}))\mathbf{U}_r\right) \tag{23}$$

The resulting Petrov-Galerkin, hyper-reduced scheme can then be written

$$\begin{cases} \text{initial data } \mathbf{v}^0 \text{ given,} \\ \bar{\mathbf{z}}^1 = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| \mathbf{H}\mathbf{P}^T(\mathbf{A}^1 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{U}_r\mathbf{z} - \Delta t_1^{-1}\mathbf{W}_r^T\mathbf{M}\mathbf{v}^0 - \mathbf{W}_r^T\mathbf{d}^1 \right\|, \\ \bar{\mathbf{z}}^i = \underset{\mathbf{z}\in\mathbb{R}^m}{\operatorname{argmin}} \left\| \mathbf{H}\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\bar{\mathbf{z}}^{i-2}, \mathbf{U}_r\bar{\mathbf{z}}^{i-1}))\mathbf{U}_r\mathbf{z} - \right. \\ \qquad\qquad \left. - \Delta t_i^{-1}\mathbf{W}_r^T\mathbf{M}\mathbf{U}_r\bar{\mathbf{z}}^{i-1} - \mathbf{W}_r^T\mathbf{d}^i \right\|, \quad i \geqslant 2, \end{cases} \tag{24}$$

where $\mathbf{H} = (\mathbf{P}^T\mathbf{W}_r)^+$ denotes a general Moore-Penrose pseudoinverse of the $l_w \times n_w$ matrix $\mathbf{P}^T\mathbf{W}_r$ with $m \leqslant n_w \leqslant l_w \leqslant N$, and the index selection matrix $\mathbf{P}$ contains indices from either DEIM or gappy POD.

Remark 3. *The collection in* (23) *is the analog of snapshot selection strategy 3 in [12]. With it, the semi-implicit Petrov-Galerkin approximation* (24) *can be shown to be consistent under the assumption that* $(\mathbf{P}^T\mathbf{W}_r)^+\mathbf{P}^T(\mathbf{A}^i + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{i-2}, \mathbf{U}_r\mathbf{z}^{i-1}))\mathbf{U}_r$ *remains non-singular and identical time step sizes are used for the state-reduced and hyper-reduced simulations.*

We will refer to (24) as PG-DEIM or PG-GPOD below, depending on the type of hyper-reduction technique used.

In practice, the snapshot collection method in equation (23) may be too storage intensive. In this work, we resort to a less expensive, though not necessarily consistent strategy that saves the action of the matrix on the reduced state. In other words, we propose saving snapshots of the matrix component of the residual inside the argmin operator already evaluated at the minimizing state found from the state-reduced (PG-NPOD) model. This strategy is summarized in form of a new snapshot matrix $\mathbf{T}^{PG}$:

$$\begin{aligned}\mathbf{T}^{PG} = \big((\mathbf{A}^1 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{U}_r\mathbf{z}^1, (\mathbf{A}^2 + \mathbf{N}(\mathbf{v}^0, \mathbf{U}_r\mathbf{z}^1))\mathbf{U}_r\mathbf{z}^2, \dots, \\ (\mathbf{A}^M + \mathbf{N}(\mathbf{U}_r\mathbf{z}^{M-2}, \mathbf{U}_r\mathbf{z}^{M-1}))\mathbf{U}_r\mathbf{z}^M\big)\end{aligned} \tag{25}$$

The fine-scale analog for this strategy, which is expected to be even less consistent but cheaper, is

$$\mathbf{T}^{PG}_{fs} = \big((\mathbf{A}^1 + \mathbf{N}_0(\mathbf{v}^0))\mathbf{u}^1, (\mathbf{A}^2 + \mathbf{N}(\mathbf{v}^0, \mathbf{v}^1))\mathbf{v}^2, \dots, (\mathbf{A}^M + \mathbf{N}(\mathbf{v}^{M-2}, \mathbf{v}^{M-1}))\mathbf{u}^M\big) \tag{26}$$

We conclude this section by noting that in case of more than one variable in a system of PDEs, we can perform reduction for each component of the nonlinearity just as was done for reduction of the state in §3.1.4.

4 Numerical experiments

To evaluate the Galerkin and Petrov-Galerkin reduced-order models detailed above, we consider two classical test problems. The first is a one-dimensional Burgers equation example, which we use to explore consistency of the reduction schemes in detail. The second is a well-known circular dam break [19]. In both cases, the semi-discrete and discrete systems have the same basic forms (9) and (11), respectively, with the caveat that the Burgers approximation only involves the 11 block.

4.1 1D viscous Burgers equation

We begin with Burgers' equation on the unit interval with constant initial and boundary data

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} - \nu\frac{\partial^2 u}{\partial t^2} = 0, & x \in (0,1),\ t > 0 \\ u(0,t) = 1, & t \geqslant 0, \\ \nu\frac{\partial u}{\partial x}\big|_{x=1} = 0, & t \geqslant 0, \\ u(x,0) = 0, & x > 0. \end{cases} \tag{27}$$

For the computational mesh, we use $N = 101$ uniformly spaced nodes. We set the viscosity to $\nu = 0.01$ and time step to $\Delta t = 0.01$.

4.1.1 Galerkin-based state reduction

We first perform a fine-scale run until $T = 2$ to collect the state snapshots. If we perform a non-thin SVD on the snapshot set, we produce a basis $\mathbf{U}$ that reaches the fine-scale dimension N. The singular values in log-scale are shown in Fig. 1. We also measure the relative error in the state-reduced solution

$$\varepsilon_{POD} = \frac{\max_i \|\mathbf{v}^i - \mathbf{U}_r\mathbf{z}^i\|_2}{\max_i \|\mathbf{v}^i\|_2} \tag{28}$$

for each simulation. Fig. 2 plots ε_{POD} against the number of modes m for the state variable u. The lowest error value achieved over the modal spectrum $1 \leqslant m \leqslant 50$ is approximately $3.4 \cdot 10^{-5}$. The graphs of the fine-scale and Galerkin-based reduced-order (NPOD) solution are shown for two instances of time in Figs. 3 and 4.

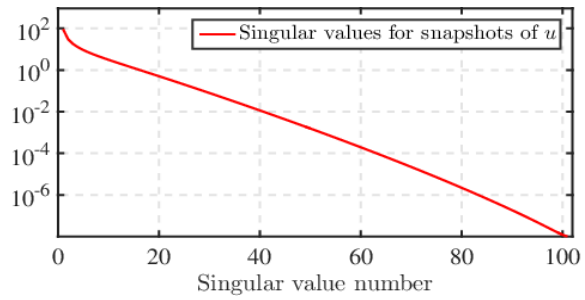


Figure 1: Singular values for state, u.

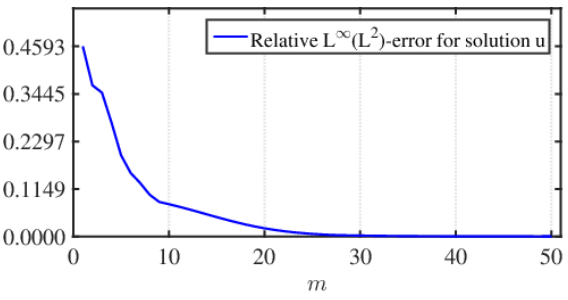


Figure 2: ε_{POD}

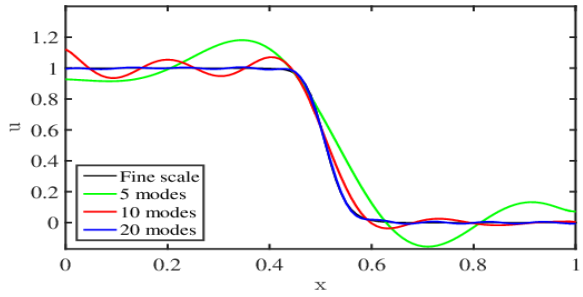


Figure 3: Solution curves at $t = 1$

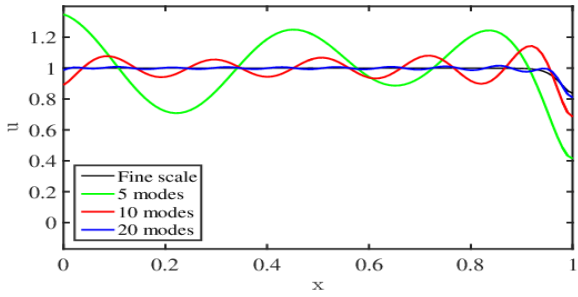


Figure 4: Solution curves at $t = 2$

4.1.2 Galerkin-based hyper-reduction

Next, the reduced-order model with hyper-reduction is investigated. During the fine-scale run of the previous subsection, the nonlinearity snapshots

$$\mathbf{T}_{fs}^{G} = (\mathbf{N}(\mathbf{v}^0)\mathbf{v}^1, \mathbf{N}(\mathbf{v}^1)\mathbf{v}^2, \cdots, \mathbf{N}(\mathbf{v}^{199})\mathbf{v}^{200})$$

were collected. During the state-reduced NPOD run with $m = 10$ modes ($\mathbf{U}_r \in \mathbb{R}^{101\times 10}$) we additionally collected snapshots as shown:

$$\mathbf{T}^{G} = (\mathbf{N}(\mathbf{v}^0)\mathbf{U}_r\mathbf{z}^1, \mathbf{N}(\mathbf{U}_r\mathbf{z}^1)\mathbf{U}_r\mathbf{z}^2, \cdots, \mathbf{N}(\mathbf{U}_r\mathbf{z}^{199})\mathbf{U}_r\mathbf{z}^{200})$$

Empirical bases were found via SVD and are denoted as $\mathbf{W}_{fs}$ and $\mathbf{W}$, respectively. Their singular values in log-scale are shown in Figs. 5 and 6.

DEIM and gappy POD sample index selection algorithms from [20] were applied to both, and the hyper-reduced models were constructed and tested. During the hyper-reduced simulations, the relative error between the hyper-reduced and NPOD solutions

$$\varepsilon_{hyp} = \frac{\max\limits_i \|\mathbf{U}_r\mathbf{z}^i - \mathbf{U}_r\bar{\mathbf{z}}^i\|_2}{\max\limits_i \|\mathbf{U}_r\mathbf{z}^i\|_2} = \frac{\max\limits_i \|\mathbf{z}^i - \bar{\mathbf{z}}^i\|_2}{\max\limits_i \|\mathbf{z}^i\|_2} \tag{29}$$

was recorded for varying $\mathbf{W}_r$. The results are shown in Figs. 7 and 8.

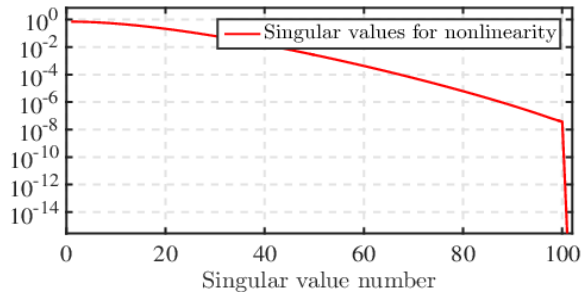


Figure 5: $\mathbf{T}^G_{fs}$ singular values

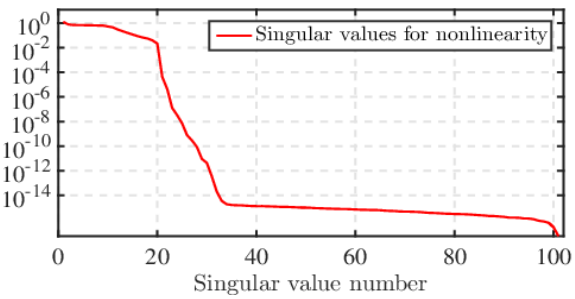


Figure 6: $\mathbf{T}^G$ singular values

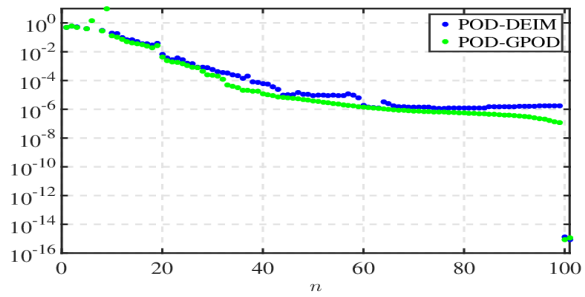


Figure 7: ε_{hyp} for $\mathbf{T}^G_{fs}$

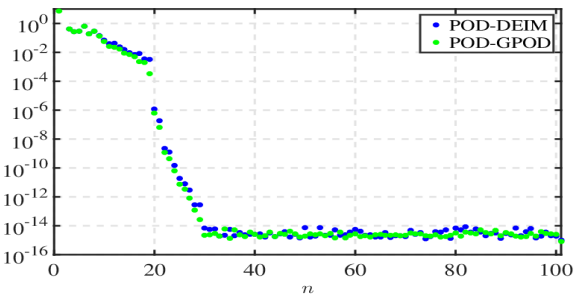


Figure 8: ε_{hyp} for $\mathbf{T}^G$

We observe the following. First, the consistency of the model with the fine-scale-based snapshot set $\mathbf{T}^G_{fs}$ is only achieved by reaching the fine-scale dimension for the nonlinearity, which is a trivial case of consistency, since this corresponds to an isomorphic change of the degrees of freedom. In other words, since the basis $\mathbf{W}_{fs}$ is less representative of the nonlinearity recorded in $\mathbf{T}^G$ than $\mathbf{W}$, the error stagnates before the rank of the $\mathbf{T}^G_{fs}$ is reached (the rank is $m = N - 1$ due to the constant Dirichlet boundary condition, as is also confirmed by Fig. 5). Second, the decay of the singular values toward machine precision at around $n_w \approx 32$ for $\mathbf{T}^G$, shown in Fig. 6 implies that the basis captures the nonlinearity throughout the simulation at this number of modes. We then see the expected drop in error to machine precision with $n_w \approx 32$ when using $\mathbf{T}^G$ as the snapshot set. Third, we see overall better accuracy of the gappy POD hyper-reduction strategy as opposed to DEIM, which is consistent with the results in [20]. Qualitatively similar graphs were obtained when $\mathbf{U}_r$ had $m = 5$ and 20 columns for the state reduction.

4.1.3 Petrov-Galerkin-based state reduction

We next consider the Petrov-Galerkin model without hyper-reduction (PG-NPOD). We use the same bases $\mathbf{U}$ for the state $\mathbf{v}$ as for the NPOD model, i.e., those obtained during the fine-scale run of (27) with final time $T = 2$. The solution curves obtained from the PG-NPOD model (not shown) were comparable to those presented in Figs. 3-4, for 5, 10, and 20 modes. We again calculated the relative $L^\infty(0, T; L^2)$-error (28) for the changing number of state modes, m, between 1 and 50. The lowest error value occurred when $m = 50$ and was equal to $3.4 \cdot 10^{-5}$. For the relatively smooth solution behavior in Fig. 3–4, the differences in accuracy between the NPOD and PG-NPOD approximations were negligible.

4.1.4 Petrov-Galerkin-based hyper-reduction

We finally try hyper-reduction for the Petrov-Galerkin reduced model given by equation (24). This time, we collected the nonlinearity snapshots $\mathbf{T}_{fs}^{PG}$ via (26) from the fine-scale runs, and $\mathbf{T}^{PG}$ from (25) using $m = 10$ during the state-reduced (PG-NPOD) runs. From these, we again used SVD to obtain the bases $\mathbf{W}_{fs}$ and $\mathbf{W}$, respectively. The singular values are shown in Figs. 9 and 10, respectively. The errors (29) coming from the runs of the hyper-reduced models,

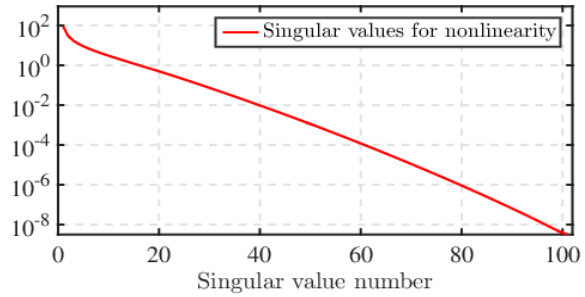


Figure 9: $\mathbf{T}_{fs}^{PG}$ singular values

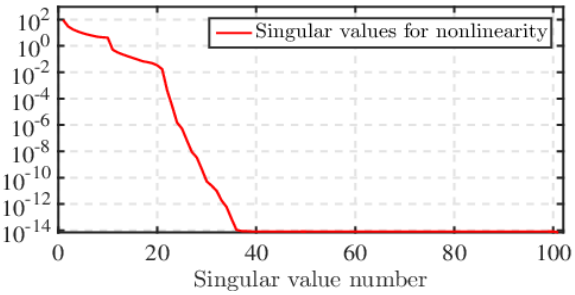


Figure 10: $\mathbf{T}^{PG}$ singular values

PG-DEIM and PG-GPOD, are presented in Figs. 11 and 12. Note that in the figures, the modal range starts with $n_w = m = 10$, due to the restriction that the hyper-reduced model have a unique least-squares solution.

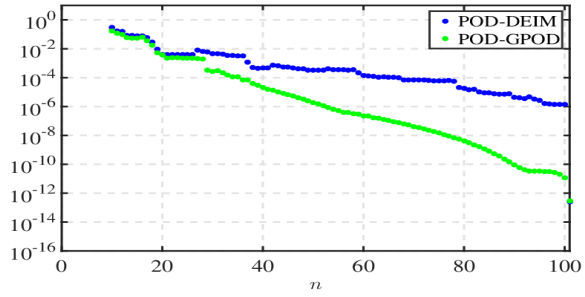


Figure 11: ε_{hyp} for $\mathbf{T}_{fs}^{PG}$

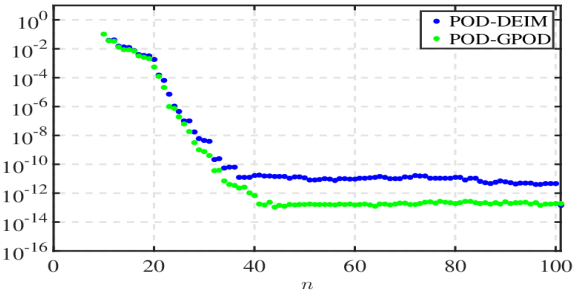


Figure 12: ε_{hyp} for $\mathbf{T}^{PG}$

Unlike the Galerkin-based models, here neither of the snapshot sets guarantees consistency, unless the entire basis is exhausted. Indeed, the drop to machine precision for the fine-scale snapshot set $\mathbf{T}_{fs}^{PG}$ occurs only when the fine-scale dimension is reached. The basis $\mathbf{W}_{fs}$ with gappy POD predicts the solution from the Petrov-Galerkin state-reduced model (PG-NPOD) much better than when the DEIM technique is used. The results with $\mathbf{T}^{PG}$ are much better with both hyper-reduction techniques, although the DEIM error curve demonstrates a small irreducible error even after reaching the effective rank of $\mathbf{T}^{PG}$, unlike gappy POD, which manages to reach machine precision accuracy even though the snapshot set $\mathbf{T}^{PG}$ is not (provably) consistent.

4.2 2D shallow water

In this section, the FEM scheme from equation (9) discretized in time via (11) is used for the simulation of an SWE problem, for which the geometry, topography, and the initial conditions were taken from [19]. The domain is a square $[0, 50] \times [0, 50]$, meshed uniformly by 20,000

equal triangles having sides $1/2$, $1/2$ and $\sqrt{2}/2$. This mesh generates $N_n = 10201$ nodes in total, and that results in three times as many, i.e., $N = 30603$ degrees of freedom for the entire FEM-discretized system. The topography function $\mathbf{b} \equiv -1$ throughout the domain with gravity $g = 9.81$, and we additionally set $\nu = 10^{-4}$ for viscosity and $n_{mn} = 1.5 \cdot 10^{-4}$ for the Manning coefficient. For the initial conditions, the elevation was given in a form of a cylindrical dam in the center of the square. In terms of depth,

$$u_1^0 = \begin{cases} 10, & \text{for } \sqrt{(x-25)^2 + (y-25)^2} \leqslant 11, \\ 1, & \text{otherwise} \end{cases}$$

and $u_2^0 = u_3^0 = 0$. Finally, wall boundary boundary conditions were assumed on all four sides

The discontinuity implied by the initial condition is expected to be handled during the time integration by the shock-capturing term that is incorporated in the scheme (9). As time evolution develops, the dam breaks into a wave traveling radially from the center of the square, and then reflects from the boundaries. The fine-scale simulation, as well as the subsequent reduced-order simulations, were performed with $\Delta t = 0.005$ until $T = 5$, which corresponds to 1000 steps. The results of the fine-scale simulation for different time instances are shown in Figs. 13–16. Since the depth in the fine solution remains bounded well away from 0 (the vacuum or dry state)

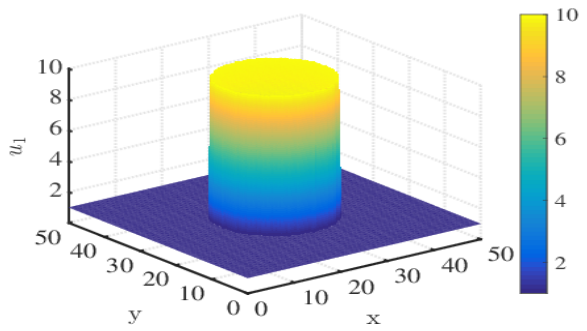


Figure 13: Initial depth

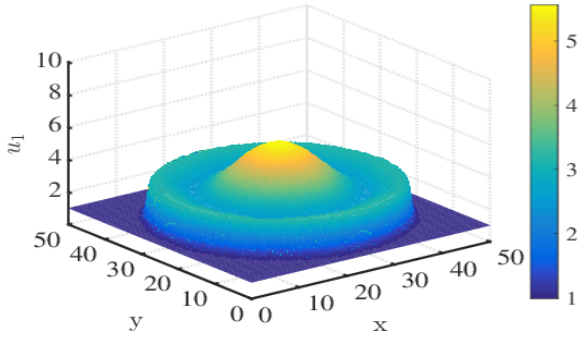


Figure 14: Depth at $t = 1.25$

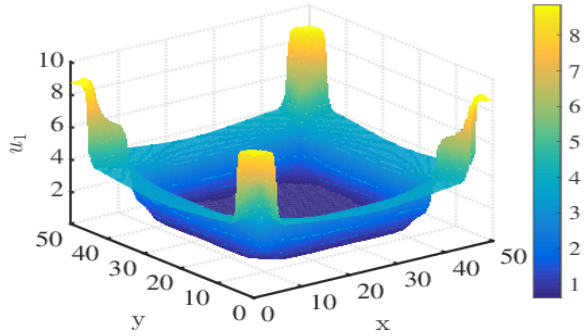


Figure 15: Depth at $t = 3.5$

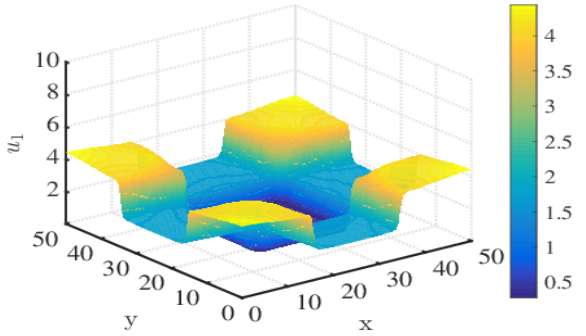


Figure 16: Depth at $t = 5$(final time)

in the fine scale simulations, we require that the various model reduction techniques considered maintain $h \geqslant 0$. Otherwise, we mark the simulation as failed, rather than engage a wetting and drying routine.

During the fine-scale simulations, state snapshots, $\mathbf{S}$, were collected separately for each solution component. The nonlinear snapshots $\mathbf{T}_{fs}^{G}$ and $\mathbf{T}_{fs}^{PG}$ were also collected separately for each equation in the SWE system. We will use a superscript $i = 1, 2, 3$ to distinguish the snapshot matrices when necessary below.

While calculations were performed with both the Galerkin and Petrov-Galerkin reduced models, we present only the Petrov-Galerkin results below for brevity. As before, the SVD bases calculated from $\mathbf{T}_{fs}^{PG}$ will be denoted $\mathbf{W}_{fs}^{i}$ with corresponding index $1 \leqslant i \leqslant 3$, while those from $\mathbf{T}^{PG}$ will be denoted $\mathbf{W}^{i}$.

4.2.1 State-reduced Petrov-Galerkin model

The solution snapshots for each of the three variables were decomposed via SVD to obtain bases $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$, all of size 10201×1000, for these variables respectively. The global basis is then made of $\mathbf{U}_i$ placed along the diagonal block-wise, as in (19). We can then write the PG-NPOD update at each time step as

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{pmatrix}^{k} = \underset{\mathbf{z}\in\mathbb{R}^{3m}}{\operatorname{argmin}} \left\| \left(\mathbf{A}^k + \mathbf{N}^{k-1,k-2}\right) \mathbf{U}_r \mathbf{z} - \frac{1}{\Delta t_k} \mathbf{M} \mathbf{U}_r \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{pmatrix}^{k-1} \right\|. \tag{30}$$

The depth solution obtained from the PG-NPOD scheme with $m = 50$ is shown in Figs. 17 and 18, together with absolute values of the components of $\mathbf{v}_1 - \mathbf{U}_{1,r}\mathbf{z}_1$ for depth at the same time instances. The similarity of Figs. 17-18 to the fine-scale solution Figs. 14-16 is apparent. The relative error for the depth $h = \mathbf{u}_1$, L^∞ in time and L^2 in space, ε_{POD}, was 0.0504.

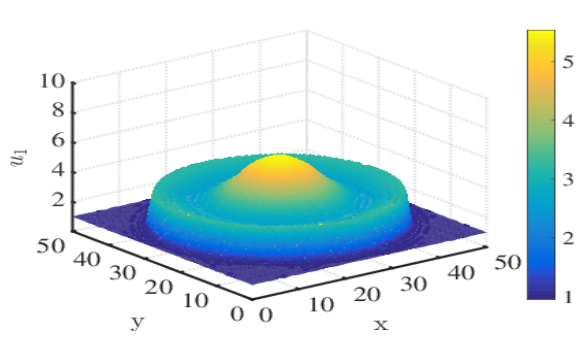


Figure 17: PG-NPOD solution at $t = 1.25$

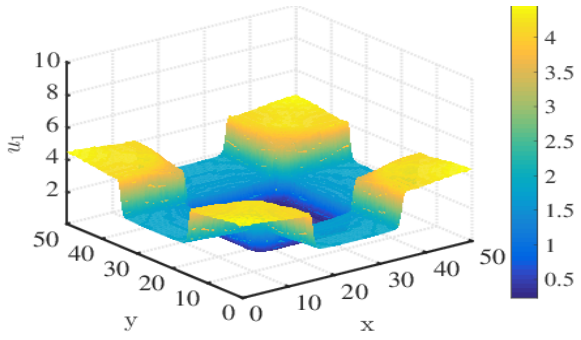


Figure 18: PG-NPOD solution at final time $t = 5$

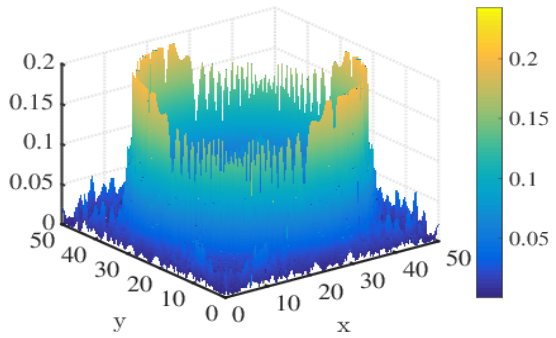


Figure 19: PG-NPOD absolute error at $t = 1.25$

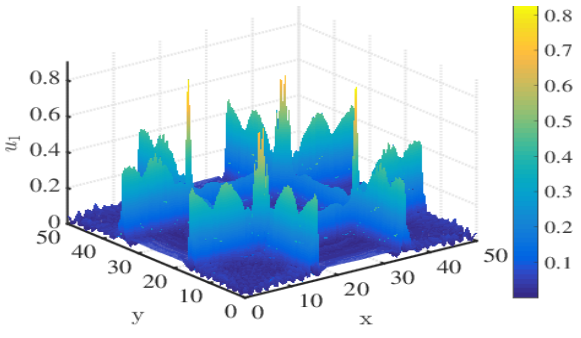


Figure 20: PG-NPOD absolute error at $t = 5$

4.3 Hyper-reduction

Four scenarios in total were used for the hyper-reduction assessment of the Petrov-Galerkin (GNAT-based) scheme. We considered: DEIM and gappy POD for $\mathbf{T}_{fs}^{PG}$ computed from the

fine-scale simulations as well as $\mathbf{T}^{PG}$ computed from PG-NPOD simulations with $m = 50$ modes for each component of the state variable, $\mathbf{u}$. In other words, we had a reduced basis for each state component, $\mathbf{U}_i \in \mathbb{R}^{10201\times 50}$, $i = 1, 2, 3$. We used the same number of nonlinear modes, n_w, for each snapshot basis $\mathbf{W}^i$ and $\mathbf{W}^i_{fs}$ for $i = 1, 2, 3$, and $50 \leqslant n_w \leqslant 1000$. We again measured the error (29) as before but for simplicity report only the error for the depth variable, u_1.

4.3.1 Petrov-Galerkin projection approximation

The DEIM method did not perform well for the bases $\mathbf{W}_{fs}$. No simulation completed the full simulation to $T = 5$ without violating the depth-positivity constraint for n_w in the range $50 \leqslant n_w \leqslant 1000$. The results with using $\mathbf{T}^{PG}$ were slightly better. Within the range $50 \leqslant n_w \leqslant 1000$ sampled in increments of 10, six runs successfully completed the simulation. The errors for these runs are shown in Table 4.3.1. Note, the error values are multiplied by 100.

# of hyper-modes n_w	210	230	520	550	650	740
error $\varepsilon_{hyp} \cdot 10^2$	6.36	6.53	4.587	5.12	2.5	2.86

Table 1: ε_{hyp} for PG-DEIM using snapshots $\mathbf{T}^{PG}$

As with DEIM, gappy POD hyper-reduction for the fine-scale bases $\mathbf{W}_{fs}$, did not produce a simulation that completed to $T = 5$ without violating the positivity constraint. On the other hand, gappy POD hyper-reduction using the bases $\mathbf{W}$ obtained from the PG-NPOD simulations was significantly better, showing both higher accuracy and more robustness. The relative error, ε_{hyp}, for $10 \leqslant n_w \leqslant 1000$ (sampled in multiples of 10) is shown in Fig. 21.

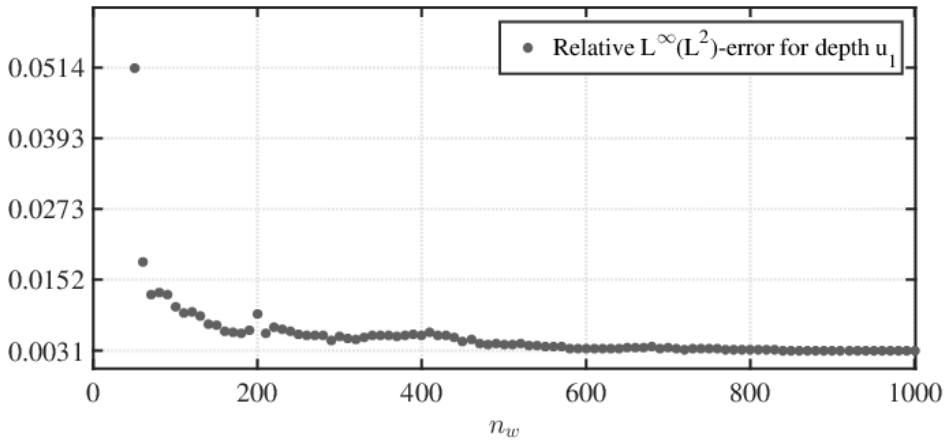


Figure 21: ε_{hyp} for the PG-GPOD model with varying number of empirical modes, n_w

5 CONCLUSION

In this work, we presented a fine-scale scheme for the SWE based on a stabilized FEM approximation and semi-implicit time integration. We then formulated reduced-order models

for this approximation based on both Galerkin and Petrov-Galerkin projection. We introduced hyper-reduction to the models via both DEIM and gappy POD techniques and considered alternative snapshot selection strategies and their consistency.

We performed two sets of numerical experiments to explore the schemes' performance. Numerical experiments for viscous Burgers equation clearly supported theoretical predictions concerning consistency of both the Galerkin and Petrov-Galerkin frameworks. The shallow water example was significantly more challenging but confirmed the trends observed in the Burgers equation results. Specifically,

- Irreducible errors were observed in all scenarios when a basis generated from fine-scale simulations directly was used.
- For Galerkin projection, the consistency predictions were confirmed by simulations with untruncated bases that were constructed from NPOD computations.
- Gappy POD hyper-reduction was more robust than DEIM in both the Galerkin and Petrov-Galerkin frameworks.
- The POD-GPOD and PG-GPOD approximations using snapshots $\mathbf{T}^G$ and $\mathbf{T}^{PG}$, respectively, are the best candidates for robust online solvers in sharp-front SWE problems like the dam-break example considered here.

References

[1] V. Aizinger and C. N. Dawson. A discontinuous Galerkin method for two-dimensional flow and transport in shallow water. *Advances in Water Resources*, 25(1):67–84, 2002.

[2] D. Amsallem, M. Zahr, Y. Choi, and C. Farhat. Design optimization using hyper-reduced-order models. *Structural and Multidisciplinary Optimization*, 51(4):919–940, 2015. ISSN 1615-147X. doi: 10.1007/s00158-014-1183-y. URL `http://dx.doi.org/10.1007/s00158-014-1183-y`.

[3] A.C. Antoulas. *Approximation of large-scale dynamical systems*, volume 6. SIAM, 2005.

[4] A.C. Antoulas and D.C. Sorensen. Approximation of large-scale dynamical systems: An overview. *International Journal of Applied Mathematics and Computer Science*, 11(5): 1093–1121, 2001.

[5] M. Barrault, Y. Maday, N.C. Nguyen, and A.T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[6] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.

[7] S.W. Bova and G.F. Carey. A symmetric formulation and SUPG scheme for the shallow-water equations. *Advances in Water Resources*, 19(3):123–131, 1996.

[8] E. Burman and A. Ern. Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection-diffusion-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 191:3833–3855, 2002.

[9] T. Butler, L. Graham, D. Estep, C.N. Dawson, and J.J. Westerink. Definition and solution of a stochastic inverse problem for the Mannings n parameter field in hydrodynamic models. *Advances in Water Resources*, 78:60–79, 2015.

[10] M.A. Cardoso, L.J. Durlofsky, and P. Sarma. Development and application of reduced-order modeling procedures for subsurface flow simulation. *International Journal for Numerical Methods in Engineering*, 77(9):1322–1350, 2009.

[11] K. Carlberg, Charbel B.-M., and C. Farhat. Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011.

[12] K. Carlberg, C. Farhat, and D. Amsallem. The GNAT method for nonlinear model reduction: effective implementation to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623–647, 2013.

[13] S. Chaturantabut and D.C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[14] R. Everson and L. Sirovich. Karhunen-Loéve procedure for gappy data. *Journal of the Optical Society of America A*, 12(8):1657–1664, 1995.

[15] T.J.R. Hughes and M. Mallet. A new finite element formulation for computational fluid dynamics: III. The generalized streamline operator for multidimensional advective-diffusive systems. *Computer Methods in Applied Mechanics and Engineering*, 58(3):305–328, 1986.

[16] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, and J.-B. Quincy. The variational multiscale method – a paradigm for computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 166(1):3–24, 1998.

[17] W. Hundsdorfer and J. G. Verwer. *Numerical solution of time-dependent advection-diffusion-reaction equations*. Springer-Verlag Publishers, Berlin, Germany, 2003.

[18] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, New York, NY, 2002.

[19] R. Liska and B. Wendroff. 2d shallow water equations by composite schemes. *International Journal for Numerical Methods in Fluids*, 30(4):461–479, 1999.

[20] A. Lozovskiy, M.W. Farthing, C.E. Kees, and E. Gildin. POD-based model reduction for stabilized finite element approximations of shallow water flows. *Journal of Computational and Applied Mathematics*, page In press, 2016.

[21] P. Ortiz. Non-oscillatory continuous FEM for transport and shallow water flows. *Computer Methods in Applied Mechanics and Engineering*, 223–224:55–69, 2012.

[22] C.B. Vreugdenhil. *Numerical methods for shallow-water flow*. Kluwer Academic Publishers, 1992.

[23] K. Willcox. Unsteady flow sensing and estimation via the gappy Proper Orthogonal Decomposition. *Computers and Fluids*, 35(2):208–226, 2006.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EFFICIENT TWO-DIMENSIONAL SIMULATION MODELS FOR HYDRAULIC AND MORPHODYNAMIC TRANSIENTS

Pilar Garcia-Navarro[1], Javier Murillo[1], Mario Morales-Hernandez[1], Carmelo Juez[2], and Asier Lacasta[1]

[1](1) Computational Hydraulics Group. University of Zaragoza. Spain.
e-mail: pigar@unizar.es, jnurillo@unizar.es, mmorales@unizar.es, alacasta@unizar.es

[2] EPFL, Lausanne, Switzerland.
e-mail: carmelo.juez@epfl.ch

Keywords: CUDA; GPU; Landslides; Numerical modeling; Shallow Flow; Coulomb forces; bed load transport, Exner equation.

Abstract. *Recent advances in the simulation of shallow flows over mobile bed have shown that accurate and stable results in realistic problems can be provided if an appropriate coupling between the shallow water equations (SWE) and the Exner equation is performed. In this way the computational cost may become unaffordable in situations involving large time and space scales. Therefore, for restoring the numerical efficiency, the coupling technique is simplified, not decreasing the number of waves involved in the Riemann problem but simplifying their definitions. The effects of the approximations made are tested against experimental data which include transient problems over erodible bed. The simplified model is formulated under a general framework able to insert any desirable discharge solid load formula. Also, the movement of poorly sorted material over steep areas constitutes a hazardous environmental problem. Computational tools help in the understanding and predictions of such landslides. The main drawback is the high computational effort required for obtaining accurate numerical solutions due to the high number of cells involved. However, recent advances in massive parallelization techniques for 2D hydraulic models are able to reduce computer times by orders of magnitude making 2D applications competitive and practical for operational flood prediction in large river reaches. Moreover, high performance code development can take advantage of general purpose and inexpensive Graphical Processing Units (GPU), allowing to run 2D simulations more than 100 times faster than old generation 2D codes, in some cases.*

1 INTRODUCTION

Landslides play an important role in the evolution of landscape and constitute an important environmental topic. They can be responsible for dramatic civil damages and that is the reason why the building of defenses and barriers is required. The computational tools are a suitable partner for developing a careful design of such elements. Over recent years, reliable predictions of the spreading of granular material have been obtained [21, 22, 17, 3] and numerical results have been validated with respect to series of experiments based on granular dry flows [**?**, 8, 23, 14, 15]. In particular, [19, 9] have recently presented a robust finite volume upwind scheme which includes the presence of steep slopes leading to obtain promising results.

Under the presence of complex topography or the presence of hydraulic structures, the use of 2D or 3D hydrodynamic models may be required. 2D depth averaged models are widely accepted for most practical purposes in complex cases. These models provide predictions for the water depth and the two-dimensional, depth averaged, flow velocity field at the cost of a fine topographic representation. The bed evolution is frequently computed through the Exner equation. Asynchronous techniques are based on the assumption that morphodynamic time scales are not relevant enough for altering the hydrodynamic variables within the interval of a computational time step. Therefore, the fluid mass and momentum equations are solved apart from (decoupled of) the Exner equation. Conversely, synchronous procedures assume that changes in the morphodynamic and hydrodynamic quantities take place within the same time scale, i.e. equations for both phases are solved at the same time and with the same time restriction. As stated in [10], unsteady flows with a wide range of hydrodynamic and morphodynamic situations can only be properly tackled by means of a synchronous technique.

Once the forecasting capacity of the computational tool has been reached another important concern is the improvement of the efficiency in terms of the computational cost. This type of geophysical flow involves the study of huge domains where the accuracy of the results is tied to the resolution of the Digital Terrain Model considered. Hence, a high number of cells is usually needed in the simulation. For this reason, the hardware GPU emerges as a promising strategy for handling this environmental and up to date problem.

In terms of scientific computation, the last four decades have followed Moore's law [16] where the number of transistors on a chip increased exponentially. This integration allowed to obtain faster and faster applications just recompiling the code for this new processors. Unfortunately, power has become the primary design constraint for chip designers, where both energy and power dissipation create a technological barrier for the integration capacity [4]. Nevertheless, Multi-Core micro architecture together with an adequate programming model (OpenMP is one of the most extended) brings a chance to exploit the parallelism of some parts of the code [24]. Moreover, Multi-Core paradigm has a large power consumption rate when performing small tasks and, for these purposes the Many-Core systems appears to be a very interesting option [2]. Many-Core architectures are those composed of smaller and not so complex cores that, usually have special purposes. Industrial implementation of this solution has been obtained in the field of Graphical Processing, where several efforts have been devoted to make more powerful devices. Indeed, this technology has been historically oriented to a very particular task of performing shading operations when rendering graphics. The purpose of the present work is to apply this hardware to the simulation of hazardous and high time consuming geophysical flows.

2 MATHEMATICAL MODEL

2.1 Equations

The mathematical model considered for reproducing the landslides phenomenon is based on the shallow flow equations, where the general three-dimensional conservation laws are depth averaged. The pressure distribution is considered hydrostatic and as frictional terms, only Coulomb type friction forces are assumed so that the 2D hydrodynamic equations are written in global coordinates as follows:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} + \frac{\partial \mathbf{G}(\mathbf{U})}{\partial y} = \mathbf{S}_\tau + \mathbf{S}_b \tag{1}$$

where

$$\mathbf{U} = (h, hu, hv)^T \tag{2}$$

are the conserved variables with h representing granular material depth in the z coordinate and (u, v) the depth averaged components of the velocity vector. The fluxes are given by

$$\begin{aligned} \mathbf{F} &= \left(hu, hu^2 + \frac{1}{2} g_\psi h^2, huv \right)^T \\ \mathbf{G} &= \left(hv, huv, hv^2 + \frac{1}{2} g_\psi h^2 \right)^T \end{aligned} \tag{3}$$

with $g_\psi = g \cos^2 \psi$ and ψ the direction cosine of the bed normal with respect to the vertical. The physical basis of this gravity projection is explained in [9] and it is of utmost importance for not ruining the numerical predictions when the simulation involves the presence of steep slopes.

The term $\mathbf{S}_\tau$ notes the frictional effects in the bed, and is defined as

$$\mathbf{S}_\tau = \left(0, -\frac{\tau_{b,x}}{\rho}, -\frac{\tau_{b,y}}{\rho} \right)^T \tag{4}$$

with $\tau_{b,x}, \tau_{b,y}$ the bed shear stress in the x and y directions respectively and ρ the density of the granular mass. Since the geophysical flows considered in this work are dense, the main rheological properties are governed by the frictional forces. These interactions between the sand grains are computed by means of the Manning and/or Coulomb laws depending on the case. The latter is based on the internal friction angle of the material, θ_b.On the other hand, the term $\mathbf{S}_b$ is defined for gathering the information relative to the pressure force exerted over the bottom.

Thanks to the hyperbolic character of (1) it is possible to obtain a Jacobian matrix, $\mathbf{J_n}$, which is built by means of the flux normal to a direction given by the unit vector $\mathbf{n}$, $\mathbf{E_n} = \mathbf{F}n_x + \mathbf{G}n_y$,

$$\mathbf{J_n} = \frac{\partial \mathbf{E_n}}{\partial \mathbf{U}} = \frac{\partial \mathbf{F}}{\partial \mathbf{U}} n_x + \frac{\partial \mathbf{G}}{\partial \mathbf{U}} n_y \tag{5}$$

whose components are

$$\mathbf{J}_n = \begin{pmatrix} 0 & n_x & n_y \\ (g_\psi h - u^2) n_x - uvn_y & vn_y + 2un_x & un_y \\ (g_\psi h - v^2) n_y - uvn_x & vn_x & un_x + 2vn_y \end{pmatrix} \tag{6}$$

The eigenvalues of this Jacobian matrix constitute the basis of the upwind technique which is detailed in the next subsection.

On the other hand, the bed evolution is modeled through the Exner equation, which is basically a movable bed continuity equation where the bed level time variations are due to the solid fluxes which cross the control volume. In this work the authors only focus on highly concentrated bed-load phenomena and, consequently, the 2D Exner equation is:

$$\frac{\partial z}{\partial t} + \xi \frac{\partial q_{s,x}}{\partial x} + \xi \frac{\partial q_{s,y}}{\partial y} = 0 \tag{7}$$

where $\xi = \frac{1}{1-p}$, p is the material porosity and $q_{s,x}$, $q_{s,y}$ are the solid fluxes. They are computed as a function of excess bed shear stress with respect to the critical value and taking into account the bed shear stress direction. This bedload transport is often expressed through the following dimensionless parameter:

$$\Phi = \frac{|\mathbf{q}_s|}{\sqrt{g_\psi (s-1) d_m^3}} \tag{8}$$

where $s = \rho_s / \rho_w$ is the ratio of solid material (ρ_s) over water (ρ_w) densities, and d_m is the grain median diameter. According to the numerical assessment performed in [9] the empirical Smart (1984) formula is chosen for computing the dimensionless bedload discharge as follows:

$$\Phi = 4 \ (d_{90}/d_{30})^{0.2} \ F \ S^{0.1} \theta^{1/2} (\theta - \theta_c^S) \tag{9}$$

where S is the velocity vector projected over the bed slope vector, as in [9], for distinguishing between positive and negative sloping beds. On the other hand d_{90} and d_{30} are grain diameter values for which 90% and 30% of the weight of a nonuniform sample is finer respectively. F is the Froude number, θ is the dimensionless shear stress and θ_c^S is the critical shear stress. This formula is only applied when the shear stress is larger than the critical shear stress. Otherwise there is no sediment transport.

2.2 Numerical method

The numerical scheme is constructed by defining an approximate Jacobian matrix $\widetilde{\mathbf{J}}$ at each k edge between neighboring cells defined through the normal flux $\mathbf{E_n}$ so that the volume integral in the cell at time t^{n+1} is expressed as:

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{k=1}^{NE} \sum_{m=1}^{3} (\widetilde{\lambda}^- \widetilde{\alpha} - \widetilde{\beta}^-)_k^m \widetilde{\mathbf{e}}_k^m l_k \frac{\Delta t}{A_i} \tag{10}$$

The superscript minus in (10) implies that only the incoming waves are considered for updating the flow variables of each cell, defining $\widetilde{\lambda}^- = \frac{1}{2}\left(\widetilde{\lambda} - \left|\widetilde{\lambda}\right|\right)$. Further, special care is considered when calculating wet/dry fronts. The strategy proposed is based on enforcing positive values of interface discrete water depths coming from a detailed study of the Riemann problem [18]. When they become negative, the numerical values of the friction and bed slope source terms is reduced instead of diminishing the time step.

Equation (7) is also integrated in a grid cell Ω_i. Using Gauss theorem:

$$\frac{\partial}{\partial t} \int_{\Omega_i} z d\Omega + \oint_{\partial \Omega_i} q_{s\mathbf{n}} dl = 0 \tag{11}$$

where $q_{s\mathbf{n}} = (q_{s,x}n_x + q_{s,y}n_y)$.

Assuming a piecewise representation of the variable z and that the second integral can be written as the sum of fluxes across the cell edges, the bed level is updated as:

$$z_i^{n+1} = z_i^n - \sum_{k=1}^{NE} \xi q^*_{s\mathbf{n},k} \frac{\Delta t \; l_k}{A_i} \tag{12}$$

where:

$$q^*_{s\mathbf{n},k} = \begin{cases} q_{s\mathbf{n},i} & \text{if} \quad \widetilde{\lambda}_s > 0 \\ q_{s\mathbf{n},j} & \text{if} \quad \widetilde{\lambda}_s < 0 \end{cases} \tag{13}$$

where $q_{s\mathbf{n},i}$ and $q_{s\mathbf{n},j}$ are the bed load discharge computed at the neighboring cells i, j, and $\widetilde{\lambda}_s$ is the numerical bed celerity estimated as:

$$\widetilde{\lambda}_s = \frac{\delta q_{s\mathbf{n},k}}{\delta z_k} \tag{14}$$

with $\delta q_{s\mathbf{n},k} = q_{s\mathbf{n},j} - q_{s\mathbf{n},i}$ and $\delta(z_k) = z_j - z_i$.

2.2.1 Stability criteria

The explicitly updated conserved variables are defined through the fluxes obtained within each cell, so, the computational time step has to be chosen small enough for ensuring a stability region. Traditionally, the numerical stability has been controlled through a dimensionless parameter, CFL,

$$\Delta t = CFL \, \frac{\min(\chi)}{\max|\widetilde{\lambda}^m|} \qquad\qquad CFL \leq 0.5 \tag{15}$$

where χ is a relevant distance between neighboring cells [18] and $\widetilde{\lambda}^m$ are the hydrodynamic celerities. The stability criterion is revisited for including a discrete estimation of the bed celerity, $\widetilde{\lambda}_s$, as in [10],

$$\Delta t = CFL \, \frac{\min(\chi)}{\max|\widetilde{\lambda}^m, \widetilde{\lambda}_s|} \qquad\qquad CFL \leq 0.5 \tag{16}$$

With this numerical strategy, the stability condition takes into consideration the most restrictive numerical wave speed coming from the hydrodynamical and morphodynamical solvers. The resulting global time step is used for updating the whole set of conserved hydrodynamic and morphological variables in the system of equations.

3 NUMERICAL RESULTS

3.1 Landslide in a practical application

The test considered for the validation of the GPU implementation is based on a real topography of a catchment [5]. The Arnas catchment is located in the northern Spain Pyrenees, in the Borau valley, and has a surface of 2.84 km^2, ranging in altitude from around 900 to 1340 meters. Geologically, the catchment lies over Eocene flysch formations and has suffered land use and coverage changes in recent decades, generating a mixed vegetation cover which ranges from forest patches, dense and open shrubs, grassland cover and bare land. The assumption made by the authors is that the part of the bare terrain is composed by poorly sorted material and the idea

is to verify the maximum run out and potential consequences of a massive mobilization of that material. In Figure 1 is plotted the fixed bed rock and onto it the moving granular material. Due to the large dimensions of the catchment, the number of cells involved in the calculus is over 869000, making this type of phenomena a suitable candidate for GPU strategy.

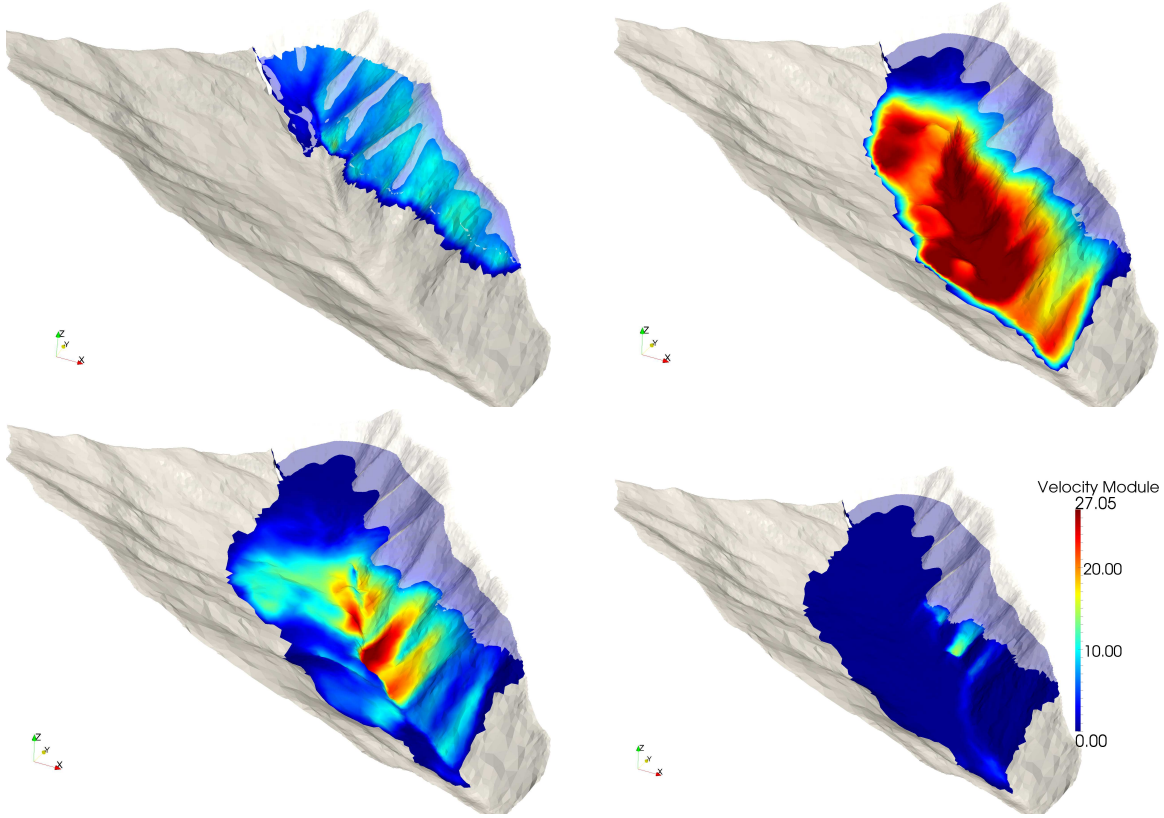


Figure 1: Evolution of the depth with the velocity (m/s) at time $t = 1s$. (top-left), $t = 2s$. (top-right), $t = 4s$. (bottom-left) and $t = 54s$. (bottom-right). The transparent element corresponds to the initial condition.

4 Computational load

The necessity of refined mesh for tests 1, 2 and 3 is justified by the important effect of the bed slope in the phenomena and then, an accurate representation of the topography is required. This means that the time-step length restricted by the CFL condition, is very low and then many time-steps are required to complete the simulation. Moreover, each time-step has a very high computational cost because of the number of cells. On the other hand, the last test case does not require such refinement because of the uncertainty of the data acquisition. Indeed, the mesh has been designed using the most accurate available LIDAR data for the topography, being the resolution of 5mx5m.

In order to analyze the computational load of the numerical engine as well as the gain obtained using both Single-Core and Multi-Core approaches, a sequential version using an *Intel Core i7 3770k@3.5 GHz* is compared against an OpenMP (4 Threads) parallel version and a GPU version without taking into account the improvements proposed on the paper (not optimized) and against an optimized version running on a *NVIDIA Tesla c2075* GPU. All the implementations have been tested in the previous four cases and the computational cost as well as the speed-ups of the parallel versions are highlighted in Table 1 and in Figure 2.

Taking into account both factors, small time-step size and high number of cells, the cost of the simulations for the sequential version of the numerical engine for tests cases 1 2 and 3 is three orders of magnitude larger than the real time. On the other hand, the parallel Multi-Core version, with 4 cores of the same CPU, can accelerate the computation between 2.25 and 2.65 times. The GPU improves the simulation cost between 34.88 and 49.40 times compared with the sequential version. Moreover, if the mesh is reordered, the computational cost is smaller and a speed-up between 49.96 and 59.85 is obtained.

The main reason for discrepancies on the speed-up of cases 1, 2 and 3 against test case 4 is that the number of calculations in the latter is higher compared with the number of calculations for the other cases, i.e. there are more cells that satisfy $h > 0$ in test 4. Therefore, the more cells are involved in the calculus, the larger speed-ups can be obtained with the GPU. This is also described in [13].



Figure 2: Computational time (upper) using logarithmic scale and speed-up (lower) for the compared implementations

		Seq.	4 Cores		GPU Std.		GPU Opt	
Case	n_{cells}	$t(s)$	$t(s)$	s_{up}	$t(s)$	s_{up}	$t(s)$	s_{up}
Test 1	319354	191.15	81.63	**2.34**	5.48	**34.88**	3.83	**49.96**
Test 2	458684	2274.13	912.94	**2.49**	64.03	**35.52**	45.59	**49.89**
Test 3	670940	5217.70	2319.48	**2.25**	140.68	**37.09**	104.12	**50.11**
Test 4	869149	22929.15	8657.59	**2.65**	464.16	**49.40**	383.13	**59.85**

Table 1: Detail of computational cost and speed-up for the compared implementations using the four cases.

4.1 TOUS DAM BREAK

To test large spatial domains that require a high number of cells for flood warning/hazard prediction the dam failure of Tous dam is proposed [1]. Tous dam is the last flood control structure of the Júcar River basin in the central part of the Mediterranean coast of Spain. During the 20th and the 21st October 1982 a particular meteorological condition led to extremely heavy rainfall. As a result the Júcar River basin suffered flooding all along and the Tous Dam failed with devastating effects downstream. The first affected town was Sumacárcel, about 5 km downstream of Tous Dam, lying at the toe of a hill on the right bank of Júcar river [1]. The terrain is moderately mountainous and most of the buildings lie on a slope that partially protected them from the flood. The ancient part of the village, however, is located closer to the river course and was completely flooded, with high water marks reaching between 6 m and 7 m.

The DTM model used in this work was generated by CEDEX in 1998 [1]. From this information a numerical mesh with $3 \cdot 10^5$ cells has been defined. This computational domain covers most of the original DTM, starting just after the dam location and finishing approximately 1 km downstream of Sumacárcel. The mesh has been refined in the dam area and in the village area (Fig. 3) for providing an adequate resolution for the hydraulic structures and the buildings. It is stressed that the decrease in the cell size leads to an increment in the simulation time since the stability criterion is more restrictive. This is also described in [11].

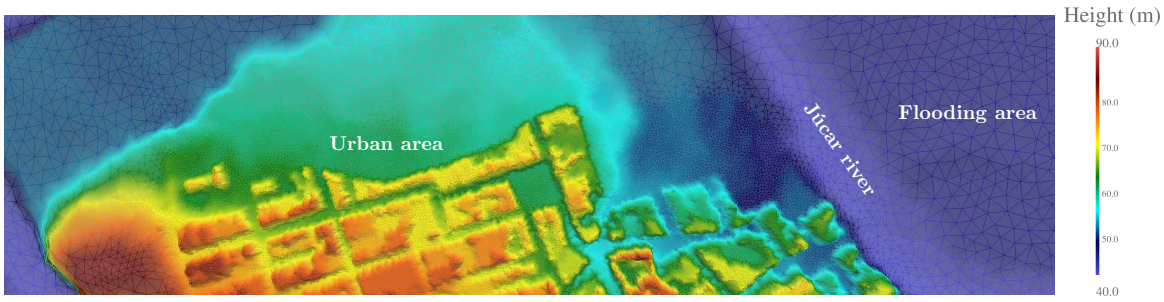


Figure 3: Detail of the simulation mesh at the village area nearby

The cause of the dam break was overtopping/dam-breaching, due to intense rainfall, and its later erosion and collapse. The height of the dam crest was 98.5 m and before reaching this level the discharge facilities of the dam were opened in order to evacuate the huge amount of incoming water. To reproduce this situation, the authors have considered the water elevation records together with the reservoir rating curves for simulating the spillway procedure, i.e. a water discharge of 3568 m^3s^{-1} is considered for obtaining the initial condition. Once the crest level is reached, a dam breach starts and it causes the erosion and collapse process. Hence, an outflow discharge emerging from the dam creates the traveling wave which is the responsible for the flooding event, i.e. it is the key information for the prediction of this event. In previous studies [1], since the morphodynamic change of the dam was not modeled, a tuning synthetic discharge, based on several assumptions, was estimated. Finally, at the outlet boundary, downstream of the domain, the flow was let to exit freely without imposing any conditions, as no information was provided.

On the other hand, following [1], a Manning coefficient of 0.030 $sm^{-1/3}$ has been set for the whole river bed reach and, additionally, an increased roughness coefficient of 0.1 $sm^{-1/3}$ has been defined in two zones close to the village with dense orange trees. The mean sediment

diameter involved in the erosion process has been set to 0.02 m. As the ground in the town area was fully paved with concrete the flood did not erode it. The real time simulated has been 11.1 hours from the beginning of the dam overtopping.

In Fig. 4 the breach evolution of the dam is plotted at several times. The flow overtopping causes the inception of the erosion at the front edge of the dam crest. As the breach increases in size the flow is accelerated and a severe erosion occurs. Consequently, the water discharge in the breach also augments. The earthfill material is grabbed by the flow and it is settled downstream the dam creating a sediment tongue which migrates towards the riverbed. At the end of the event the morphology of the dam area has changed completely and an important fraction of the dam has been completely removed, which is in agreement with the photos taken after the event and provided in [1].

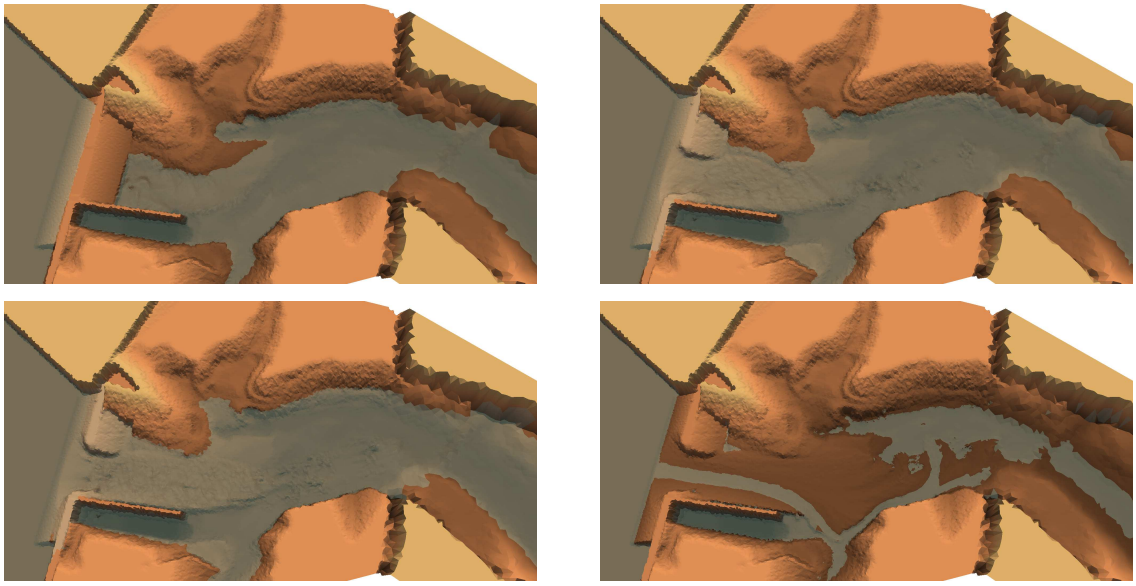


Figure 4: Initial condition (Top-Left) and evolution of the erosion process at t=1.3 hours (Top-Right), t=2.7 hours (Bottom-left) and at final stage (t=11.1 hours) (Bottom-right)

The evolution of the computed flooding can be seen in full plan view in Fig. 5 at times t =0, 1.3, 2.7 and 11.1 hours considering the time t=0 when the water surface level inside the reservoir has reached the dam crest and the overtopping is about to start. The flow advances towards the village filling the riverbed capacity and, consequently, inundating the floodplain areas nearby.

Thanks to the work described in [1], there are field data for the estimation of: (i) the maximum and minimum levels reached by the flood wave or (ii) a unique level for the water surface at different locations within the town, for evaluating the quality of the simulations. This estimation was performed considering a range of values within which it was completely ensured that the water reached that level. The location of the gauging points is shown in Fig. 6. Figure 7 displays the water depth recorded at several locations in Sumacárcel village together with the numerical predictions. There is a good agreement between the field data and the estimated depth, since most of the probes reach the range, between the maximum and minimum, estimated during the event. This agreement is attributed to the adequate simulation of the erosion process at the Tous dam.

It is also important to highlight that, by coupling the hydrodynamic and the breach erosion phenomena, less assumptions are required. This may be relevant in practical applications but is

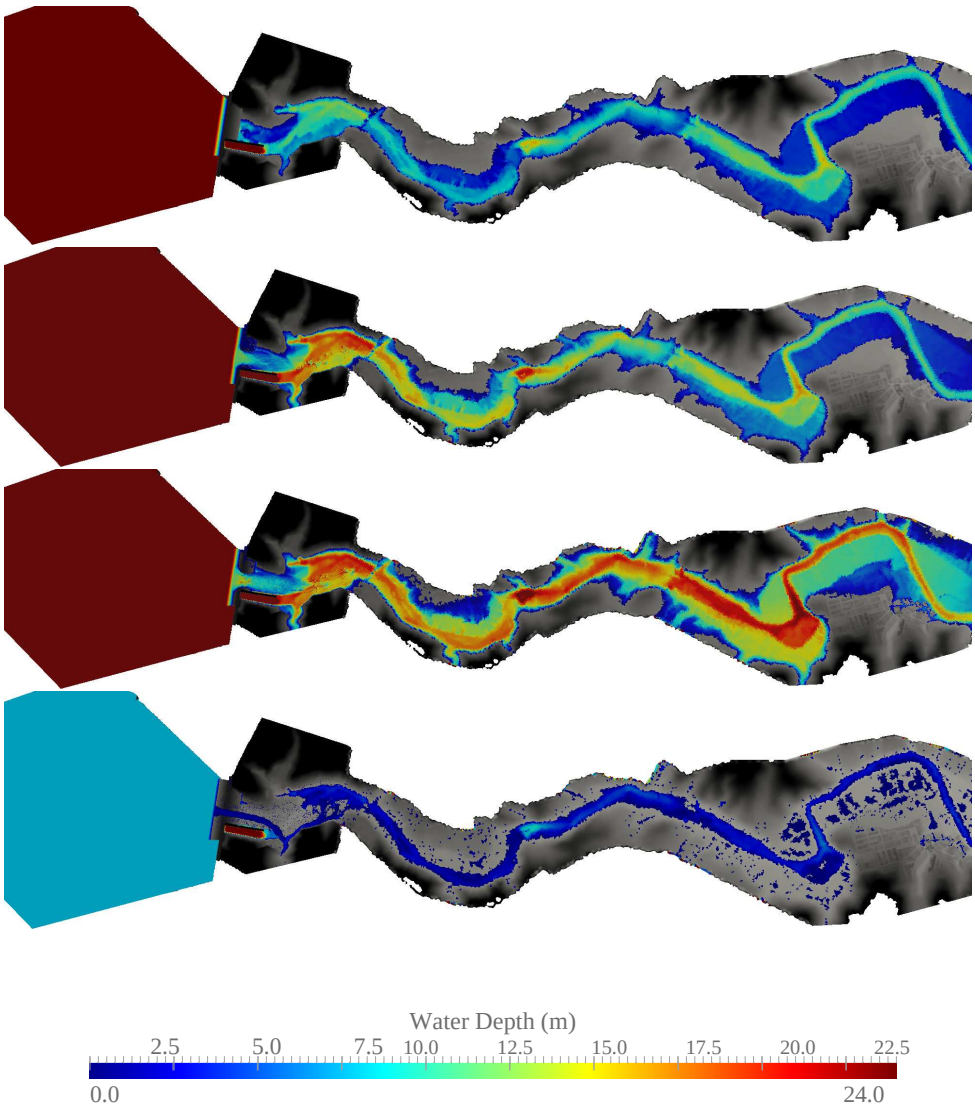


Figure 5: Water depth evolution along the valley at times t=0, 1.3, 2.7, 11.1 hours from top to bottom

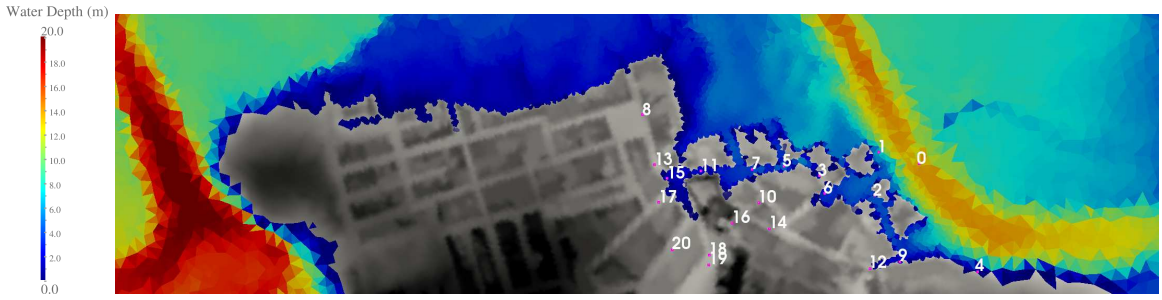


Figure 6: Detail of the location of the gauging points

costly in computational terms. For instance, in [1] a synthetic hydrograph based on a detailed analysis of how the dam failed was proposed. However, thanks to the GPU capabilities it is possible to couple the hydrodynamics and the dam erosion for obtaining directly the hydrograph which is the responsible for the later flooding event. In Fig. 8 both hydrographs, the synthetic

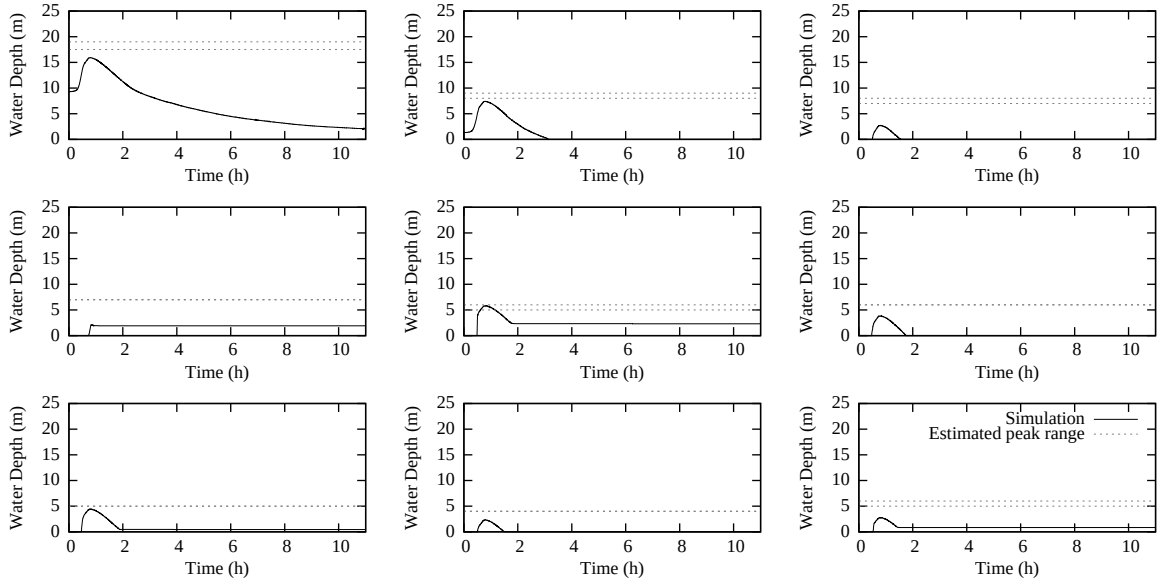


Figure 7: Water depth numerical predictions at several locations in Sumacárcel village and estimated range provided in [1] for gauges 1, 2, 3, 4, 6, 7, 11, 12 and 15 (from top to bottom and from left to right), see Fig. 6

and the computed one in the dam-breach, are plotted. It is remarkable that the peak discharge observed by means of the simulation, Q_{peak}= 14568.09 m^3s^{-1}, is very close to the peak discharge estimated in [1], where Q_{peak}= 15000 m^3s^{-1}. Conversely, the computed discharge is less sustained in time. This difference is probably because the inlet tributaries of the reservoir have been neglected. Since this effect has not been taken into account, in [1] there is not a fair estimation of the magnitude of these inlet tributaries, only the water contained in the reservoir at the beginning of the event is allowed to outflow in the simulation.

The evolution of the dam-breach is also plotted in Fig. 8 using the same cross section used to evaluate the discharge. It can be observed that most of the process has occurred within the first 1500 s, i.e. during the peak discharge. After t=1500 s changes in bed morphology are less violent.

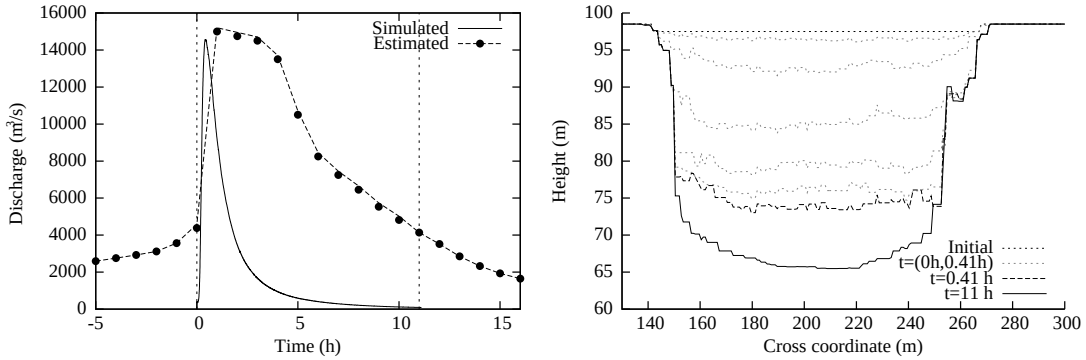


Figure 8: (Left) Comparison of the hydrograph generated due to the dam failure using the presented implementation against the hydrograph estimated in [1]. Simulated window is highlighted considering the time interval between $t = 0$ and $t = 11$ hours. (Right) Evolution of the dam-breach from $t = 0$ to $t = 0.41$ hours (peak discharge) each 0.07 hour and $t = 11.0$ hours (final state)

The execution time is summarized in Table 2. In this case, only the parallel CPU version has been benchmarked due to the huge execution time required for the single-core CPU version.

Table 2: Detail of execution time and speed-up for the compared implementations

4 Cores	GPU	
t	t	S_{up}
207 h 7 min	8h 7 min	25.25

The GPU reduces the simulation effort 25 times compared with the 4-Core version allowing an efficient simulation and accurate prediction. It is important to take into account that, in this case, the improvement has been increased compared against the previous cases where the GPU accelerates the computation of the OpenMP solution in a 20 factor. This effect has been previously reported in hydrodynamic simulation in [12] and it is due to the large number of elements included in the calculation. Thanks to the GPU capabilities it has been affordable to locally refine the mesh in the breach area and provide an adequate design for the initial breach which provides the dam-breaching discharge. Therefore, several possibilities can be addressed in the same day which is a noticeable advance when comparing with the computational effort based on CPU.

5 ACKNOWLEDGEMENTS

This work was partially supported and funded by the Spanish Ministry of Science and Technology under research projects CGL2015-66114-R, and by Diputación General de Aragón, DGA, through FEDER funds.

6 CONCLUSIONS

- The new opportunities given by the GPU implementation have been described for the analysis of several situations where the morphodynamic effects are relevant.
- For this purpose, the shallow water equations in combination with the Exner equation have been discretized in Finite Volumes and the numerical schemes implemented to run on a GPU card.
- This model allows to properly represent the propagation of bed and surface waves over realistic bathymetries in affordable computation time even when considering large domains and retaining a high level of accuracy.
- Unstructured meshes have been considered, since this grid topology is the only one which avoid misleading preferential flow directions.
- The computational times have been compared with those obtained when considering Single-Core and Multi-Core processors. The GPU implementation and specially, the application of cell and wall ordering algorithms, have driven to obtain noticeable improvements in the speed-up of the test cases. The GPU implementation provides a peak speedup of 50.
- This saving of time allows to address large-number-of-cells, large-time and large-space scenarios, strengthening preventive measures and enhancing response capacities.
- This opens the possibility of facing the sediment transport analysis in a particular location for several years or the geomorphological changes in domains of a regional-size.

REFERENCES

[1] F. Alcrudo and J. Mulet. Description of the Tous dam break case study (Spain). *Journal of Hydraulic Research*, 45(Extra Issue):45–57,2007.

[2] S. Borkar. Thousand core chips: A technology perspective. In *Proceedings of the 44th Annual Design Automation Conference*, DAC '07, pages 746–749, New York, NY, USA, 2007. ACM.

[3] F. Bouchut, E. D. Fernández-Nieto, A. Mangeney, and P. Y. Lagrée. On new erosion models of Savage-Hutter type for avalanches. *Acta Mechanica*, 199(1-4):181–208, 2008.

[4] R.G. Dreslinski, M. Wieckowski, D. Blaauw, D Sylvester, and T. Mudge. Near-threshold computing: Reclaiming moore's law through energy efficient integrated circuits. *Proceedings of the IEEE*, 98(2):253–266, Feb 2010.

[5] J. García-Ruiz, J. Arnaez, S. Beguería, M. Seeger, C. Marti-Bono, D. Regues, N. Lana-Renault, and S. White. Runoff generation in an intensively disturbed, abandoned farmland catchment, Central Spanish Pyrenees. *Catena*, 59:79–92, 2005.

[6] Peter N Glaskowsky. Nvidia's fermi: The first complete GPU computing architecture. *A white paper prepared under contract with NVIDIA Corporation*, (September):126, 2009.

[7] J.M.N.T. Gray, K. Wieland, and K. Hutter. Gravity-driven free surface flow of granular avalanches over complex basal topography. *Proc. Royal Soc. London*, Ser. A, 455:1841, 1999.

[8] R.M. Iverson and R.P. Denlinger. Flow of variably fluidized granular masses across three-dimensional terrain. A Coulomb mixture theory. *Journal of Geophysical Research*, 106(B1):537–552, 2001.

[9] C. Juez, J. Murillo, and P. García-Navarro. 2D simulation of granular flow over irregular steep slopes using global and local coordinates. *Journal of Computational Physics*, 255:166–204, 2013.

[10] C. Juez, J. Murillo and P. García-Navarro. A 2D weakly-coupled and efficient numerical model for transient shallow flow and movable bed. *Advances in Water Resources*, 71:93–109,2014.

[11] C. Juez, A. Lacasta, J. Murillo and P. García-Navarro. An efficient GPU implementation for a faster simulation of unsteady bed-load transport. *Journal of Hydraulic Research*, 1–14,2016.

[12] A. Lacasta, M. Morales-Hernández, J. Murillo, and P. García-Navarro. An optimized gpu implementation of a 2d free surface simulation model on unstructured meshes. *Advances in Engineering Software*, 78:1–15, 2014.

[13] A. Lacasta, C. Juez, J. Murillo, and P. García-Navarro. An efficient solution for hazardous geophysical flows simulation using GPUs. *Computers & Geosciences* 78:63–72, 2015.

[14] E. Lajeunesse, A. Mangeney-Castelnau, and J. P. Villote. Spreading of a granular mass on a horizontal plane. *Physics of Fluids*, 16:2371–2381, 2004.

[15] A. Mangeney, O. Roche, O. Hungr, N. Mangold, G. Faccanoni, and A. Lucas. Erosion and mobility in granular collapse over sloping beds. *Journal of Geophysical Research*, 115:F03040, 2010.

[16] G.E. Moore. No exponential is forever: but "forever" can be delayed! [semiconductor industry]. In *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, 1:20–23, 2003.

[17] L. Moretti, A. Mangeney, Y. Capdeville, E. Stutzmann, C. Huggel, D. Schneider, and F. Bouchut. Numerical modeling of the Mount Steller landslide flow history and of the generated long period seismic waves. *Geophysical Research Letters*, 39:L16402, 2012.

[18] J. Murillo and P. García-Navarro. Weak solutions for partial differential equations with source terms: Application to the shallow water equations. *Journal of Computational Physics*, 229:4327–4368, 2010.

[19] J. Murillo and P. García-Navarro. Wave Riemann description of friction terms in unsteady shallow flows: Application to water and mud/debris floods. *Journal of Computational Physics*, 231:1963–2001, 2012.

[20] NVIDIA. *NVIDIA CUDA C Programming Guide*, Jun 2011.

[21] M. Pirulli, M. Bristeau, A. Mangeney-Castelnau, and C. Scavia. The effect of the earth pressure coefficients on the runout of granular material. *Environmental Modelling and Software*, 22:1437–1454, 2007.

[22] M. Pirulli and A. Mangeney. Results of Back-Analysis of the Propagation of Rock Avalanches as a Function of the Assumed Rheology. *Rock Mechanics and Rock Engineering*, 41:59–84, 2008.

[23] O. Pouliquen and Y. Forterre. Friction law for dense granular flows: application to the motion of a mass down a rough inclined plane. *Journal of Fluid Mechanics*, 453:133–151, 2002.

[24] S. Sharma and K. Gupta. Parallel performance of numerical algorithms on multi-core system using openmp. In Natarajan Meghanathan, Dhinaharan Nagamalai, and Nabendu Chaki, editors, *Advances in Computing and Information Technology*, volume 177 of *Advances in Intelligent Systems and Computing*, pages 279–288. Springer Berlin Heidelberg, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ADDRESSING THE CHALLENGES OF IMPLEMENTATION OF HIGH-ORDER FINITE-VOLUME SCHEMES FOR ATMOSPHERIC DYNAMICS ON UNSTRUCTURED MESHES

Panagiotis Tsoutsanis[1], **Dimitris Drikakis**[2]

[1]Fluid Mechanics and Computational Sciences,Cranfield University
Cranfield MK43 0AL, United Kingdom
e-mail: panagiotis.tsoutsanis@cranfield.ac.uk

[2] University of Strathclyde
Glasgow G1 1XW, United Kingdom
e-mail: dimitris.drikakis@strath.ac.uk

Keywords: WENO, Unstructured Meshes, Non-Hydrostatic, Rising Bubble

Abstract. *The solution of the non-hydrostatic compressible Euler equations using Weighted Essentially Non-Oscillatory (WENO) schemes in two and three-dimensional unstructured meshes, is presented. Their key characteristics are their simplicity; accuracy; robustness; non-oscillatory properties; versatility in handling any type of grid topology; computational and parallel efficiency. Their defining characteristic is a non-linear combination of a series of high-order reconstruction polynomials arising from a series of reconstruction stencils. In the present study an explicit TVD Runge-Kutta 3^{rd}-order method is employed due to its lower computational resources requirement compared to implicit type time advancement methods. The WENO schemes (up to 5^{th}-order) are applied to the two dimensional and three dimensional test cases: a 2D rising thermal bubble. The scalability and efficiency of the schemes is also investigated.*

1 Introduction

Understanding the atmospheric processes both qualitatively and quantitatively on regional and global scales is of great value for scientists and policy decision makers. Tools utilised by the scientific community to gain insight into the complex atmosphere-climate system, include the numerical weather prediction (NWP) and climate models. The main objective is to accurately simulate weather and climate variations, trends, patterns etc. All the models utilised for this purpose have a distinctive component in common, the *dynamical core*, which solves the fluid dynamics equations. In other words, it is the model component that drives atmospheric motions, hence with a profound impact on the accuracy of the predictions. The leap occurring in computing processing power the last decade, provides a motivation to resolve atmospheric processes on finer scales and higher resolution than happened before. There is currently a trend of either enhancing the majority of the dynamical cores of established models or develop new dynamical cores that utilise highly sophisticated algorithms or set of equations that are tailored for finer resolution simulations in larger computing facilities. A number of initiatives have been taken to address this bottleneck and develop the next generation of dynamical cores. Such initiatives include the UK Met Office and European Centre for Medium-Range Weather Forecasts Even Newer Dynamics for General Atmospheric Modelling of the Environment (ENDGame) [51]; the National Center for Atmospheric Research High-Order Methods Modelling Environment [34]; the National Oceanic and Atmospheric Administration AM3 model [35] and the Nonhydrostatic Icosahedral Model (NIM) [36]; the Frontier Research Center for Global Change Nonhydrostatic Icosahedral Atmospheric Model, (NICAM) [41]; the Non-hydrostatic Unified Model of the Atmosphere (NUMA) [17]; and the Max Planck Institute for Meteorology and Deutscher Wetterdienst ICOsahedral Non-hydrostatic general circulation models, (ICON) [31].

This paper focuses on the the numerics of the dynamical core aspect of atmospheric models employed for weather and climate predictions for well established test cases. Since high performance computing systems enable the analysis on finer scales and higher resolution, the topography should be efficiently and accurately represented, a property typical for unstructured grids. In the past, finite volume numerical methods have been introduced in the structured grid context resulting in robust schemes that enjoyed a wide recognition in various disciplines, including the weather and climate prediction community, e.g. [27, 28, 45].

The first generation numerical methods for unstructured grids exhibited lower accuracy and were computationally more demanding than structured grids. However, the numerical methods for unstructured grids have matured and numerous elegant approaches [10, 11, 18, 21, 29, 37, 39, 47, 49, 50, 56–58] and algorithms have been developed in the finite volume framework for a wide range of applications for Computational Fluid Dynamics(CFD). Other state-of-the-art approaches are also available, such as the Discontinuous Galerkin (DG) [6, 7, 11, 52, 57, 64], and Spectral Finite-Volume (SFV) methods [5, 53–55, 58, 63] that have been successfully applied for CFD applications. For the finite volume framework the first class of high-resolution methods developed for unstructured grids included the ENO type schemes [1, 46], followed by the WENO type schemes [13, 23, 42, 43]. In the WENO case, the high-order accuracy was achieved by non-linearly combining a series of high-order reconstruction polynomials arising from a series of reconstruction stencils. Recently, a class of WENO type methods [47, 49]

has been successfully extended to hybrid unstructured meshes with various geometrical shapes such as tetrahedrals, hexahedrals, prisms, and pyramids. The schemes can achieve the very high order of spatial accuracy across interfaces between cells of different types, and at the same time essentially non-oscillatory profiles are produced for discontinuous solutions. This gives greater flexibility to handle complex geometrical shapes in an efficient and accurate manner.

For the atmospheric dynamics there is also an overwhelming number of recent state-of-the-art approaches that utilise high-order schemes for either structured grid or unstructured dynamical cores [4, 7, 14, 15, 24, 30, 32, 44, 59–61]. Regarding the high-order schemes (higher than 3rd order of spatial accuracy) for unstructured dynamical cores all of the approaches are based on the Discontinuous Galerkin framework and the Lagrange-Galerkin Spectral Element Method[7, 15, 26, 32, 60]. The fact that WENO schemes have been successfully applied to a range of smooth and discontinuous flows [10, 11, 50], including the Reynolds Averaged Navier Stokes (RANS) equations [2, 3, 8], motivates the application of these schemes in the context of non-hydrostatic Euler equations.

The aim of this paper is to assess the performance and suitability of WENO-class higher-order finite volume schemes in conjunction with unstructured meshes for well established test cases for non-hydrostatic Euler equations. The schemes can be used in any type of conforming grids.

The paper is organised as follows. In Section 2 the governing equations of a dry non-hydrostatic atmosphere are presented. Section 3 outlines the spatial discretisation techniques WENO reconstruction, numerical fluxes, source terms and temporal discretisation. The results obtained with various schemes for different test cases, including discussion on the parallel performance, are presented in Section 4 . Section 5 summarises the conclusions drawn from the present study.

2 Governing Equations

The compressible inviscid Euler equations employed by the CFD community where mass, momentum and total energy are the conserved quantities, are rarely used for atmospheric studies, with a few notable exceptions [16]. The primary reason for not being that common in atmospheric studies is the additional computational step required to compute potential temperature from total energy. However, the benefits of using these set of equations are their formally conservative nature as reported in [16]. The compressible non-hydrostatic Euler equations without moisture effects are considered in the following form:

$$\frac{\partial}{\partial t}\mathbf{U}+\frac{\partial}{\partial x}\mathbf{F}(\mathbf{U})+\frac{\partial}{\partial y}\mathbf{G}(\mathbf{U})+\frac{\partial}{\partial z}\mathbf{H}(\mathbf{U})=\mathbf{S}\,, \tag{1}$$

where $\mathbf{U}$ is the vector of the conserved variables, $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$ are the flux vectors in x, y and z directions of Cartesian coordinates and $\mathbf{S}$ is a gravity source function given by

$$\mathbf{U} = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ \rho w \\ E \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ \rho uw \\ u\left(E + p\right) \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} \rho w \\ \rho wu \\ \rho wv \\ \rho w^2 + p \\ w\left(E + p\right) \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \rho v \\ \rho vu \\ \rho v^2 + p \\ \rho vw \\ v\left(E + p\right) \end{bmatrix}.$$

$$\mathbf{S} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\rho g \\ 0 \end{bmatrix}.$$

where ρ is density; u, v, w are the velocity components in the x, y and z directions of Cartesian coordinates, respectively; p is pressure; $E = \rho c_v T + (1/2)\rho(u^2 + v^2 + w^2) + \rho g y$ is the total energy per unit mass; γ is the ratio of specific heats; g the gravitational constant; gy is the geopotential height; T is the temperature; $R = c_p - c_v$ is the gas constant, where c_p is the specific heat at constant pressure and c_v is the specific heat at constant volume. The ideal gas law is used with $\gamma = 1.4$ throughout this work, with the pressure p given by

$$p = p_s \left(\frac{\rho R \theta}{p_s} \right)^{\gamma}, \tag{2}$$

where p_s is the atmospheric pressure at sea level, and θ is the potential temperature

$$\theta = \frac{T}{\pi}, \tag{3}$$

with π being the Exner pressure provided by the following expression

$$\pi = \left(\frac{p}{p_s} \right)^{\frac{R}{c_p}}. \tag{4}$$

The following equations relate the Exner pressure π to the density ρ, and the potential temperature θ to the total energy per unit mass E:

$$\rho = \frac{p_s}{R\theta} \pi^{\frac{c_v}{R}}, \tag{5}$$

and

$$E = \rho c_v \theta \pi + (1/2)\rho(u^2 + v^2 + w^2) + \rho g y. \tag{6}$$

The Exner pressure and potential temperature can be split in their mean and perturbed parts as follows

$$\theta\left(x,y,z,t\right)=\bar{\theta}\left(y,t\right)+\theta'\left(x,y,z,t\right), \tag{7}$$

and

$$\pi\left(x,y,z,t\right)=\bar{\pi}\left(y,t\right)+\pi'\left(x,y,z,t\right), \tag{8}$$

where the hydrostatic balance of their mean values is given by

$$c_p\bar{\theta}\frac{d\bar{\pi}}{dy}=-g. \tag{9}$$

Integrating in space over a mesh element V_i , the following semi-discrete finite-volume method is obtained

$$\frac{d}{dt}\mathbf{U}_i+\frac{1}{|V_i|}\oint_{\partial V_i}\mathbf{F}_n dA=\mathbf{S_i}\ , \tag{10}$$

and using the rotational invariance property of the Euler equations [48]

$$\mathbf{F}_n\left(\mathbf{U}\right)=\mathbf{F}\left(\mathbf{U}\right)n_x+\mathbf{G}\left(\mathbf{U}\right)n_y+\mathbf{H}\left(\mathbf{U}\right)n_z=\mathbf{T^{-1}F}\left(\mathbf{TU}\right)\ , \tag{11}$$

where $\mathbf{n}=(n_x,n_y,n_z)$ is the outward unit normal vector; $\mathbf{U}_i(t)$ are the cell averages of the solution at time t; $\mathbf{F}_n$ is the projection of the flux tensor on the normal direction; $\mathbf{T}$ is the rotation matrix; and $\mathbf{T^{-1}}$ its inverse [48] ; and the source term $\mathbf{S_i}$ is given by

$$\mathbf{S_i}=\frac{1}{|V_i|}\int_{V_i}S(x,y,z,t)dxdydz, \tag{12}$$

Assuming that the element's surface consists of L faces, with the spatial index i being omitted for simplicity,

$$\partial V_i=\sum_j^L A_j,$$

($\mathbf{n}_j$ denoting the outward unit vector for face A_j), then the integral over the element boundary ∂V_i can be split into the sum of integrals over each face, resulting in the following expression:

$$\frac{d}{dt}\mathbf{U}_i=\mathbf{R}_i, \tag{13}$$

where

$$\mathbf{R}_i=-\frac{1}{|V_i|}\sum_{j=1}^{L}\int_{A_j}\mathbf{F}_{n,j}dA+\mathbf{S_i}=-\frac{1}{|V_i|}\sum_{j=1}^{L}\mathbf{K}_{ij}+\mathbf{S_i}.$$

The numerical flux $\mathbf{K}_{ij}$ corresponding to face j of the cell V_i is the surface integral of the projection of the tensor of fluxes onto $\mathbf{n}_j$. In a numerical method the exact integral expression for the numerical flux $\mathbf{K}_{ij}$ for the face j of a cell V_i is approximated by a suitable Gaussian

numerical quadrature where the wall bounded elements quadrature points are depicted with different color corresponding to different weights

$$\mathbf{K}_{ij} = \int_{A_j} \mathbf{F}_{n,j} dA = \sum_{\beta} \mathbf{F}_{n,j} \left(\mathbf{U}(\mathbf{x}_\beta, t)\right) \omega_\beta |A_j|, \tag{14}$$

where the subscript β corresponds to different Gaussian integration points $\mathbf{x}_\beta$ and weights ω_β over the face A_j .

3 Numerical Framework

In this section the numerical framework for the solution of the compressible non-hydrostatic Euler equations is presented with respect to spatial and temporal discretisation.

3.1 Spatial discretisation

The spatial discretisation is based on the approach of [47, 49], which is suitable for unstructured meshes with various types of element shapes in 2D and 3D, and it has been previously used successfully for laminar, transitional and turbulent flows [50]. Therefore, only the key characteristics of this approach are going to be described in this paper. The first step in the approximation of the governing equations concerns the spatial discretisation of the domain $\Omega \subset R^3$ into E number of conforming elements V_i, with the index i ranging from 1 to E number of elements. The elements can be any combination of hexahedrals, pyramids, prisms, and tetrahedrals as shown in Figure 1 in 3D, and triangles or quadrilaterals in 2D. It must be noted that the present WENO schemes are not limited by conforming meshes since they can be used with non-conforming meshes which brings additional benefits such as adaptive mesh refinement [12].

The combination of all elements in the spatial domain is given by

$$\Omega = \bigcup_{i=1}^{E} V_i. \tag{15}$$

In the context of the finite volume framework that is employed in the present study, the data is represented by cell averages of conserved variable $u(x, y, z)$ in each element V_i

$$\overline{u}_i = \frac{1}{|V_i|} \int_{V_i} u(x, y, z)\, dV, \tag{16}$$

where $|V_i|$is the volume of the element. To achieve high-order, accurate spatial discretisation, high-order accurate point-wise values of the solution need to be recovered from the cell averages. The reconstruction problem can thus be reformulated as follows: For a target cell V_0 we build a high-order polynomial $p_i(x, y, z)$ that has the same cell average as u on the target cell

$$\overline{u}_i = \frac{1}{|V_i|} \int_{V_i} u(x, y, z)\, dV = \frac{1}{|V_i|} \int_{V_i} p_i(x, y, z)\, dV. \tag{17}$$

The point-wise values of the conserved variable u in each cell are approximated by a polynomial of a desired order of accuracy r. The polynomial uses the cell averages of $u(x, y, z)$ on the target cell V_i as well as averages $\bar{u}_m$ from neighbouring cells, V_m. The key

3.1.1 Reconstruction algorithm

The main steps for the reconstruction process are the following:

1. For all the non tetrahedral cells the cell is decomposed into tetrahedral elements. It must be stressed that the decomposition to tetrahedral cells, is done only for the transformation purposes to the reference space in order to remove any scaling effect of the least square system to be solved.

2. Choose one of the resulting decomposed elements.

3. Transform the chosen decomposed element from the physical space described by the Cartesian coordinates x, y, z into a reference space described by ξ, η, ζ.

4. Based on the Jacobian matrix of the transformation of the chosen decomposed element, map the coordinates of the entire element into the reference space described by coordinates ξ, η, ζ.

5. Based on the same Jacobian, all the elements in the stencil are transformed to the reference space and their volumes, and barycentres positions are recomputed in the new reference space.

Let $\mathbf{v}_{ij}$, $j = 1, 2, \ldots J_i$ be the vertices of the considered (general) element, which can be either tetrahedral, hexahedral, prismatic or pyramidal. Let also $\mathbf{w}_1 = (x_1, y_1, z_1)$, $\mathbf{w}_2 = (x_2, y_2, z_2)$, $\mathbf{w}_3 = (x_3, y_3, z_3)$, $\mathbf{w}_4 = (x_4, y_4, z_4)$ be the four vertices of one of the tetrahedrals of this element. Obviously, these vertices are between the $\mathbf{v}_{ij}$ ones. The transformation from the Cartesian coordinates $x,\ y,\ z$ into a reference space $\xi,\ \eta,\ \zeta$ is given by the following equations

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + J \cdot \begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix}, \tag{18}$$

with the Jacobian matrix given by

$$J = \begin{bmatrix} x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ y_2 - y_1 & y_3 - y_1 & y_4 - y_1 \\ z_2 - z_1 & z_3 - z_1 & z_4 - z_1 \end{bmatrix}. \tag{19}$$

Using an inverse mapping the element V_i can be transformed to the element V_i' in the reference co-ordinate system

$$\mathbf{v}'_{ij} = J^{-1} \cdot (\mathbf{v}_{ij} - \mathbf{w}_1), \quad j = 1, 2, \ldots J_i. \tag{20}$$

The spatial average of $u(x,y,z)$ does not change during transformation

$$\bar{u}_i = \frac{1}{|V_i|}\int_{V_i} u(x,y,z)\,dV \equiv \frac{1}{|V_i'|}\int_{V_i'} u(\xi,\eta,\zeta)\,d\xi d\eta d\zeta. \tag{21}$$

For performing the reconstruction on the target element V_i, we form the central reconstruction stencil $\mathcal{S}$ as described previously, consisting of $M+1$ elements, including the target element V_i

$$\mathcal{S} = \bigcup_{m=0}^{M} V_m, \tag{22}$$

where the index m refers to the local numbering of the elements in the stencil, with the element with index 0 being the considered cell i. The r^{th} order reconstruction polynomial at the transformed cell V_0' is sought as an expansion over local polynomial basis functions $\phi_k(\xi,\eta,\zeta)$:

$$p(\xi,\eta,\zeta) = \sum_{k=0}^{K} a_k\phi_k(\xi,\eta,\zeta) = \bar{u}_0 + \sum_{k=1}^{K} a_k\phi_k(\xi,\eta,\zeta), \tag{23}$$

where a_k are degrees of freedom and the upper index in the summation of expansion K is related to the order of the polynomial r by the expression $K = \frac{1}{6}(r+1)(r+2)(r+3)-1$ for 3D and $K = \frac{1}{2}(r+1)(r+2)-1$ for 2D . The conservation condition (16) imposes an important constraint on the basis functions: they must have zero mean value over the cell V_0'. On purely tetrahedral meshes hierarchical orthogonal reconstruction basis functions [10, 57], defined on the reference element, satisfy this requirement automatically. Since our general cells are not necessarily transformed onto a unit tetrahedron or cube, we need to construct basis functions ϕ_k in such a way that the condition is satisfied identically, irrespective of the degrees of freedom. The basis functions are defined as follows:

$$\phi_k(\xi,\eta,\zeta) \equiv \psi_k(\xi,\eta,\zeta) - \frac{1}{|V_0'|}\int_{V_0'} \psi_k\,d\xi d\eta d\zeta, \quad k = 1,2,\ldots \tag{24}$$

Any type of orthogonal polynomial basis function can be utilised; however, in the present study generic polynomial basis functions are used

$$\{\psi_k\} = \xi,\ \ \eta,\ \ \zeta,\ \ \xi^2,\ \ \eta^2,\ \ \zeta^2,\ \ \xi\cdot\eta,\ \ \xi\cdot\zeta,\ \ \zeta\cdot\eta,\ \ \xi\cdot\eta\cdot\zeta\ldots$$

To find the unknown degrees of freedom a_k for each cell V_m' from the stencil the cell average of the reconstruction polynomial $p(\xi,\eta,\zeta)$ must be equal to the cell average of the solution $\bar{u}_m$:

$$\int_{E_m'} p(\xi,\eta,\zeta)\,d\xi d\eta d\zeta = |V_m'|\bar{u}_0 + \sum_{k=1}^{K}\int_{V_m'} a_k\phi_k\,d\xi d\eta d\zeta = |V_m'|u_m, \quad m = 1,\ldots \tag{25}$$

Denoting the integrals of the basis function k over the cell m in the stencil, the vector of right-hand side by A_{mk} and $\mathbf{b}$ are given by

$$A_{mk} = \int_{V'_m} \phi_k \, d\xi d\eta d\zeta, \quad b_m = |V'_m|(\bar{u}_m - \bar{u}_0)$$

We can rewrite the equations for degrees of freedom a_k in a matrix form as

$$\sum_{k=1}^{K} A_{mk} a_k = b_m, \quad m = 1, 2, \dots M. \tag{26}$$

The three-dimensional integrals on the left-hand side of (26) are calculated using Gaussian quadratures of appropriate order. In general, to compute the degrees of freedom a_k we need at least K cells in the stencil, different from the target cell E_0 . However, the use of the minimum possible number of cells in the stencil $M \equiv K$ results in a scheme which may become unstable on general meshes [9, 21, 22, 49]. It is, therefore, recommended to use more cells in the stencil than the minimum required number. Although it is usually sufficient to use 50% more cells, for mixed-element meshes it is safer to increase the stencil further. We typically select $M = 2 \cdot K$.

Since the resulting system becomes over-determined, the least-square procedure is invoked to solve it. The least-square reconstruction is obtained by seeking the minimum of the following functional

$$\mathcal{F} = \sum_{m=1}^{M} w_m \cdot \left(\sum_{k=1}^{K} A_{mk} a_k - b_m \right)^2,$$

where the weights w_m are squared reciprocals of the distance between cells E'_0 and E_m. The advantage of the weighted least square reconstruction is that the influence of the data further from the considered E'_0 is reduced, although a central least square reconstruction is materialised with the weights ω_m being equal to unity. Minimisation of $\mathcal{F}$ gives a linear system for finding a_k :

$$\sum_{k=1}^{K} C_k a_k = \sum_{m=1}^{M} A_{mp} \, w_m b_m, \quad C_k = \left(\sum_{m=1}^{M} w_m A_{mk} A_{mp} \right), \quad p = 1, \dots K \tag{27}$$

A QR decomposition method is employed to solve this system of equations. The coefficients of the resulting linear symmetric matrix A are pre-computed and stored for each element during the pre-processing stage of the calculation, thus increasing the computational efficiency of the method. If the mesh was to be refined these would automatically require the need to re-compute and store this matrix. Having solved numerically the linear system, we can form the reconstruction polynomial (23).

3.1.2 WENO reconstruction

The main characteristic of the WENO reconstructions is the use of reconstruction polynomials from several different stencils and their combination in a non-linear manner. In WENO schemes the actual reconstructed value is a convex combination of reconstructed values from

stencils, with nonlinear (solution-adaptive) WENO weights [23, 42, 62]. These nonlinear weights are constructed from the linear (constant) weights by taking into account the smoothness of the solution in each of the reconstruction stencils. The resulting methods are uniformly high-order accurate, while maintaining non-oscillatory behaviour in regions with sharp gradients. The WENO reconstruction used in the present study is based on the implementation of [49], where the reader is referred to for further details.

3.1.3 Numerical fluxes and Source term

Since for each computational cell the point-wise values of the conserved vector $\mathbf{U}$ are replaced by high-order reconstruction polynomials the numerical flux for the face A_j of cell V_i is discontinuous at each Gaussian point β. The values $\mathbf{U}_\beta^-$ and $\mathbf{U}_\beta^+$ correspond to the reconstructed value of cell V_i and its adjacent neighbour $V_{i'}$. In upwind finite-volume methods the discontinuity is replaced at each Gaussian integration point by using a monotone function of left and right boundary extrapolated values so that (14) can be re-written as

$$\mathbf{K}_{ij} \approx \sum_\beta \mathbf{F}_{n,j}\left(\mathbf{U}_\beta^-, \mathbf{U}_\beta^+\right)\omega_\beta |A_j|. \tag{28}$$

The function $\tilde{\mathbf{F}}_{n,j}\left(\mathbf{U}_\beta^-, \mathbf{U}_\beta^+\right)$ is the Riemann solver [48]. Employing the rotational invariance property of the Euler equations [48], for each face A_j the normal projection of the flux tensor $\mathbf{F}_{n,j}$ is replaced by

$$\mathbf{F}_{n,j} = \mathbf{T^{-1}F}\left(\mathbf{T}_j\mathbf{U}\right), \tag{29}$$

where $\mathbf{T}_j$ is the rotation matrix for face j. Re-writing (28) for $\mathbf{K}_{ij}$ gives

$$\mathbf{K}_{ij} = \sum_\beta \mathbf{F}_{n,j}\left(\mathbf{U}_\beta^-, \mathbf{U}_\beta^+\right)\omega_\beta |A_j| = \sum_\beta \mathbf{T^{-1}F}\left(\hat{\mathbf{U}}_L, \hat{\mathbf{U}}_R\right)\omega_\beta |A_j|, \tag{30}$$

where $\hat{\mathbf{U}}_j$ is the rotated conserved variable and

$$\hat{\mathbf{U}}_L = \mathbf{T}_j\mathbf{U}_\beta^-, \quad \hat{\mathbf{U}}_R = \mathbf{T}_j\mathbf{U}_\beta^+.$$

The flux function for the Gaussian point β results in the one-dimensional Riemann initial value problem

$$\frac{\partial}{\partial t}\hat{\mathbf{U}} + \frac{\partial}{\partial s}\hat{\mathbf{F}} = \mathbf{0}, \quad \hat{\mathbf{F}} = \mathbf{F}(\hat{\mathbf{U}}), \quad \hat{\mathbf{U}}(s,0) = \begin{array}{ll} \hat{\mathbf{U}}_L, & s<0, \\ \hat{\mathbf{U}}_R, & s>0 \end{array} \tag{31}$$

In the present study the HLLC Riemann solver of Toro [48] is employed. The numerical source term $\mathbf{S_i}$ is approximated by using a Gaussian quadrature rule of the same order of accuracy as the order of the polynomial of the reconstruction

$$\mathbf{S_i} \approx \frac{1}{|V_i|}\sum_\beta \mathbf{S}(x_\beta, y_\beta, z_\beta, t)w_\beta, \tag{32}$$

where $x_\beta, y_\beta, z_\beta$ is the volumetric Gaussian quadrature points and w_β is the Gaussian weights of r-order of accuracy.

3.2 Temporal discretisation

Having constructed the numerical fluxes $\mathbf{F}_{n,j}$ as expressed in the semi-discrete conservative formulation, the next step involves the advancement of the solution in time. The explicit SSP Runge-Kutta 3^{rd}-order method [19] has been employed for the time integration

$$\left.\begin{aligned} \mathbf{U_i}^1 &= \mathbf{U_i}^n + \Delta t \cdot \mathbf{R}_i\left(\mathbf{U_i}^n\right) \\ \mathbf{U_i}^2 &= \tfrac{3}{4}\mathbf{U_i}^n + \tfrac{1}{4}\mathbf{U_i}^1 + \tfrac{\Delta t}{4} \cdot \mathbf{R}_i\left(\mathbf{U}^1\right) \\ \mathbf{U_i}^{n+1} &= \tfrac{1}{3}\mathbf{U_i}^n + \tfrac{2}{3}\mathbf{U_i}^2 + \tfrac{2\Delta t}{3} \cdot \mathbf{R}_i\left(\mathbf{U_i}^2\right) \end{aligned}\right\} \tag{33}$$

with $\mathbf{R}_i$ given by (13).

The time step Δt is computed as follows

$$\Delta t = K \ \min_i \frac{h_i}{S_i \cdot V_i}, \tag{34}$$

where h_i is the radius of the inscribed sphere of each cell i and V_i its corresponding volume, $K \leq 1/3$ is the CFL number for unsplit finite-volume schemes [48] , and S_i is the maximum propagation speed in each cell i given by

$$S_i = sp_x \cdot n_x + sp_y \cdot n_y + sp_z \cdot n_z, \tag{35}$$

where

$$sp_x = |u + a|\,, \quad , sp_y = |v + a|\,, \quad sp_x = |w + a|\,,$$

with $\mathbf{n} = (n_x, n_y, n_z)$ being the outward unit normal vector and a is the speed of sound.

4 Results

This section presents the results obtained with the WENO schemes for a series of 2D and 3D test cases of a non-hydrostatic atmosphere.

4.1 2D Robert Smooth Bubble

After its introduction [40], the Robert smooth bubble problem is considered as a standard benchmark in NWP. The evolution of a warm bubble in a constant potential temperature environment is investigated, and since the bubble is warmer than the ambient air, it rises and due to shearing motion it forms a mushroom shaped structure as shown in Figure 3. The two-dimensional, non-hydrostatic compressible Euler equations (1) are numerically solved in the computational domain $1000m \times 1500m$ with $t \in [0, 800]\,s$ with no flux boundary conditions. The initial condition [40] corresponds to a warm bubble centered at $(500, 260)\,m$, in hydrostatic balance, and a perturbation of the potential temperature to trigger the movement of the warm bubble as follows:

$$\theta' = \begin{cases} 0 & for\ r < r_c \\ 0.25\left(1 + cos\left(\frac{\pi r}{r_c}\right)\right) & for\ r \geq r_c \end{cases},$$

where r is the distance from the centre of the bubble

$$r = \sqrt{(x - x_c)^2 + (z - z_c)^2},$$

where π is the trigonometric constant; $r_c = 250m$ is the radius of the bubble; and the potential temperature $\bar{\vartheta}$ is constant at $300K$. The unstructured meshes used consist of arbitrary triangular elements as shown in Figure 2 and have resolutions of $25m$ and $6m$. The WENO3 and WENO5 schemes using characteristics based reconstructions are used, with a CFL number of 0.9.

From the obtained computational results as shown in Figure 3 and Figure 4 it can be noticed that long-wave oscillations of the interface occur forming smaller vortices, the spatial symmetry of the solution is completely lost and that higher-order WENO schemes resolve more of these small scale vortices.

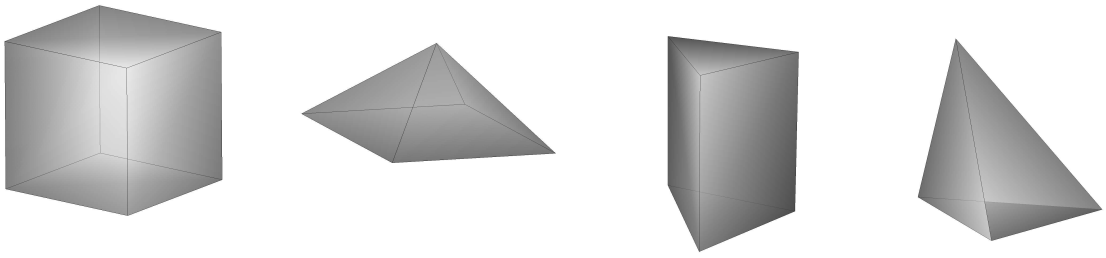
Regarding the long-wave oscillations of the interface it can be attributed to the the shear flow interaction between the interface and the surrounding unperturbed air, which in turn give rise to the Kelvin-Helmholtz instabilities that lead to turbulent structures. There have been numerous extensive numerical experiments [20, 33, 40, 60] to investigate this problem and to date the exact solution remains unknown.

The spatial symmetry of the solution is lost due to the fact that the mesh is not symmetric, the initial profile is approximated by a Gaussian quadrature rule of the same order as the polynomial order used for the reconstruction, and finally due to the multidimensional reconstruction of the present algorithms. The fact that each time the mesh is refined, or a higher-order scheme (due to the Gaussian quadrature) is used the initial perturbations change slightly. In return these perturbations according to [20] grow exponentially in time, hence the perturbations have larger spatial scales that the initial perturbations.

As the grid resolution is refined or as the spatial order of the WENO schemes is increased the numerical dissipation is reduced, hence more structures are resolved. The question is whether or not these structures are just artificial or are part of the correct solution is an issue that requires further investigation, particularly in the absence of experiments. The WENO schemes for this discontinuous problem remain stable as witnessed by [38]. However, for other approaches that do not possess discontinuity capturing capabilities a form of artificial viscosity would be required both for stabilising and for obtaining a converged solution [33]. At the same time it can be realised that some of these perturbations may be part of the correct solution [33, 60], hence controlling the amount of artificial viscosity can be quite challenging.

4.2 3D Robert Smooth Bubble

The three dimensional, non-hydrostatic compressible Euler equations (1) are numerically solved in the computational domain $1000m \times 1000m \times 1500m$ and $t \in [0, 450]\, s$ in conjunction with NFBC. The initial condition [40] corresponds to a warm bubble centered at



(a) Hexahedral (b) Pyramid (c) Prism (d) Tetrahedral

Figure 1: 3D Element shapes

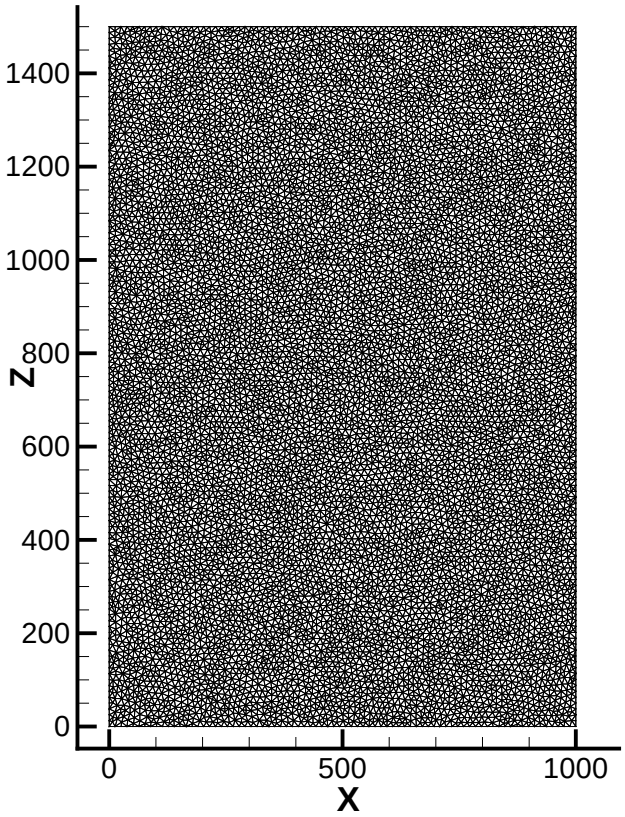
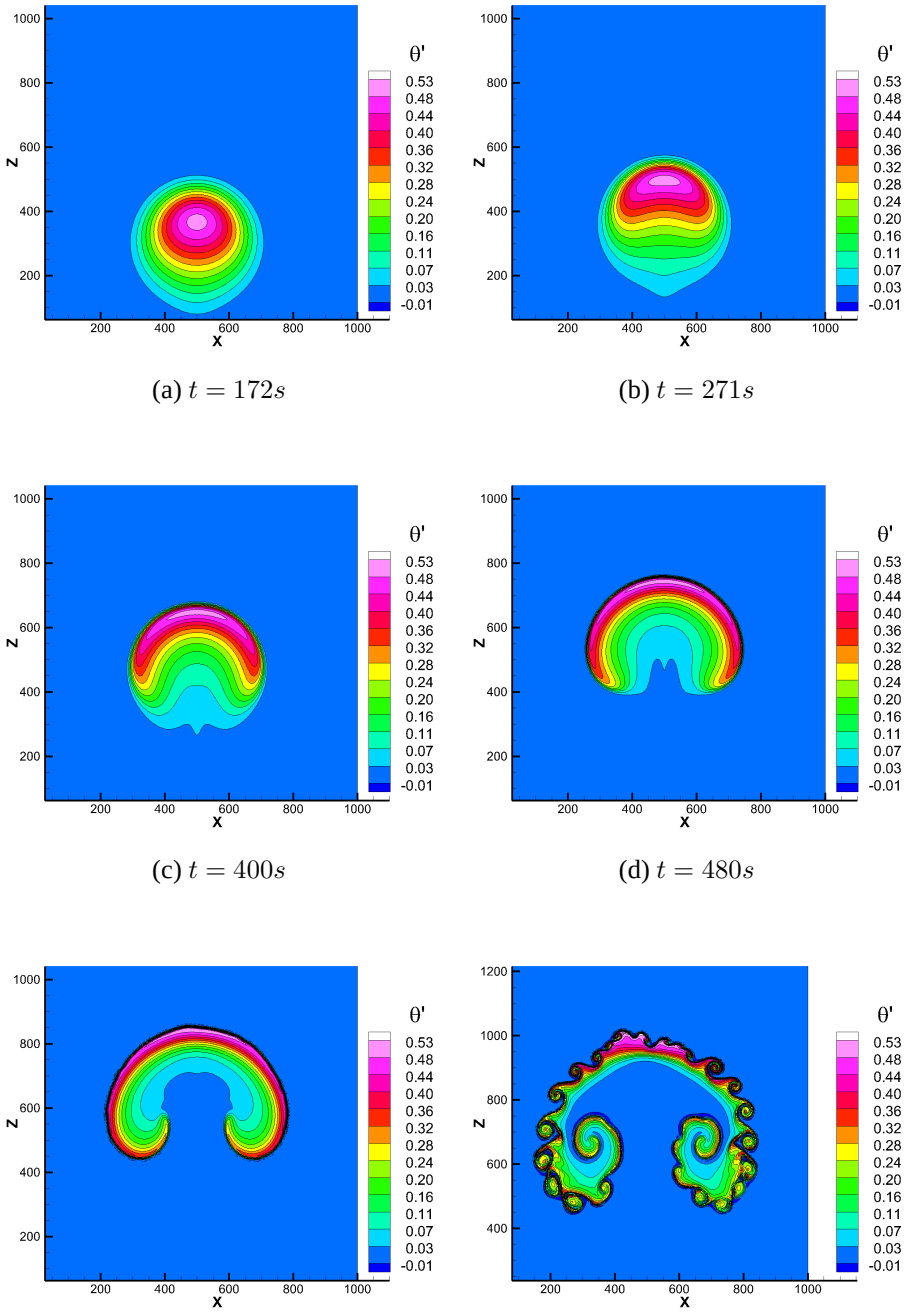


Figure 2: Unstructured mesh used for the rising thermal bubble case.

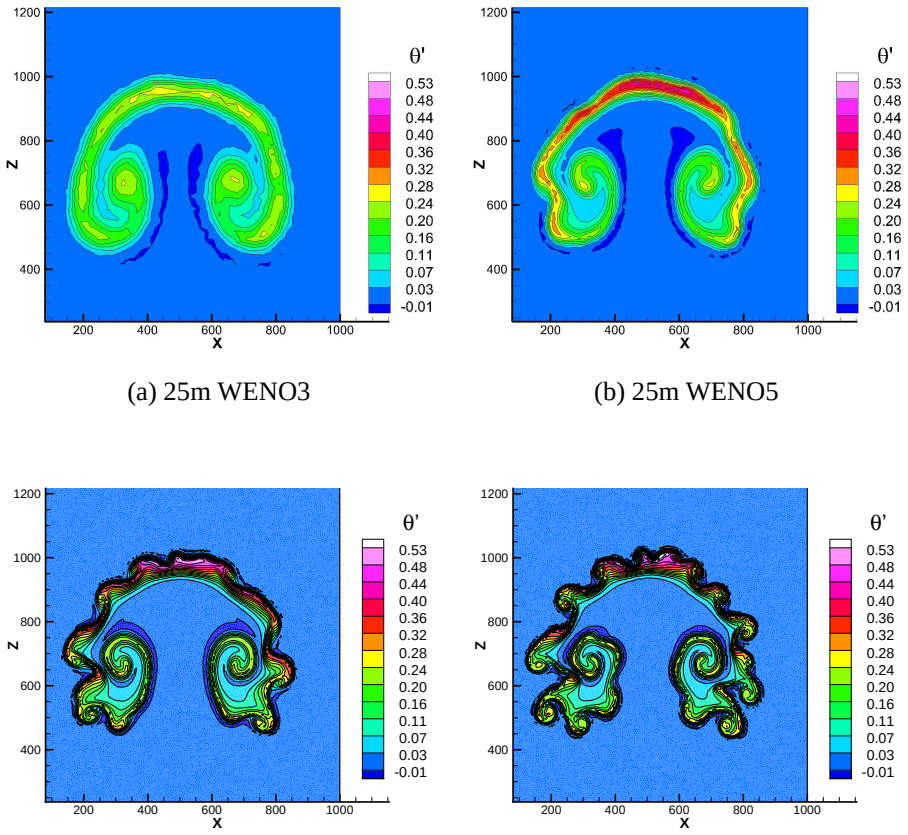


(a) $t = 172s$ (b) $t = 271s$

(c) $t = 400s$ (d) $t = 480s$

(e) $t = 600s$ (f) $t = 800s$

Figure 3: Potential temperature perturbation $\theta'(K)$ at various instants for the 2D Robert smooth bubble test on an unstructured mesh of $6m$ resolution using a WENO5 scheme.



(a) 25m WENO3

(b) 25m WENO5

(c) 6m WENO3

(d) 6m WENO5

Figure 4: Potential temperature perturbation $\theta'(K)$ at $t = 800sec$ 2D Robert smooth bubble test at different grid resolutions with WENO3 and WENO5 schemes.

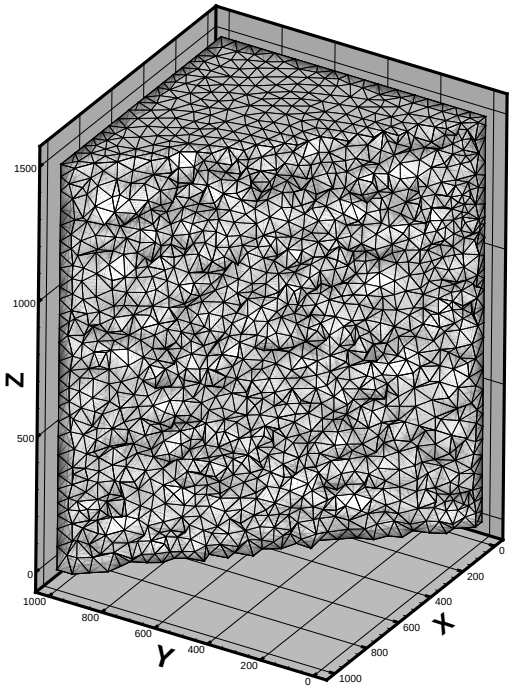


Figure 5: Cutaway of Unstructured mesh used for the 3D Robert smooth bubble test.

$(500, 500, 260)\, m$, in hydrostatic balance, and a perturbation of the potential temperature to trigger the movement of the warm bubble as follows:

$$\theta' = \begin{cases} 0 & for\ r < r_c \\ 1.25\left(1 + cos\left(\frac{\pi r}{r_c}\right)\right) & for\ r \geq r_c \end{cases}$$

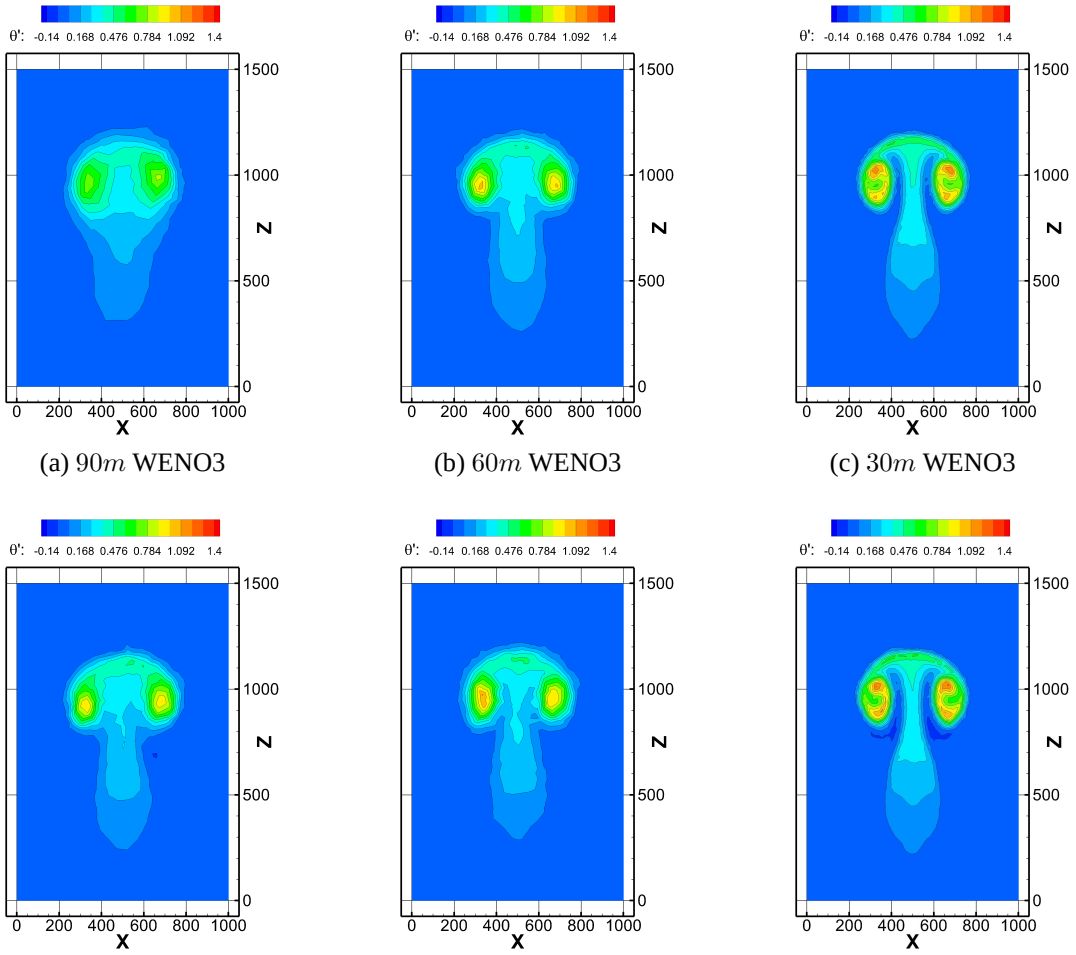
where r is the distance from the centre of the bubble

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2},$$

π is the trigonometric constant; $r_c = 250$ is the radius of the bubble; and the potential temperature $\bar{\vartheta}$ is constant at $300K$. Three unstructured meshes comprising of $19,175$, $64,000$, and $210,212$ elements have been employed using tetrahedral elements Figure 5.

The main objective of this 3D test problem is to investigate if the performance of the WENO schemes at extremely low spatial resolution $30 \sim 100m$, which is more realistic in terms of under resolved grid-arrangements used in NWP. The WENO3 and WENO5 schemes using characteristics based reconstructions are used with a CFL number of 0.9.

From the obtained computational results Figure 6, it can be noticed that the solution does not exhibit the same instabilities of the 2D test case, although the spatial symmetry of the solution is lost and that the higher-order WENO schemes provide sharper profile solutions. The non-existence of Kelvin-Helmholtz instabilities was expected due to the coarse grid resolution.



(a) $90m$ WENO3 (b) $60m$ WENO3 (c) $30m$ WENO3

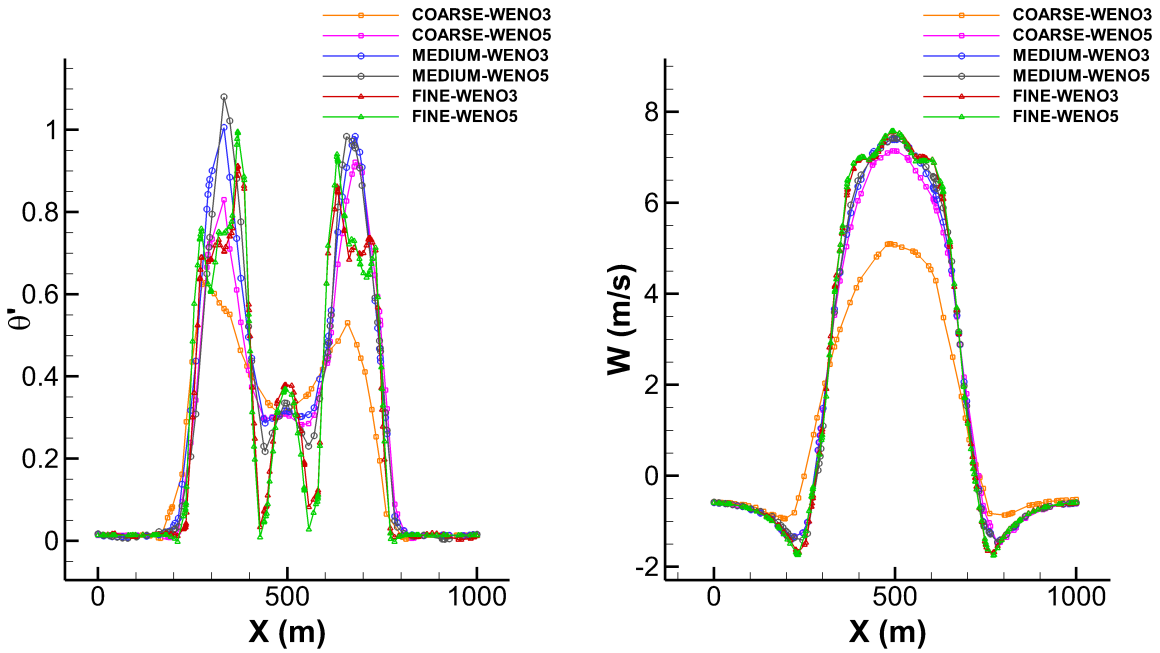
(d) $90m$ WENO5 (e) $60m$ WENO5 (f) $30m$ WENO5

Figure 6: Potential temperature perturbation $\theta'(K)$ at $t = 300sec$ and $y = 500m$ for the 3D Robert smooth bubble test.

Despite the coarse resolution, the solution breaks up its symmnetry. The asymmetry of solution is more pronounced at coarser grid resolutions. This behaviour is attributed to the multidimensional character of the numerical reconstruction and is a well known property of high-order upwind schemes exhibited particularly in flows associated with vortical structures and separation. The Potential temperature perturbation $\theta'(K)$ and vertical velocity $\mathbf{W}'(m/s)$ profiles (Figure 7) seem to trend towards convergence; although the solution is not expected to converge without any explicit viscosity due to the inviscid nature of the problem.

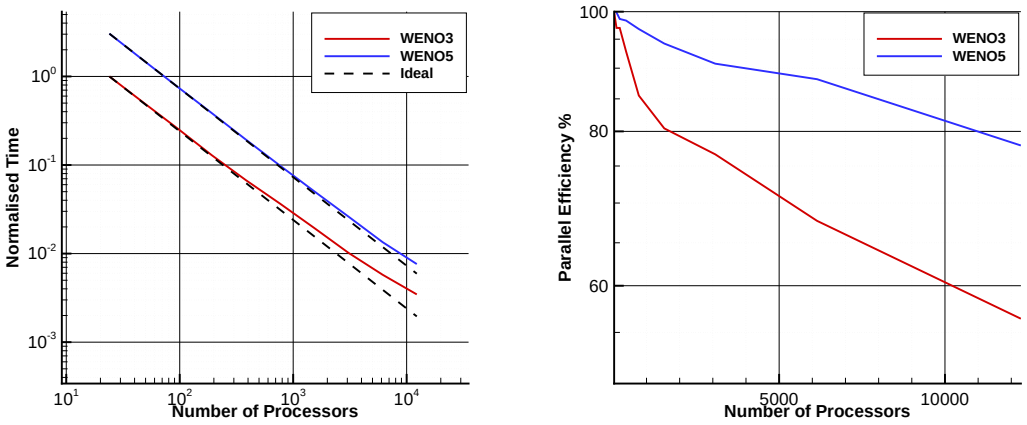
4.3 Parallel Performance

The requirements for atmospheric modelling on a regional or global scale have significantly increased in the last decade. Parallelisation of existing computational methods and software is an active area of research. Previous studies [50] showed that the ratio of computational time



(a) $\theta'(K)$ (b) $\mathbf{W}(m/s)$

Figure 7: Potential temperature perturbation $\theta'(K)$ and vertical velocity $\mathbf{W}(m/s)$ at $t = 300sec$ $y = 500m$ and $z = 964.5m$ for the Robert smooth bubble test.



(a) Scalability (b) Efficiency

Figure 8: Scalability and efficiency of WENO3 and WENO5 methods

over communication time is proportional to the order of accuracy of the WENO scheme. For WENO schemes with accuracy greater than third, the computational time is one order of magnitude greater than the communication time required for each iteration. One of the advantages of explicit methods used here is that they can be easily parallelised using domain decomposition. The software package used for partitioning uniform and hybrid unstructured meshes is the METIS software package [25]. The present numerical methods are employed in an unstructured flow solver (UCNS3D) capable of handling hybrid meshes in three-dimensions.

The Robert smooth bubble using the finest resolution, unstructured mesh with $210,212$ elements, was utilised as a test problem to assess the scalability of the developed numerical schemes. Although this grid resolution for this sample test case is relatively coarse when compared to the grid resolutions employed for climate and numerical weather prediction, our primary aim for is to demonstrate how much dependent the parallel performance is on numerical schemes implemented in the same software framework. The performance tests were conducted at the ARCHER UK National Supercomputing Service at EPCC. The results were obtained using the latest Intel FORTRAN compiler and Intel MPI library.

The time required for each time-step for the WENO3 and WENO5 schemes with respect to the number of processors is illustrated in Figure 8. The time is normalised with the time required for the WENO3 scheme.

The results confirm that high order methods scale better than lower order methods due to the fact that the ratio of computational time over communication time is greater in the lower order methods. Using 6144 processors, the parallel efficiency for the different schemes is: WENO3 (67%) and WENO5 (87%). It is pointed out that for the WENO5 scheme 92% of the computational time is spent in the reconstruction process. The exchange of the updated values of the stencil elements takes less than 3% of the total communication time since the largest communication requirement is the exchange of the reconstructed values for each quadrature point of halo-cells. The WENO high-order schemes have a greater computational cost, but

are expected to be more accurate when problems with large time scales are involved such as numerical climate modelling. For numerical climate modelling where projections of the climate for centuries are studied, WENO schemes are expected to have the strength of sufficiently small dissipation errors that will not contaminate the projected solution as a lower-order numerical scheme.

When using $6,144$ processors, each processor holds approximately 35 elements and on $12,288$ processors, each processor has less than 18 elements. The latter seems extremely low, however, it may become common practice in the future when exaflop systems will become available. These results also illustrate that improving the scalability and efficiency of numerical weather prediction models in order to take advantage of massive parallel systems, does not only require the optimisation of the computational software (software engineering), but also requires re-designing numerical methods in order to make most of the parallel performance.

5 Conclusions

We presented the implementation of high-order finite-volume schemes for atmospheric dynamics on unstructured meshes. Combining the features of unstructured meshes to accurately represent complex topologies that can be extended to adapt the resolution (dynamically or statically) in certain flow regions and using high-order numerical methods that utilise only local information, results in highly-efficient and highly-accurate numerical frameworks for numerical weather and climate modelling.

The present study discussed the implementation of WENO numerical schemes in this context and assessed their performance in terms of accuracy, robustness, efficiency and scalability. The results demonstrate the aforementioned properties of the schemes in the two-dimensional and three-dimensional rising bubble test cases, and the associated challenges of multidimensional discretisations. The WENO reconstruction can be utilised as a building-block in a dynamical core that is not limited by the type of meshes, or the type of the governing equations. Future work will concern application of the schemes in regional and global atmospheric simulations.

Acknowledgments

The authors acknowledge the computing time on the UK national high-performance computing service ARCHER that was provided through the UK turbulence consortium in the framework of the EPSRC grant EP/G069581/1. Panagiotis Tsoutsanis acknowledges PRACE for awarding him access to resource SUPERMUC and HazelHen based in Germany at Munich and Stuttgart respectively and the support of Thomas Ponweiser from RISC Software GmbH, Austria to the technical work is gratefully acknowledged.

References

[1] R. Abgrall. On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. *Journal of Computational Physics*, 114(1):45–58, 1994.

[2] A. F. Antoniadis, D. Drikakis, I. Kokkinakis, P. Tsoutsanis, and Z. Rana. High-order methods for hypersonic shock wave turbulent boundary layer interaction flow. In AIAA,

editor, *20th AIAA International Space Planes and Hypersonic Systems and Technologies Conference*, number AIAA 2015-3524, Glasgow, Scotland, July 2015.

[3] A. F. Antoniadis, P. Tsoutsanis, and D. Drikakis. Numerical accuracy in rans computations of high-lift multi-element airfoil and aircraft configurations. In AIAA, editor, *53rd AIAA Aerospace Sciences Meeting*, number AIAA 2015-0317, Kissimmee, Florida, January 2015.

[4] S. Brdar, M. Baldauf, A. Dedner, and R. Klofkorn. Comparison of dynamical cores for nwp models: comparison of cosmo and dune. *Theoretical and Computational Fluid Dynamics*, 27:453–472, 2013.

[5] C. Breviglieri, A.F.S. Maximiliano, E. Basso, and J.L.F. Azevedo. Improved high-order spectral finite volume method implementation for aerodynamic flows. 2009.

[6] B. Cockburn and C.-W. Shu. Runge-kutta discontinuous galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16(3):173–261, 2001.

[7] J.M. Dennis, R.D. Nair, H.M. Tufo, M. Levy, and T. Voran. Development of a scalable global discontinuous galerkin atmospheric model. *Int. J. Comp. Sci. Eng.*

[8] D. Drikakis, A. F. Antoniadis, P. Tsoutsanis, I. Kokkinakis, and Z. Rana. Azure: An advanced cfd software suite based on high-resolution and high-order methods. In AIAA, editor, *53rd AIAA Aerospace Sciences Meeting*, number AIAA 2015-0813, Kissimmee, Florida, January 2015.

[9] M. Dumbser and M. Kaser. Arbitrary high order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems. *Journal of Computational Physics*, 221(2):693–723, 2007.

[10] M. Dumbser, M. Kaser, V.A. Titarev, and E.F. Toro. Quadrature-free non-oscillatory finite volume schemes on unstructured meshes for nonlinear hyperbolic systems. *Journal of Computational Physics*, 226(1):204–243, 2007.

[11] M. Dumbser, D.S. Balsara, E.F. Toro, and C.-D. Munz. A unified framework for the construction of one-step finite volume and discontinuous galerkin schemes on unstructured meshes. *Journal of Computational Physics*, 227(18):8209–8253, 2008.

[12] M. Dumbser, A. Hidalgo, and O. Zanotti. High order space-time adaptive ader-weno finite volume schemes for non-conservative hyperbolic systems. *Computer Methods in Applied Mechanics and Engineering*, 268:359–387, 2014.

[13] O. Friedrich. Weighted essentially non-oscillatory schemes for the interpolation of mean values on unstructured grids. *Journal of Computational Physics*, 144(1):194–212, 1998.

[14] F. X. Giraldo, J. F. Kelly, and E. M. Constantinescu. Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (numa). *SIAM Journal of Scientific Computing*, 35:B1162–B1194, 2013.

[15] F.X. Giraldo and M. Restelli. A study of spectral element and discontinuous galerkin methods for the navier-stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases. *Journal of Computational Physics*, 227(8):3849–3877, 2008.

[16] F.X. Giraldo, M. Restelli, and M. Louter. Semi-implicit formulations of the navier-stokes equations: Application to nonhydrostatic atmospheric modeling. *SIAM Journal on Scientific Computing*, 32(6):3394–3425, 2010.

[17] Giraldo, F. Non-hydrostatic Unified Model of the Atmosphere (NUMA), 2015. URL `http://faculty.nps.edu/fxgirald/projects/NUMA/Introduction_to_NUMA.html`.

[18] C. Ollivier Gooch and M. Van Altena. A high-order-accurate unstructured mesh finite-volume scheme for the advection-diffusion equation. *Journal of Computational Physics*, 181(2):729–752, 2002.

[19] S. Gottlieb. On high order strong stability preserving runge-kutta and multi step discretizations. *Journal of Scientific Computing*, 25(112):105–128, 2005.

[20] W.W. Grabowski and T.L. Clark. Cloud-environment interface instablility: rising thermal calculations in two spatial dimensions. *Journal of Atmospheric Sciences*, 48(1):527–546, 1991.

[21] A. Haselbacher. A weno reconstruction algorithm for unstructured grids based on explicit stencil construction. pages 3369–3378, 2005.

[22] G. Hu, R. Li, and T. Tang. A robust weno type finite volume solver for steady euler equations on unstructured grids. *Communications in Computational Physics*, 9(3):627–648, 2011.

[23] G. S Jiang and C. W Shu. Efficient implementation of weighted eno schemes. *Journal of Computational Physics*, 126(1):202–228, 1996.

[24] D. Kalise and I. Lie. Modeling and numerical approximation of a 2.5d set of equations for mesoscale atmospheric processes. *Journal of Computational Physics*, 231:7274–7298, 2012.

[25] G. Karypis and V. Kumar. Multilevel k-way partitioning scheme for irregular graphs. *J. Parallel Distrib. Comput.*, 48:96–129, 1998.

[26] J.F. Kelly, F.X. Giraldo, and G. Jost. Scalability of semi-implicit time integrators for nonhydrostatic galerkin-based atmospheric models on large scale clusters. pages 561–565, 2011.

[27] B. Van Leer. Towards the ultimate conservative difference scheme. ii. monotonicity and conservation combined in a second-order scheme. *J. Comput. Phys.*, 14:361–370, 1974.

[28] B. Van Leer. Towards the ultimate conservative difference scheme. v. a second-order sequel to godunov's method. *J. Comput. Phys.*, 32(1):101–136, 1979.

[29] W.A. Li and Y.X. Ren. Quadrature-free non-oscillation finite volume scheme for solving navier-stokes equations on unstructured grids. volume 1376, pages 639–641, 2011.

[30] S. Marras, M. Moragues, M. Vzquez, O. Jorba, and G. Houzeaux. A variational multiscale stabilized finite element method for the solution of the euler equations of nonhydrostatic stratified flows. *Journal of Computational Physics*, 236(1):380–407, 2013.

[31] Max Planck Institute for Metereology and Deutscher Wetterdienst. ICOsahedral Non-hydrostatic General Circulation Model, 2015. URL `http://www.mpimet.mpg.de/en/science/models/icon.html`.

[32] A. Muller, J. Behrens, F. X. Giraldo, and V. Wirth. Comparison between adaptive and uniform discontinuous galerkin simulations in dry 2d bubble experiments. *Journal of Computational Physics*, 235:371–393, 2013.

[33] A. Muller, J. Behrens, F.X. Giraldo, and V. Wirth. Comparison between adaptive and uniform discontinuous galerkin simulations in dry 2d bubble experiments. *Journal of Computational Physics*, 235:371–393, 2013.

[34] National Center for Atmospheric Research. High-Order Methods Modeling Environment, 2015. URL `http://www.homme.ucar.edu/`.

[35] National Oceanic Atmospheric Administration. Am3, 2015. URL `http://www.gfdl.noaa.gov/am3`.

[36] National Oceanic Atmospheric Administration. Nonhydrostatic Icosahedral Model NIM, 2015. URL `http://www.esrl.noaa.gov`.

[37] X. Nogueira, L. Cueto-Felgueroso, I. Colominas, F. Navarrina, and M. Casteleiro. A new shock-capturing technique based on moving least squares for higher-order numerical schemes on unstructured grids. *Computer Methods in Applied Mechanics and Engineering*, 199(37-40):2544–2558, 2010.

[38] M.R. Norman, R.D. Nair, and F.H.M. Semazzi. A low communication and large time step explicit finite-volume solver for non-hydrostatic atmospheric dynamics. *Journal of Computational Physics*, 230(4):1567–1584, 2011.

[39] C.F. Ollivier-Gooch. Quasi-eno schemes for unstructured meshes based on unlimited data-dependent least-squares reconstruction. *Journal of Computational Physics*, 133(1):6–17, 1997.

[40] A. Robert. Bubble convection experiments with a semi-implicit formulation of the euler equations. *Journal of the Atmospheric Sciences*, 50(13):1865–1873, 1993.

[41] M. Satoh, T. Matsuno, H. Tomita, H. Miura, T. Nasuno, and S. Iga. Nonhydrostatic icosahedral atmospheric model (nicam) for global cloud resolving simulations. *Journal of Computational Physics*, 227(7):3486–3514, 2008.

[42] J. Shi, C. Hu, and C.-W. Shu. A technique of treating negative weights in weno schemes. *Journal of Computational Physics*, 175(1):108–127, 2002.

[43] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439–471, 1988.

[44] P. K. Smolarkiewicz, Szmelter J., and A. A. Wyszogrodzki. An unstructured-mesh atmospheric model for nonhydrostatic dynamics. *Journal of Computational Physics*, 254: 184–199, 2013.

[45] P.K. Smolarkiewicz. Multidimensional positive definite advection transport algorithm: An overview. *International Journal for Numerical Methods in Fluids*, 50(10):1123–1144, 2006.

[46] D. Stanescu and W.G. Habashi. Essentially nonoscillatory euler solutions on unstructured meshes using extrapolation. *AIAA Journal*, 36(8):1413–1416, 1998.

[47] V.A. Titarev, P. Tsoutsanis, and D. Drikakis. Weno schemes for mixed-element unstructured meshes. *Communications in Computational Physics*, 8(3):585–609, 2010.

[48] Eleuterio F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics - A Practical Introduction*. Berlin: Springer-Verlag, 1997.

[49] P. Tsoutsanis, V.A. Titarev, and D. Drikakis. Weno schemes on arbitrary mixed-element unstructured meshes in three space dimensions. *Journal of Computational Physics*, 230 (4):1585–1601, 2011.

[50] P. Tsoutsanis, A.F. Antoniadis, and D. Drikakis. Weno schemes on arbitrary unstructured meshes for laminar, transitional and turbulent flows. *Journal of Computational Physics*, 256:254–276, 2014.

[51] UK MET Office and ECMWF. Even Newer Dynamics for General atmospheric modelling of the environment (ENDGame), 2015. URL `http://www.metoffice.gov.uk/research/areas/dynamics/endgame`.

[52] A. Uranga, P.-O. Persson, M. Drela, and J. Peraire. Implicit large eddy simulation of transition to turbulence at low reynolds numbers using a discontinuous galerkin method. *International Journal for Numerical Methods in Engineering*, 87(1-5):232–261, 2011.

[53] Z.J. Wang. Spectral (finite) volume method for conservation laws on unstructured grids. basic formulation. *Journal of Computational Physics*, 178(1):210–251, 2002.

[54] Z.J. Wang and Y. Liu. Spectral (finite) volume method for conservation laws on unstructured grids. ii. extension to two-dimensional scalar equation. *Journal of Computational Physics*, 179(2):665–697, 2002.

[55] Z.J. Wang, L. Zhang, and Y. Liu. Spectral (finite) volume method for conservation laws on unstructured grids iv: Extension to two-dimensional systems. *Journal of Computational Physics*, 194(2):716–741, 2004.

[56] W.R. Wolf and J.L.F. Azevedo. High-order eno and weno schemes for unstructured grids. *International Journal for Numerical Methods in Fluids*, 55(10):917–943, 2007.

[57] Z. Xu, Y. Liu, and C.-W. Shu. Hierarchical reconstruction for discontinuous galerkin methods on unstructured grids with a weno-type linear reconstruction and partial neighboring cells. *Journal of Computational Physics*, 228(6):2194–2212, 2009.

[58] Z. Xu, Y. Liu, and C.-W. Shu. Hierarchical reconstruction for spectral volume method on unstructured grids. *Journal of Computational Physics*, 228(16):5787–5802, 2009.

[59] C. Yang and X. C. Cai. A scalable fully implicit compressible euler solver for mesoscale nonhydrostatic simulation of atmospheric flow. *SIAM Journal of Scientific Computing*, 36: S23–S47, 2014.

[60] L. Yelash, A. Muller, M. Lukacova-Medvid'ova, F.X. Giraldo, and V. Wirth. Adaptive discontinuous evolution galerkin method for dry atmospheric flow. *Journal of Computational Physics*, 268:106–133, 2014.

[61] G. Zangl, D. Reinert, P. Ripodas, and M. Baldauf. The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 2014.

[62] Y. T. Zhang and C. W. Shu. Third order weno scheme on three dimensional tetrahedral meshes. *Communications in Computational Physics*, 5(2-4):836–848, 2008.

[63] Y. Zhou and Z.J. Wang. Implicit large eddy simulation of transitional flow over a sd7003 wing using high-order spectral difference method. 2010.

[64] J. Zhu, J. Qiu, C.-W. Shu, and M. Dumbser. Runge-kutta discontinuous galerkin method using weno limiters ii: Unstructured meshes. *Journal of Computational Physics*, 227(9): 4330–4353, 2008.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELLING OF LONG WAVES (SEICHES) IN CASCAIS BAY

Vera Bras[1], **Antonio Trigo-Teixeira**[2]

[1] Instituto Superior Tecnico
Lisbon University
AV. Rovisco Pais, 1 1049-001 Lisboa
vera.bras@tecnico.ulisboa.pt

[2] CERIS – Instituto Superior Tecnico
Lisbon University
AV. Rovisco Pais, 1 1049-001 Lisboa
trigo.teixeira@tecnico.ulisboa.pt

Keywords: Seiches, harbour resonance, periodogram, pass-band filter, numerical modelling, CGWAVE.

Abstract. *This study is focused on the analysis and characterization of long waves (seiches) in Cascais Bay-Portugal, where the phenomenon is known to occur. For this purpose, time series records of water surface elevation (WSE) from the tidal gauge were analyzed in time and frequency domains. The periodicities of the long waves were found, firstly with a plotting of time series, where two waves appear immediately: the astronomic tide (T=12h 25min) and a long wave with period of approximately (T=12min), and secondly with the application of the spectral FFT analysis to obtain the periodogram. Before the application of the FFT, the WSE elevation time series was first filtered for the astronomic signal using harmonic analysis and then a pass-band of second order filter was used to removal the noise. The frequency analysis allows identification of hidden periodicities and conclude that most of the spectral energy was concentrated on a single frequency. Finally, some statistics were calculated, which indicated the presence of long waves in the tidal record for most of the period of analysis, without apparent relationship with the season of the year: maritime summer (May-Oct) or maritime winter (Nov-Apr).*

On a second phase the setup of a 2D finite element model to solve the mild-slope equation (MSE) was done to study the phenomenon of long wave propagation both into the bay and in a small craft harbour (Cascais Marina). The model used was CGWAVE [1] *and it was forced with long waves of different frequencies on the boundary to investigate the appearance of seiches both in the bay and in the harbour. The model permits the identification of resonant periods and the location of nodal lines (where vertical WSE oscillation is nil and water velocity is at a maximum) and ventral lines (where just the opposite occurs). The hydrodynamic drag on boats moored on nodal lines is a cause of great trouble on the Marina operation. It might be concluded that seiches occur, not only in the marina basin, but in all water body of the bay, so it is a natural phenomenon on the area.*

1 INTRODUCTION

The present work focused on the modelling of long waves (seiches) in Cascais Bay- Portugal, which is a frequent local phenomenon. It is believed that the word "seiche" derives from the Latin word "*siccus*", which means dry or expose. So, long waves (seiches) are "standing waves" as perceived by the periodic oscillations of the surface of an the enclosed basin or semi-enclosed body of water (lake, gulf, inland sea, bay, harbour) with periods from a few minutes to a few hours caused it by such phenomena as atmospheric pressure changes, winds, tidal currents and earthquakes. Usually, it is considered as a transient event because after some time subsequently to the end of the geophysical event that causes it, it eventually dissipates. However, in some places, it is also observed continuously with small amplitudes. [2], [3]

A natural or an artificial basin may offer good conditions of shelter for the open sea, particularly for wind waves (short waves), however both the geometry of the basin and their internal reflections can efficiently amplify the incident long waves, even for those with small amplitudes. So, this phenomenon of resonant oscillations is called seiches, harbour oscillations or harbour surging and is related to specific periods of the incident waves on the structure. For a basin, if its period of the oscillation is close to that the period of the incident long wave (or other possible excitation), an extreme seiche could result. Extreme seiches induce large oscillations and strong currents that might cause damage to ships and breaking moorings lines. From what was exposed, the study of long waves in coastal zones is very important. [4]–[7]

On the literature, various authors were dedicated to study of the resonant properties of harbour basins, which would be divided in two main groups: simple geometry (Miles and Munk 1961; Lee 1992; Ippen and Goda 1963; etc.) and complex geometry (Knapp and Vanoni 1945; Wilson 1960; Lee 1969; etc.). The first group provides theoretical expressions for approximated calculation of the natural period oscillation for closed and semi-enclosed basins. For the second, the use of theoretical formulas cause discrepant values of the reality because the non-linearity of the basin's shape, so physical or numerical models had to be used in these cases. Numerical models are a powerful tool for studying long waves in basins with complex geometry (real basins) and for more accurate results, when the period of the waves is higher than 400 *s* they can replace the physical models. [8], [9]

Nevertheless it was important to refer that, theoretically, in semi-enclosed basins, like bays, it's expected a nodal line across of this entrance, because the waters within the basin must communicate with waters outside. [10]

Recently, the implementation of modern's digital tidal gauge in the coast allowed a faster recording rate (range of seconds), which is a huge advantage in the use of time series data to study long waves. However, these measurements are mostly used for analyzing extreme events such as tsunamis. [11]

This study had two main objectives. The first one was to analyze the time records of water surface elevation (WSE), from the digital tidal gauge of Cascais, for the characterization of long waves (period and amplitude). These periodicities were found by making the analysis on the time and frequency domains and, finally, the results of both methods were compared.

The second main objective was to evaluate the resonance phenomenon inside of the marina, using the CGWAVE model for this hydrodynamic modelling, because the complex shape of the marina. Therefore it is possible to match the information relatively to the periods of seiches and say if the range of periods, that were observed, correspond to the resonance periods of the basin, using the numerical modelling.

2 ANALYSIS OF THE LONG WAVES (SEICHES)

2.1 Introduction

In the ocean there are different types of long waves, generated by different phenomena and with different periods and wavelengths. When waves have wavelengths much higher than the depth water, they are called long waves and have periods between 5 *min* and 24 *h* (tides). These waves have a frequency band with relatively low spectral energy, which is between the frequency band of tides and wind waves, both with higher energetics bands. [5]

The measurements of long waves are not direct; the time series from the tidal gauge can be described by:

$$\eta(t) = \bar{\eta}(t) + T(t) + R(t) \tag{1}$$

Where, $\eta(t)$ is the observed water surface elevation which varies with time; $\bar{\eta}(t)$ is the mean sea level which changes slowly with time; $T(t)$ is the astronomical tidal component and $R(t)$ is the residual component (meteorological component and noise).

So, the meteorological tide originated by the meteorological phenomena, such as atmospheric pressure variation and wind's effect, which resulted on water oscillations of just a few centimeters and it was responsible for the difference between the predicted and observed levels. Therefore, the long wave signal has to be isolated from the records observed of WSE in the tidal gauge. The procedures of this treatment of the times series and the following characterization of the seiches, in time and frequency domains, are described below.

2.2 Data processing

The upgrade of the analogic coastal tide gauge into the digital one allowed a faster sampling rate, which permitted a better identification of long waves in the records because more sampling points (*N*) provided more information and, consequently, more accuracy.

The tidal gauge of Cascais is operated by the IGEO (Portuguese Geographic Institute) and the digital time series records of WSE are available online in their website since 2007. These are discrete time series because the observations are taken on equally spaced intervals, in this case with sampling rate of $\Delta t = 5$ *s* and it results in 17280 lines of information per day. When the daily records had failures (less than 17280 lines) they were not used. For the present work 17 months were analyzed: 1st semester of 2015 and all year of 2014, except February (because it had gaps).

2.3 Time series graphs

Usually, the first step in the analysis of time series is to plot the data over time to obtain some characteristics like trend, variation in the time, periodicities and other fluctuations. Regarding this step, two types of waves were immediately apparent: the astronomic tide with a period of (T=12*h* 25*min*) and a long wave with a period of approximately (T=12 *min*).

The astronomic tide is generated by the combined effect of gravitational force that the Sun and Moon exerts on the Earth's and in the Cascais bay is of the mesotidal type (tidal range of 4.0 *m* in Spring Tides) and is semi-diurnal.

2.4 Filtering of the time series for the astronomic signal

When time series can be predicted exactly from past observations, because the successive observations are dependent, they are called deterministic, otherwise they are called stochastic. [12]

The deterministic component of the time series records of WSE corresponds to the astronomic tide and the stochastic to the meteorological tide. The first one can be predicted by a harmonic analysis, based on the movement of the Sun and the Moon, and in this step two methods were used: tide tables from IH (Portuguese Hydrograph Institute) and Wx-Tide software.

The astronomic component was determined based on these two methods, which had the same trends but different WSE. This difference can be explained with the global sea level rising, about 1.3 *mm* per year. While the tidal height from the tables does not take this factor in count, but refers that it should be added about 0.10 *m* to the predicted heights, the tidal height from the software had updated harmonic constants and, consequently, higher WSE.

The Wx-Tide just allows the minimum predicting rate of Δt= 60 *s* and, to assure that this parameter in this process is equal to the one from the tidal gauge (Δt=5 *s*), it was decided to consider the method of the tidal tables. However, an adjustment was made of about 0.10 *m*, as recommended. In the Figure 1 the results of the previous methodology for an extreme event of seiche are presented.

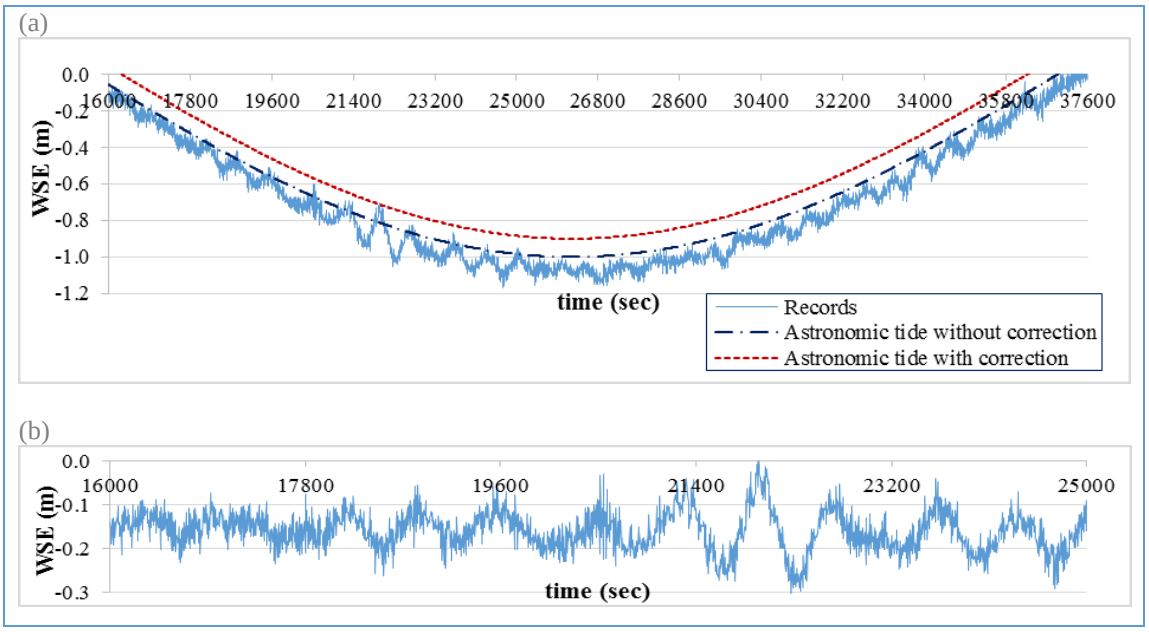


Figure 1: (a) Time series records of WSE (6*h*) from the Cascais tide gauge and astronomic trend; (b) astronomic tide removal for the time series – exert (2*h* 30*min*).

By observing the Figure 1, it is possible to see that the astronomic tide trend was removed by the difference between the records of WSE and the predicted tide by the harmonic analysis. The oscillations of the seiches had, usually, small amplitudes when compared with the tide oscillations. Thus, the long wave oscillations had a periodical trend which can be approximately described for a sinusoid, given by:

$$\eta(t) = \overline{\eta} + A\,sen(\omega t + \phi) \tag{2}$$

Where, $\eta(t)$ is the water surface elevation at time t; $\overline{\eta}$ the mean level of water surface elevation; A the amplitude of the oscillation; ω the angular velocity, which can be obtained by $2\pi/T$, where T is the period and, finally, $\varnothing$ is the phase lag relative to some defined time zero.

2.5 Pass-band filter of second order

In fact, the noise is a random signal and consequently it can be eliminated in time domain. In this phase a band-pass of second order filter (Butterworth filter) was applied to remove the noise in frequency domain. A "butter" function was used from the "Signal Processing Toolbox" - MATLAB software.

It was decided that the digital Buttherworth filter had a normalized cutoff frequency between 1/400 *Hz* and 1/900 *Hz*, which corresponds to the interval of frequencies of the seiches. This means that all signals out of this range are eliminated.

2.6 Statistical analysis

By direct observation, when regularity in the distances between peaks is observed, this period of time was counted, as well as the number of intervals between peaks. The period of this long wave is calculated by the ratio between this interval of time and the number of peaks. Half of the distance in vertical between two consecutives peaks gives the amplitude of this oscillation. [13]

In the Figure 2 some results of the statistical analysis can be observed.

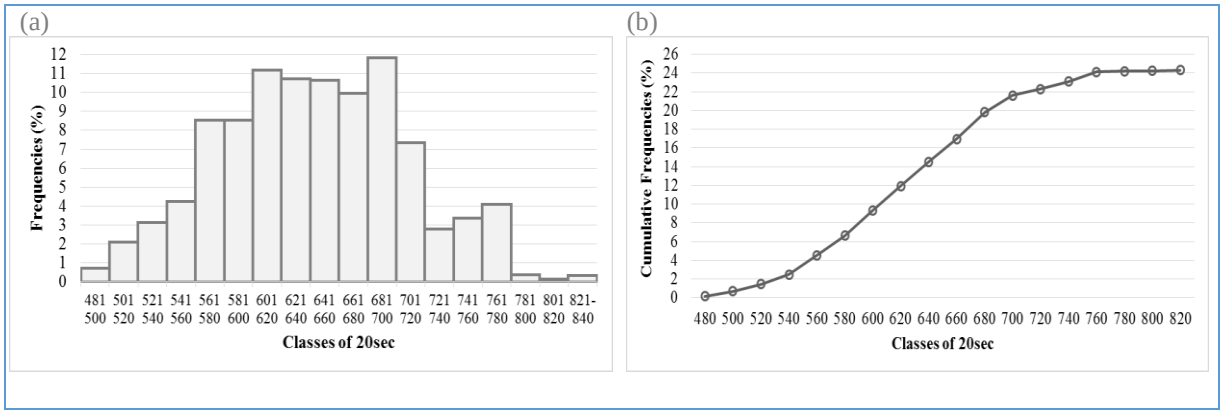


Figure 2: (a) Frequency Histogram in relation to the total time of observation of the seicha phenomenon (2420*h*) (b) cumulative frequency curve in relation to the total time of analysis (9960*h*).

According to the Figure 2 it is possible to conclude that the seiches occur with a frequency of 24.30% and the most frequent periods are in the range of 600-700 *s* (10.0-11.7 *min*). The phenomenon occurs for periods between 480 *s* and 840*s* (8-14 *min*).

Since low pressures are associated to dangerous and extreme meteorological phenomenon, an attempt to correlate them with an occurrence of seiche and pressure variations was made. However, the results did not indicate any type of relation with this possible cause of the origin of seiches in Cascais-bay.

2.7 Spectral analysis - Periodogram with the FFT algorithm

The Fourier analysis is an advantageous technique to represent time series in terms of the frequencies distribution. If certain conditions are satisfied, any function may be formulated as a sum of a series of sines and cosines of a spectral resolution ($\Delta f = 1/N\Delta t$). However, the spectral resolution depends only on the length of the time records, which means that short time series' performance is not as good. [14]

The periodogram provides the graph representation of the PSD (Power Spectral Density) *versus* spectral resolution, with a FFT algorithm.

The results of this spectral analysis were presented in Figure 3, for an extreme event of seiche and for different spectral resolutions.

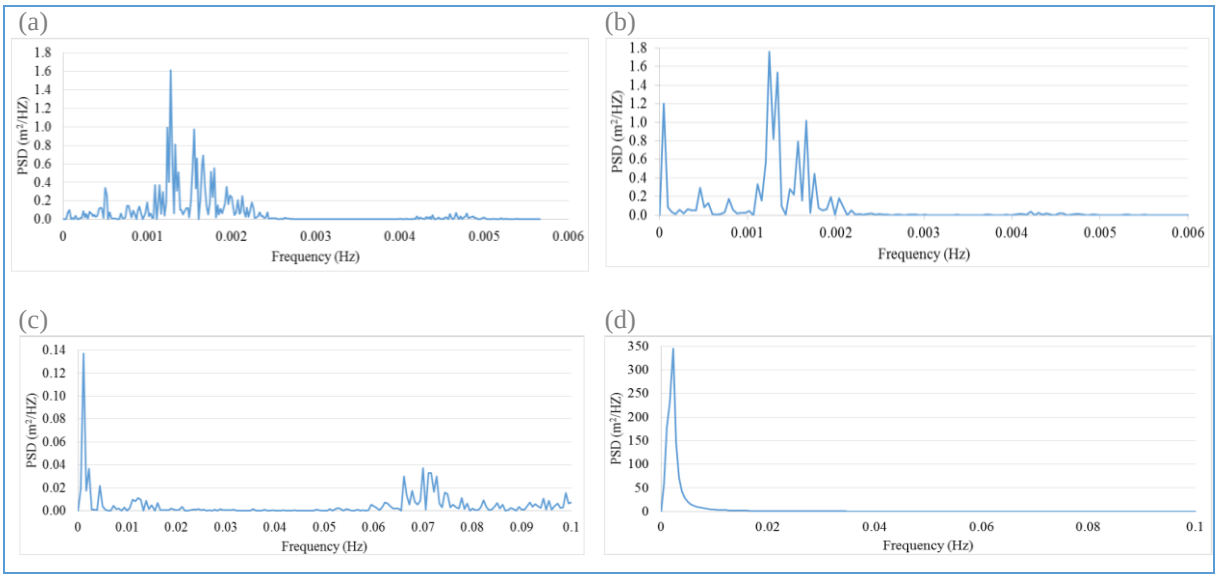


Figure 3: Identification of seicha phenomenon used the periodogram method. (a) computed to time of seiche (15*h*), with T_{peak} = 783 *s*; (b) computed to 6*h*, with T_{peak} =800 *s*; (c) computed to 30 *min*, with T_{peak}=900 *s* (d) application of the pass-band filter (time series of 30 *min*), with T_{peak} =485 *s*.

In first instance, it was observed that, for a greater number of samplings, the resolution is finer, and a correspondent spectral plot is noisier and irregular. So, a more acceptable performance is obtained with 30 *min* of times series, which provides 360 time observations but only 180 (*N/2*) were used in the FFT algorithm. Thus, after the application of the pass-band filter the noise signal (highs frequencies) was completely eliminated.

Finally, for the 30 *min* time series the spectral energy is mostly concentrated on a single frequency (low frequency) and these periods are in agreement with the ones of the direct analysis. So, one of the advantages of this analysis is the possibility of being able to easily identify the frequency where the spectral energy was concentrated – frequencies of the peak.

3 NUMERICAL MODELLING

In this section, the setup of a numerical model was done to study the phenomenon of long wave propagation, both into the bay and in a small craft harbour (Cascais marina). The model used was CGWAVE, a linear finite element method based on the elliptic mild-slope wave equation (MSE), and can simultaneously simulate the effects of refraction, diffraction, reflections by the bathymetry and structures, dissipation due to friction and breaking and nonlinear amplitude dispersion. [1]

3.1 Set up of the model

The model setup uses a detailed LIDAR bathymetry acquired recently for the coastal area and land boundaries are discretized based on the type of boundary. The ocean boundary is a semicircular arc enclosing the area of interest. For the computational simulations, the domain was defined with a linear triangular finite element mesh with 10 nodes per wavelength.

Typically, the minimum radius of the semicircle arc should be about two or three times the wavelength (*L*), which for long waves is given by:

$$L = T\sqrt{gd} \tag{3}$$

Being, *T*, period; *g*, gravity acceleration and *d*, depth of the water. According to this, for a long wave with period of T=650 *s* and a constant depth of d=1 *m*, the minimum radius obtained, for this conditions, should be between 4070 *m* and 6105 *m*.

Beside this, the distance between the open boundary and the shoreline is also important. If they are too close, the bottom effects could influence the output results from wave height and, if they are too far, it is possible that the model does not obtain sufficient bathymetry information and the simulations do not run. After several attempts, the domain was defined according to the previous specifications, with a final value for the radius of 9300 *m*.

3.2 Friction coefficient

Typically the values for friction coefficient are related with the bottom irregularity, roughness of the bottom, structures that modify the flow field and other parameters. This effect, on the period of resonant oscillations of the basin, depend slightly on water depth, decreasing as depth increases. [10]

Some attempts were made to understand the influence of the friction coefficient and it was observed that this parameter is not relevant for the output results. Regarding this conclusion, and because taking the friction in consideration is very time consuming, the model was performed for the "no friction".

3.3 Reflection coefficient

According to literature, for long waves it is possible to affirm that the reflection coefficient is unitary for the land boundaries, which means that total reflection is admitted.[15] However, for different coastline materials the reflection coefficient had values between 0 and 1 as shown in the table below. [16]

Coastline material	Vertical wall	Shore defense (rocks)	Beach
Reflection Coefficient (-)	0.90	0.25	0.10

Table 1: Reflection coefficients to coastline materials

3.4 Definition of the study scenarios

The most dangerous stages of agitation for the numerical propagations of the incident waves were associated to the wave directions of 45º and 60º, because the predominant wave direction is south-west (45º). So, the model was forced for these five scenarios, for events with the long waves amplitude's of 0.5 *m* (1 *m* of height) in the open boundary and for periods between 350 *s* and 950 *s*, with increments in time of 50 *s*.

Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Direction (º)	60	45	45	45	45
Reflection Coefficient (-)	1.0	1.0	variable	0.5	0.9

Table 2: CGWAVE boundary conditions – different scenarios

4 RESULTS AND DISCUSSION OF THE NUMERICAL MODELING

The response curve provides the resonant periods, which are the periods that correspond to the amplification factor higher than 1. It is relevant to show that the amplification factor (A) is given by:

$$A = \frac{Hi}{Hr} \tag{4}$$

Where the *Hi* is the height of the wave in a certain point and the *Hr* is the height of the incident wave (*Hr*=1*)*. When *A*>1 the wave is amplified and harbour resonance occurs. Nevertheless, it is important to note that, in the real events of resonance, these amplified waves do not have necessarily heights of 1 *m* (or more), it is a value that will give agitation indexes, since the model is linear and the height of the wave in the open ocean is not known. The response curves for five positions inside of marina are shown.

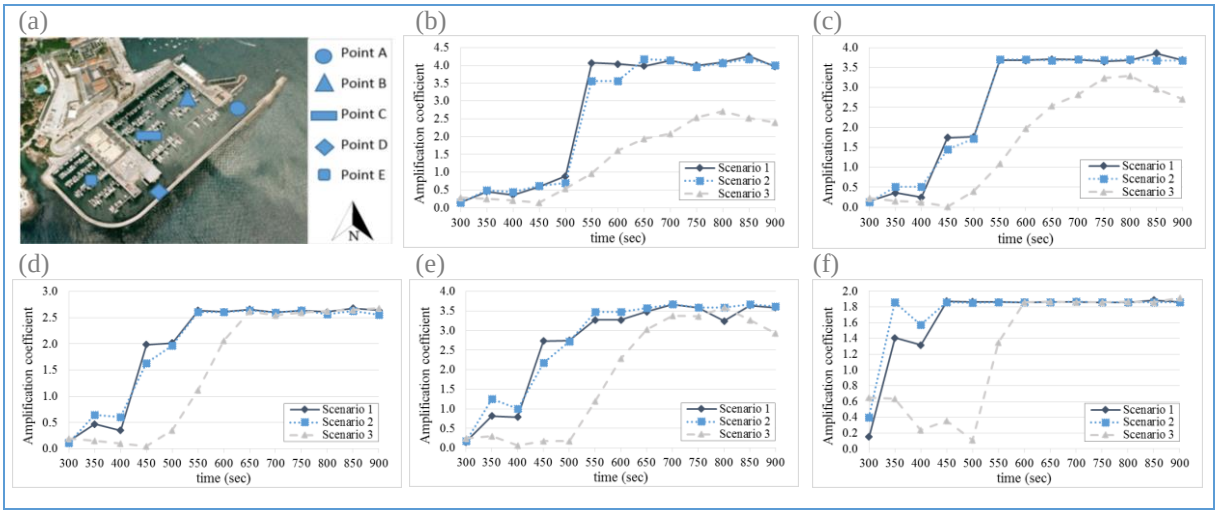


Figure 4: (a) Position of the points inside of the marina. Image from google earth, adapted. Response curves from scenarios 1, 2 and 3 in (b) Point A; (c) Point B; (d) Point C; (e) Point D; (f) Point E

Focusing in the Figure 4 it was perceived that it would be possible to decrease the interval between runs to provide a more precise curve of amplification factor but, for this problem, the analysis seems to be enough detailed.

For the range of periods that was observed in time series records of the WSE, the amplification factor is higher than 1 and consequently resonance really occurs inside of the marina. The trend of the response curve of the scenarios 1 and 2 are practically the same in both cases (maxima and minima are situated on the same range of periods), as well as their asymptotes.

The response curves of the scenario 3 also showed a similar trend in almost all positions (less for point A) that displayed a lower amplification coefficient. On the other hand, for this situation, the resonance happens on higher periods and this can be explained by a variable and smaller reflection coefficient on the boundaries.

Relatively to all of the scenarios analyzed, it is evident that, for starting periods around the 500 *s*, the resonance curve reaches an asymptote and these values of the amplification factor remain almost constant.

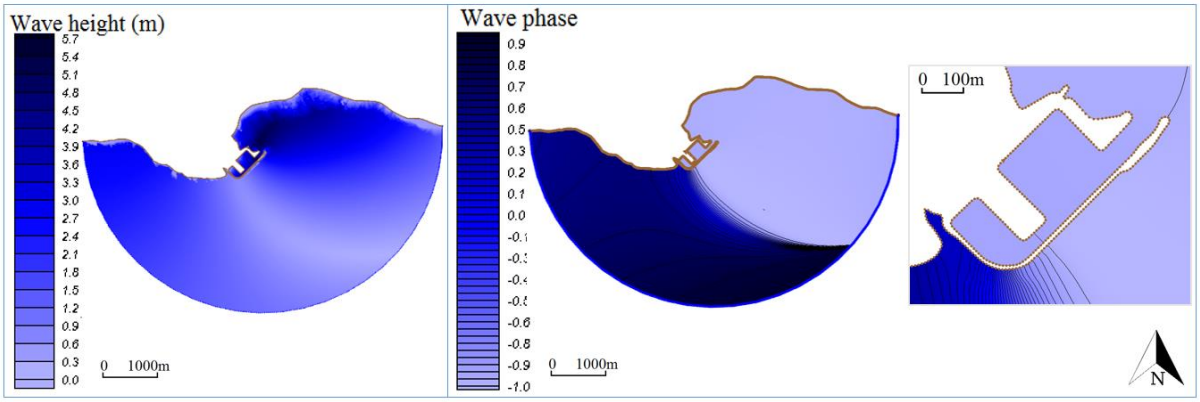


Figure 5: CGWAVE results – Cascais Bay and Marina. Scenario 1 for incident long wave of T=800 *s*.

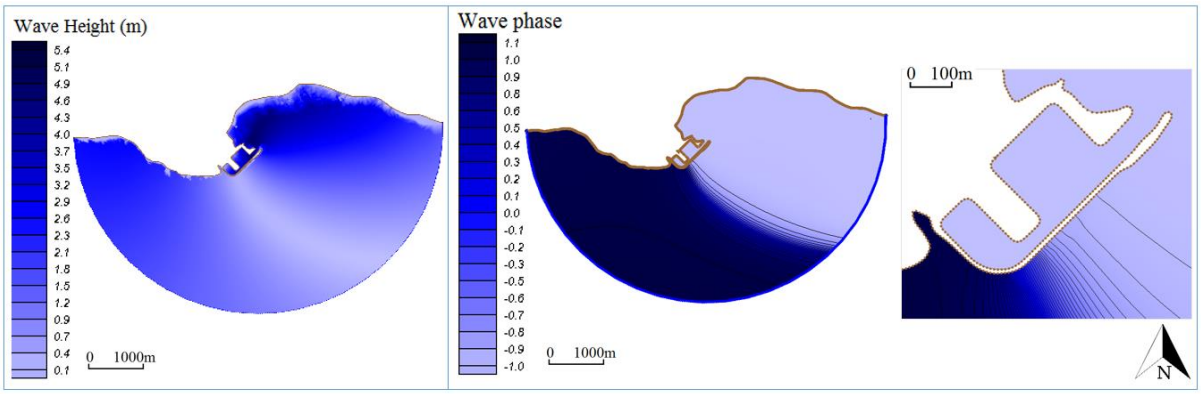


Figure 6: CGWAVE results – Cascais Bay and Marina. Scenario 2 for incident long wave of T=800 *s*.

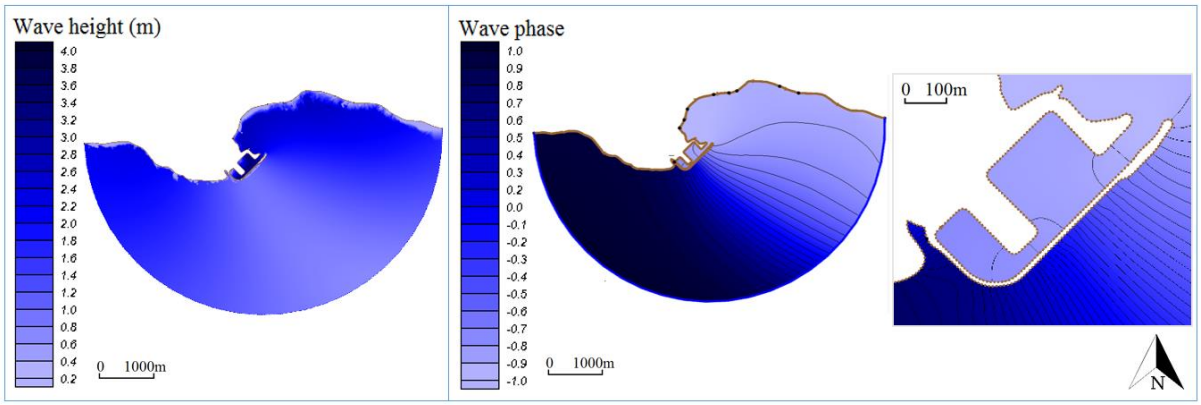


Figure 7: CGWAVE results – Cascais Bay and Marina. Scenario 3 for incident long wave of T=800 *s*.

The output results of the wave phase were used on the identification of the nodal lines (where vertical WSE oscillation is nil and the water velocity is at a maximum), which corresponds to the wave phase when it is equal to zero. In ventral lines the opposite occurs.

All scenarios are devoid of nodal lines inside the marina or in their mouth, and the single nodal line was approximately on the middle of the exterior of the marina, and it was regularly extended across the whole bay. For the resonant periods, the bay responds like a big reservoir, without great vertical oscillations of the water body or great waters velocities.

Therefore, for an instant, the water levels are practically the same in all places of the bay and a current allows, both the entrance and the exit, of the water from the bay to the inside of the marina, making the water level to fluctuate.

5 CONCLUSIONS

One of the main purposes of this work was analysis of the characteristics of the long waves in Cascais-bay. From the analysis and data processing of the time series records of WSE from the local tide gauge, it was concluded that the seiche event occurs with a frequency of 24.30% of the total observations time, for a range between 480 *s* and 840 *s* (8-14 *min*), with mean values of about 660 *s* (11 *min*). The amplitude of these oscillations are normally uniform, with ranges of a few centimeters however, in an exceptional event, the maximum observed amplitude was of 0.54 *m*.

From the statistical calculations it was observed that the presence of long waves are a frequent event (presents in 272 days) but it is very probable that these oscillations had been more frequent in Cascais-bay than the present study reports. This happens because the local tide gauge has gaps in the measurements, in extreme situations even for a month. From the pressure records from the local tide gauge, no evidence about any relation of occurrence of this phenomenon with the season of year was found. However, the amplitudes of the seiches are larger on the maritime winter (November to April).

A Matlab program was written for the acquisition of the heights of the astronomic tide by the harmonic analysis (tide tables), as well for the formulation of the periodogram method and the pass-band filter of second order (Butterworth filter). The spectral analysis allows a complementary and alternative option in relation to the direct method (time domain) for the measurement of the seiches periodicities.

Relatively to the periodogram, for exerts of 30 *min* of the time series, this peak period is in good agreement with the periods of seiches obtained by direct observation. So, these convergent results allows a confirmation of the primary periodicities of the long wave oscillations observed in Cascais-bay, as well as an extra care for hidden periodicities. Although different domains for the analysis of the period seiches were used, both methods are valid.

From the CGWAVE simulations of the long wave penetration in Cascais-Bay it was possible study two problems: the resonance phenomenon and the nodal lines position. About the resonance phenomenon, firstly the resonant periods are between values of around 400 *s* and 900 *s*. Secondly the bay response to the forced open ocean waves did not have significant variations with the incident angle. On the other hand, the reflection coefficient had some implications on the response curve, like the increase of the resonant periods on the basin response. The main advantage of this is that the highest periods are harder to occur in reality, so the occurrence of harbour resonance is less likely. So, these results were useful on showing that, even for long waves, the model is sensible to different reflection coefficients, especially for lower periods, which does not corroborate the theoretical hypothesis.

Another significant conclusion is that the periods of seiches that were obtained in the first phase of this work are in good agreement with the computed resonant periods of the bay (when A>1). This means that the periodicities of the observed oscillations really correspond to the amplification of the incident long waves from the open ocean. So, CGWAVE model gives good results for the present study.

Finally, according to the non-existence of nodal lines inside of the marina and in their mouth, it was possible to conclude that the basins oscillates in unison, so dangerous situations in moored positions and problems in the marina operation are not expected to occur.

In order to find more information about the physical processes that originates seiches in the Cascais bay, further studies are needed. An attempt to link the occurrence of seiches with the passing of atmospheric depressions did not show any correlation.

REFERENCES

[1] V. G. Panchang and B. Xu, "CGWAVE: A Coastal Wave Transformation Model for Arbitrary Domains." Tech. Report. Dept. of Civil and Environmental Engineering, University of Maine, 1995.

[2] A. B. Rabinovich, "Seiches and Harbor Oscillations," *Handb. Coast. Ocean Eng. World Sci. Publ., Singapoure,* 2009.

[3] J. Park, J. Macmahan, W. V Sweet, and K. Kotun, "Continuous seiche in bays and harbors," *Ocean Sci. Discuss*, vol. 12, pp. 2361–2394, 2015.

[4] J. W. Miles, "Harbor Seiching," *Annu. Rev. Fluid Mech.*, vol. 6, pp. 17–35, 1974.

[5] W. H. Munk, "Origin and generation of waves." Coastal Engeering. La Jolla, California: Institute of Geophysics and Scrips Institution of Oceanography - University of California, 1950.

[6] A. R. Kabiri-Samani, "Natural Frequencies of Seiche in a Closed Trapezoidal Basin with Internal Barriers," *J. Civ. Eng. Res.*, vol. 3, no. 1, pp. 22–34, 2013.

[7] J.-J. Lee, "Wave induced oscillations in harbours of arbitrary shape," *J. Fluid Mech.* , vol. 45, no. 02, pp. 375–394, 1969.

[8] U.S.Army Corps of Engineers, *Coastal Engineering Manual, Parte II-7, Harbor Hydrodynamics*. Washington DC: U.S.Army Corps of Engineers , 2008.

[9] L. S. Lillycrop, "Barbers Point Harbor, Oahu, HI, Monitoring Study." U.S. Army Corps of Engineers, Washington DC, 1993.

[10] L. C. Breaker, Y.-H. Tseng, and X. Wang, "On the natural oscillations of Monterey Bay: Observations, modeling, and origins," *Prog. Oceanogr.*, vol. 86, pp. 380–395, 2010.

[11] L. Bressan and Tinti. S., "Statistical properties of coastal long waves analysed through sea-level time-gradient functions: exemplary analysis of the Siracusa, Italy, tide-gauge data," *Nat. Hazards Earth Syst. Sci. Discuss.*, vol. 3, no. 9, pp. 5247–5286, 2015.

[12] C. Chatfield, *The Analysis of Time Series: An Introduction*, 2nd ed. New York: Chapman and Hall, 1980.

[13] F. M. Abecasis, "Estudo do problema das ondulações de longo periodo na ampliação do porto de Luanda," *"Técnica" Rev. dos alunos do I.S.T.*, vol. 303, p. 483 a 492, 1960.

[14] R. E. Thomson and W. J. Emery, Eds., *Data Analysis Methods in Physical Oceanography* , Third . Elsevier B.V. , 2014.

[15] L. H. Holthuijsen, *Waves in Oceanic and Coastal Waters*. United States of America by Cambridge University Press, New York, 2007.

[16] E. F. Thomson, H. S. Chen, and L. L. Hadley, "Validation of Numerical Model for Wind Waves and Swell in Harbours," *J. Waterw. port, Coast. Ocean enginnering*, pp. 245–257, 1996.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MESHLESS METHOD FOR SIMULATION OF PARTICLE-DRIVEN GRAVITY CURRENTS WITH OBSTACLES

Karel Kovářík[1], **Juraj Mužík**[1], **and Dana Sitányiová**[1]

[1]University of Žilina, Department of Geotechnics
Univerzitna 8215/1 Žilina, Slovakia
e-mail: karel.kovarik, juraj.muzik, dana.sitanyiova@fstav.uniza.sk

Keywords: gravity current, LBIEM, obstacle, meshless method

Abstract. *A meshless local boundary integral equation method (LBIEM) is used to analyse gravity currents flow in two-dimensional domains. The method solves the incompressible Navier–Stokes equations with a transport equation for the particle concentration. The characteristic-based split scheme is used to solve the governing equations. The LBIEM basic equations are derived via interpolation using radial basis functions. Numerical test cases of the lock-exchange channel flow with obstacles in its bed are presented here.*

1 INTRODUCTION

Gravity currents increase when a heavier fluid propagates into a lighter one. This phenomena is frequently encountered in environment applications and geohazards (like solifluction, debris flow or turbidity currents). Particle-driven gravity currents create a special group of these flows when the difference in density is caused by the concentration of suspended particles. The particles may settle or become resuspended. For this reason these currents are known as non-conservative gravity currents [1]. Moreover, particles do not exactly follow the fluid flow. When the size of particles is small the velocity of particles can be expressed as a sum of the fluid velocity and a constant settling velocity [2].

2 GOVERNING EQUATIONS

The motion of the fluid phase is governed by the Navier-Stokes equations which can be written in its primitive variables as

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}\left(u_j u_i\right) = \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} - \frac{1}{\rho}\frac{\partial p}{\partial x_i} - \frac{C}{\rho} F_i$$

where u_i is the fluid velocity vector component, *p* is the pressure, ν is the kinematic viscosity, ρ is the density of a liquid, *C* is the particle-number density (i.e. the concentration), and F_i represents Stoke's drag which is the dominant flow force on an individual particle in direction *i* [3]. The particle field can be treated in an Eulerian manner (see [3]) and is governed by the following advection-diffusion equation

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x_j \partial x_j} - \frac{\partial}{\partial x_i}\left(u_{pi} C\right) \tag{2}$$

where *D* is the coefficient of diffusion and u_{pi} is the particle velocity in the direction *i*. The particle velocity can be considered with some simplification as (see also [4])

$$u_{pi} \approx u_i + u_{si} \tag{3}$$

where u_i is the *i*-th component of the fluid velocity, u_{si} is the *i*-th component of the settling velocity of particles. When the particles are assumed to be spherical and sufficiently small the settling velocity can be expressed using Stoke's settling velocity law as

$$u_{si} = \frac{d^2(\rho_p - \rho)}{18 \mu f} g_i \tag{4}$$

Here, *d* is the particle diameter, ρ_p is the particle density, *f* is the correction for non-Stokesian drag and g_i is the gravity vector component $\mathbf{g} = \{0, -g\}$ [1].

Equations (1) and (2) can be made dimensionless using transformations of coordinates and velocities

$$\tilde{x_i} = \frac{x_i}{L} \qquad \tilde{u}_i = \frac{u_i}{u_b} \tag{5}$$

where *L* is the characteristic length and u_b is the bulk velocity defined as [2]

$$u_b = \sqrt{\tilde{g} L} \qquad \tilde{g} = g \frac{\rho_1 - \rho_0}{\rho_0} \tag{6}$$

The dimensionless concentration and pressure can be defined as

$$\tilde{C} = \frac{\rho - \rho_0}{\rho_1 - \rho_0} \qquad \tilde{p} = \frac{p}{\rho_0 u_b^2} \tag{7}$$

The dimensionless governing equations can be written as

$$\begin{aligned} \frac{\partial \tilde{u}_i}{\partial x_i} &= 0 \\ \frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\tilde{u}_j \tilde{u}_i\right) &= -\frac{\partial \tilde{p}}{\partial x_i} + \frac{1}{\sqrt{Gr}} \frac{\partial^2 \tilde{u}_i}{\partial x_j \partial x_j} - \tilde{C} e_i^g \\ \frac{\partial \tilde{C}}{\partial t} &= \frac{1}{Sc\sqrt{Gr}} \frac{\partial^2 \tilde{C}}{\partial x_j \partial x_j} - \frac{\partial}{\partial x_i}\left(\tilde{u}_{pi} \tilde{C}\right) \end{aligned} \tag{8}$$

where $e_i^g = (0, -1)$ is the unity vector pointing in the direction of gravity. Three dimensionless parameters appear in (8), namely the Grashof number *Gr*, the Schmidt number *Sc*, and the dimensionless particle velocity $\tilde{u}_{pi}$ which are defined as [3]

$$Gr = \left(\frac{u_b L}{\nu}\right)^2 \qquad Sc = \frac{\nu}{D} \tag{9}$$
$$\tilde{u}_{pi} = \tilde{u}_i + \tilde{u}_{si}$$

Using a characteristic-based split (CBS) algorithm we get the following system of equations (details see e.g. in [5] [6])

$$\begin{aligned} \tilde{u}_i^* = \tilde{u}_i^n + \Delta t \Bigg[& \frac{1}{\sqrt{Gr}} \frac{\partial^2 \tilde{u}_i}{\partial x_j \partial x_j} - \tilde{C} e_i^g - \frac{\partial}{\partial x_j}\left(\tilde{u}_j \tilde{u}_i\right) + \\ & \frac{\Delta t}{2} \tilde{u}_k \frac{\partial}{\partial x_k}\left(\frac{\partial}{\partial x_j}\left(\tilde{u}_j \tilde{u}_i\right) + \tilde{C} e_i^g\right) \Bigg]^n \end{aligned} \tag{10}$$

$$\tilde{u}_i^{n+1} = \tilde{u}_i^* - \Delta t \frac{\partial \tilde{p}^{n+1}}{\partial x_i} \tag{11}$$

$$\frac{\partial^2 \tilde{p}^{n+1}}{\partial x_i \partial x_i} = \frac{1}{\Delta t} \frac{\partial \tilde{u}_i^*}{\partial x_i} \tag{12}$$

where we used the intermediate velocity $\tilde{u}_i^*$ which does not satisfied a continuity equation in (8). Upper indexes *n* and *n+1* indicate time steps and $\Delta t = t^{n+1} - t^n$ is the length of the time interval.

In the last step the dimensionless concentration $\tilde{C}$ can be also solved using the CBS algorithm (see [6])

$$\begin{aligned} \tilde{C}^{n+1} = \tilde{C}_i^n + \Delta t \Bigg[& \frac{1}{Sc\sqrt{Gr}} \frac{\partial^2 \tilde{C}^n}{\partial x_j \partial x_j} - \frac{\partial}{\partial x_j}\left(\tilde{u}_{pj}^{n+1} \tilde{C}^n\right) + \\ & \frac{\Delta t}{2} \tilde{u}_{pk}^{n+1} \frac{\partial}{\partial x_k}\left(\frac{\partial}{\partial x_j}\left(\tilde{u}_{pj}^{n+1} \tilde{C}^n\right)\right) \Bigg] \end{aligned} \tag{13}$$

3 LOCAL BOUNDARY INTEGRAL FORMULATION

The area of interest Ω with the boundary Γ is covered by points within the area and also on the global boundary (see Fig. 1). Consider a local circular sub-domain Ω_s with boundary Λ_s centered at every point *s*. This sub-domain is regular around all of the internal points, but at the points on the global boundary this local boundary consists of a part of the global boundary intersected with the local sub-domain Γ_s (see Fig. 2). To express the local boundary integral form of the governing equations developed in the previous section in a domain Ω_s, we apply the weighting residual principle [7] to equations (10), (11), (12), and (13) to obtain the weak form. If the test function w^* is chosen to be the fundamental solution of the Laplace equation, then after integration by parts twice the following integral equations can be obtained (see also [8])

$$\int_{\Omega_s} \tilde{u}_i^* w^* d\Omega = \int_{\Omega_s} \tilde{u}_i^n w^* d\Omega - \Delta t \left[\int_{\Omega_s} \tilde{u}_j \frac{\partial \tilde{u}_i}{\partial x_j} w^* d\Omega + \int_{\Omega_s} \tilde{C} e_i^g w^* d\Omega \right. \tag{14}$$
$$+ \frac{1}{\sqrt{Gr}} \left(c_s \tilde{u}_{is} + \int_{\Lambda_s \cup \Gamma_s} \frac{\partial w^*}{\partial n} \tilde{u}_i d\Gamma - \int_{\Lambda_s \cup \Gamma_s} w^* \frac{\partial \tilde{u}_i}{\partial n} d\Gamma \right)$$
$$\left. - \frac{\Delta t}{2} \int_{\Omega_s} \tilde{u}_k \frac{\partial}{\partial x_k} \left(\tilde{u}_j \frac{\partial \tilde{u}_i}{\partial x_j} + \tilde{C} e_i^g \right) w^* d\Omega \right]^n$$

$$c_s \tilde{p}_s^{n+1} + \int_{\Lambda_s \cup \Gamma_s} \frac{\partial w^*}{\partial n} \tilde{p}^{n+1} d\Gamma - \int_{\Lambda_s \cup \Gamma_s} w^* \frac{\partial \tilde{p}^{n+1}}{\partial n} d\Gamma + \frac{1}{\Delta t} \int_{\Omega_s} \frac{\partial \tilde{u}_i^*}{\partial x_i} w^* d\Omega = 0. \tag{15}$$

$$\int_{\Omega_s} \tilde{C}^{n+1} w^* d\Omega = \int_{\Omega_s} \tilde{C}^n w^* d\Omega - \Delta t \left[\int_{\Omega_s} \tilde{u}_j^{n+1} \frac{\partial \tilde{C}^n}{\partial x_j} w^* d\Omega \right. \tag{16}$$
$$+ \frac{1}{Sc\sqrt{Gr}} \left(c_s \tilde{C}_s^n + \int_{\Lambda_s \cup \Gamma_s} \frac{\partial w^*}{\partial n} \tilde{C}^n d\Gamma - \int_{\Lambda_s \cup \Gamma_s} w^* \frac{\partial \tilde{C}^n}{\partial n} d\Gamma \right)$$
$$\left. - \frac{\Delta t}{2} \int_{\Omega_s} \tilde{u}_k^{n+1} \frac{\partial}{\partial x_k} \left(\tilde{u}_j \frac{\partial \tilde{C}^n}{\partial x_j} \right) w^* d\Omega \right]$$

Here the coefficient c_s is equal to 1 for interior points and equal to $\frac{\theta}{2\pi}$ for points on the global boundary (see Fig. 2).

The boundary $\Lambda_s \cup \Gamma_s$ is then divided into several boundary elements (see Fig. 2). The values of velocity, pressure, and concentration in node *s* can now be computed from values in the local elements. The pressure equation (15) now becomes

$$c_s \tilde{p}_s + \int_{\Lambda_s \cup \Gamma_s} \frac{\partial w_{sj}^*}{\partial n} \tilde{p}_j d\Gamma - \int_{\Lambda_s \cup \Gamma_s} w_{sj}^* \frac{\partial \tilde{p}_j}{\partial n} d\Gamma - \int_{\Omega_s} w_{sj}^* \frac{\partial \tilde{u}_i^*}{\partial x_i} d\Omega = 0. \tag{17}$$

To avoid the term $\frac{\partial \tilde{p}_j}{\partial n}$ defined on the local boundary Λ_s, we use a companion solution (see e. g. [9]). This solution satisfies the Laplace equation and is equal to the fundamental solution on the local boundary. For the 2D Laplace equation the companion solution is defined as

$$w^c = \frac{1}{2\pi} log \left(\frac{1}{r_0} \right), \tag{18}$$

where r_0 is the radius of the local domain Ω_s. Equation (17) becomes

$$c_s \tilde{p}_s + \int_{\Lambda_s \cup \Gamma_s} \frac{\partial w_{sj}^*}{\partial n} \tilde{p}_j d\Gamma - \int_{\Gamma_s} \left(w_{sj}^* - w_{sj}^c \right) \frac{\partial \tilde{p}_j}{\partial n} d\Gamma - \int_{\Omega_s} \left(w_{sj}^* - w_{sj}^c \right) \frac{\partial \tilde{u}_i^*}{\partial x_i} d\Omega = 0. \tag{19}$$

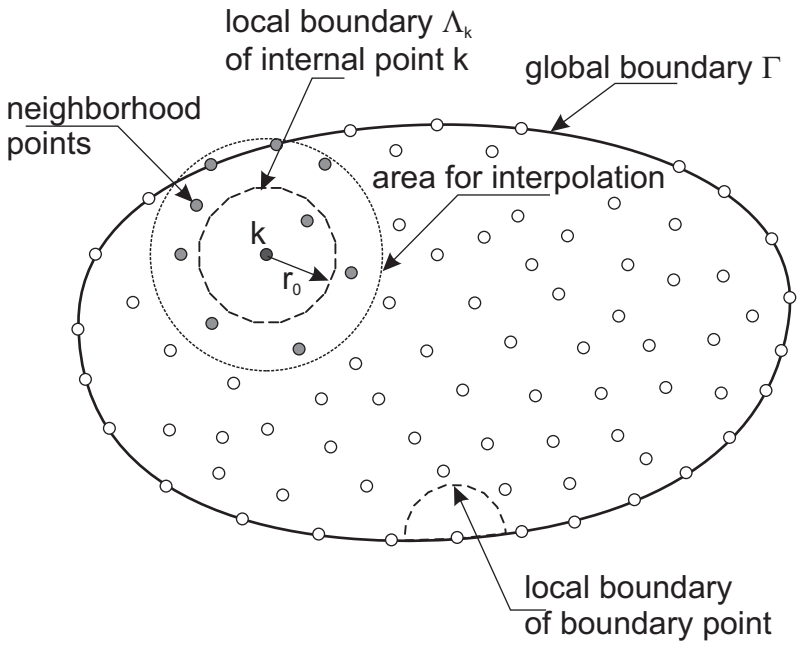


Figure 1: Points in the global area

Unknown values of the pressure and its derivatives in the local boundary element *j* can now be approximated by values of the pressure in *N* neighbouring points using RBF interpolation. The main reason for using RBF instead of the moving least squares (MLS) interpolation is that RBFs possess the delta property, making the prescription of the boundary conditions straightforward and simple. Multiquadric RBFs are used here

$$W(r_{ij}) = \sqrt{r_{ij}^2 + R^2} \tag{20}$$

where $W(r_{ij})$ is the multiquadric RBF between points *i* and *j* (see [10], [7]), r_{ij} is the Euclidean distance between these two points and *R* is the so-called shape factor of the multiquadric function. The formula of Hardy with a slight modification is applied to the local RBFs (see [10]) to find the optimal value of the shape factor, which can be computed in point *i* as

$$R_i = \frac{0.815}{N} \sum_{j=1}^{N} d_j \tag{21}$$

where d_j is the distance between the j-th local supporting point and its closest neighbouring point. A technique described in detail in [7] can be used to obtain a set of RBF shape functions Φ_{ij}, which can be used to express the pressure and its derivatives in the point *i* as

$$\tilde{p}_i = \sum_{j=1}^{N} \Phi_{ij} \tilde{p}_j \quad \frac{\partial \tilde{p}_i}{\partial x_k} = \sum_{j=1}^{N} \frac{\partial \Phi_{ij}}{\partial x_k} \tilde{p}_j \tag{22}$$

Using these shape functions causes equation (19) to become

$$c_s \tilde{p}_s + \sum_{m=1}^{N} \left(\int_{\Lambda_s \cup \Gamma_s} \frac{\partial w_{sj}^*}{\partial n} \Phi_{jm} d\Gamma \right) \tilde{p}_m - \sum_{m=1}^{N} \left[\int_{\Gamma_s} \left(w_{sj}^* - w_{sj}^c \right) \frac{\partial \Phi_{jm}}{\partial n} d\Gamma \right] \tilde{p}_m - \tag{23}$$
$$- \sum_{m=1}^{N} \left[\int_{\Omega_s} \left(w_{sj}^* - w_{sj}^c \right) \frac{\partial \Phi_{jm}}{\partial x_i} d\Omega \right] \tilde{u}_{im}^* = 0,$$

A similar procedure can be applied to equations (14) and (16). All of these equations can then be formulated in matrix form as

$$\sum_{m=1}^{N} M_{sm}\tilde{u}^{*}_{im} = \sum_{m=1}^{N} M_{sm}\tilde{u}^{n}_{im} - \Delta t \left[\sum_{m=1}^{N} S_{sm}\tilde{u}_{im} + e^{g}_{i} \sum_{m=1}^{N} M_{sm}\tilde{C}_{m} + \right. \tag{24}$$
$$+ \frac{1}{\sqrt{Gr}} \left(c_s \tilde{u}_{is} + \sum_{m=1}^{N} H_{sm}\tilde{u}_{im} - \sum_{m=1}^{N} G_{sm}\tilde{u}_{im} \right) -$$
$$\left. - \frac{\Delta t}{2} \sum_{m=1}^{N} \left(U_{sm}\tilde{u}_{im} + e^{g}_{i} F_{sm}\tilde{C}_{m} \right) \right]^{n},$$

$$\sum_{m=1}^{N} M_{sm}\tilde{u}^{n+1}_{im} = \sum_{m=1}^{N} M_{sm}\tilde{u}^{*}_{im} - \Delta t \sum_{m=1}^{N} S_{sm}\tilde{p}_{m} \tag{25}$$

$$c_s \tilde{p}_s + \sum_{m=1}^{N} H_{sm}\tilde{p}_{m} - \sum_{m=1}^{N} G_{sm}\tilde{p}_{m} - \sum_{m=1}^{N} S_{sm}\tilde{u}^{*}_{im} = 0, \tag{26}$$

$$\sum_{m=1}^{N} M_{sm}\tilde{C}^{n+1}_{m} = \sum_{m=1}^{N} M_{sm}\tilde{C}^{n}_{m} - \Delta t \left[\sum_{m=1}^{N} S_{sm}\tilde{C}_{m} + \right. \tag{27}$$
$$\left. + \frac{1}{Sc\sqrt{Gr}} \left(c_s \tilde{C}_s + \sum_{m=1}^{N} H_{sm}\tilde{C}_{m} - \sum_{m=1}^{N} G_{sm}\tilde{C}_{m} \right) - \frac{\Delta t}{2} \sum_{m=1}^{N} U_{sm}\tilde{C}_{m} \right]^{n},$$

where $\tilde{u}_{im}$, $\tilde{p}_m$, and $\tilde{C}_m$ are the values of the dimensionless velocity, pressure and concentration, respectively, in the points neighbouring the reference point *s*. The matrices M_{sm}, H_{sm}, G_{sm}, S_{sm}, and U_{sm} can be written as

$$M_{sm} = \sum_{j=1}^{n_e} \int_{\Omega_{sj}} \left(w^{*}_{sj} - w^{c}_{sj} \right) \Phi_{jm} d\Omega, \tag{28}$$

$$H_{sm} = \sum_{j=1}^{n_e} \int_{\Lambda_{sj} \cup \Gamma_{sj}} \frac{\partial w^{*}_{sj}}{\partial n} \Phi_{jm} d\Gamma,$$

$$G_{sm} = \sum_{j=1}^{n_e} \int_{\Gamma_{sj}} \left(w^{*}_{sj} - w^{c}_{sj} \right) \frac{\partial \Phi_{jm}}{\partial n} d\Gamma,$$

$$S_{sm} = \tilde{u}_{km} \sum_{j=1}^{n_e} \int_{\Omega_{sj}} \left(w^{*}_{sj} - w^{c}_{sj} \right) \frac{\partial \Phi_{jm}}{\partial x_k} d\Omega,$$

$$U_{sm} = \sum_{j=1}^{n_e} \int_{\Omega_{sj}} \frac{\partial \left(w^{*}_{sj} u_{im} \right)}{\partial x_i} \frac{\partial \left(u_{im} \Phi_{jm} \right)}{\partial x_i} d\Omega,$$

$$F_{sm} = \sum_{j=1}^{n_e} \int_{\Omega_{sj}} \frac{\partial \left(w^{*}_{sj} u_{im} \right)}{\partial x_i} \Phi_{jm} d\Omega$$

where n_e is the number of local elements, Λ_{sj} is the j-th internal local boundary element, Γ_{sj} is the j-th local boundary element situated on the intersection of the global boundary, and Ω_{sj} is the triangular area element (see Fig. 2). The mass matrix M_{sm} can be used in either a full or a lumped form. The lumped form is used here, because it eliminates a matrix inversion procedure.

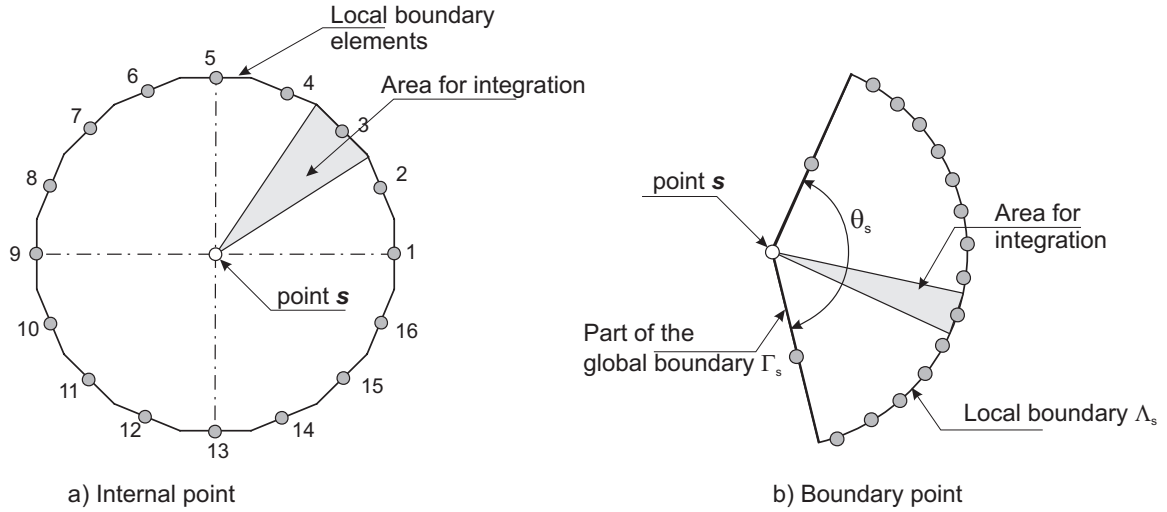


Figure 2: Local network around point

4 NUMERICAL EXAMPLES

All examples present so-called lock-exchange flow in rectangular channel when two fluids of different densities are divided by a vertical barrier at the initial time $t_0 = 0$. When this barrier is removed the denser liquid intrudes into the light one. This type of current is also frequently studied experimentally.

The schemes and boundary conditions of all these examples are presented in Fig.3.

4.1 Flow without any obstacles

The first example has a smooth bed without any obstacles and it serves as a comparing one. The rectangular computational domain of size $L = 20 \times H = 2$ is covered by the uniform grid of 800×80 points. The Grashof number $Gr = 5 \times 10^6$ is used for simulation, the Schmidt number is Sc=1 and the dimensionless settling velocity of particles is $\tilde{u}_s = 0.02$. Fig. 4 presents the results of the particle current at five dimensionless time instances. The flow is characterized by contours of particle concentration. Soon after beginning an intrusion, a front forms in the shape of a lifted nose due to the no-slip boundary condition on the bottom bed of the channel and Kelvin-Helmholtz vortices are formed at the upper part of the current.

4.2 Flow through the low obstacle

The second example presents the flow of particle current over a low rectangular obstacle (see Fig.3b). The dimensionless height of the obstacle is $d = 0.5$, i. e. $H/d = 4$. This obstacle is located in the dimensionless distance $L_s = 4.0$ from the beginning of a channel. The parameters of this example remain the same as in the first one. Fig. 5 presents contours of the particle concentration. Approximately at time $\tilde{t} = 6$ the current hits the upstream part of the obstacle and some deformation of the current starts. The whole current is divided into parts during the next time steps. The first part forms the backward current which moves back to the left part of the channel; the second part overcome the obstacle and continues moving in the right part of the channel.

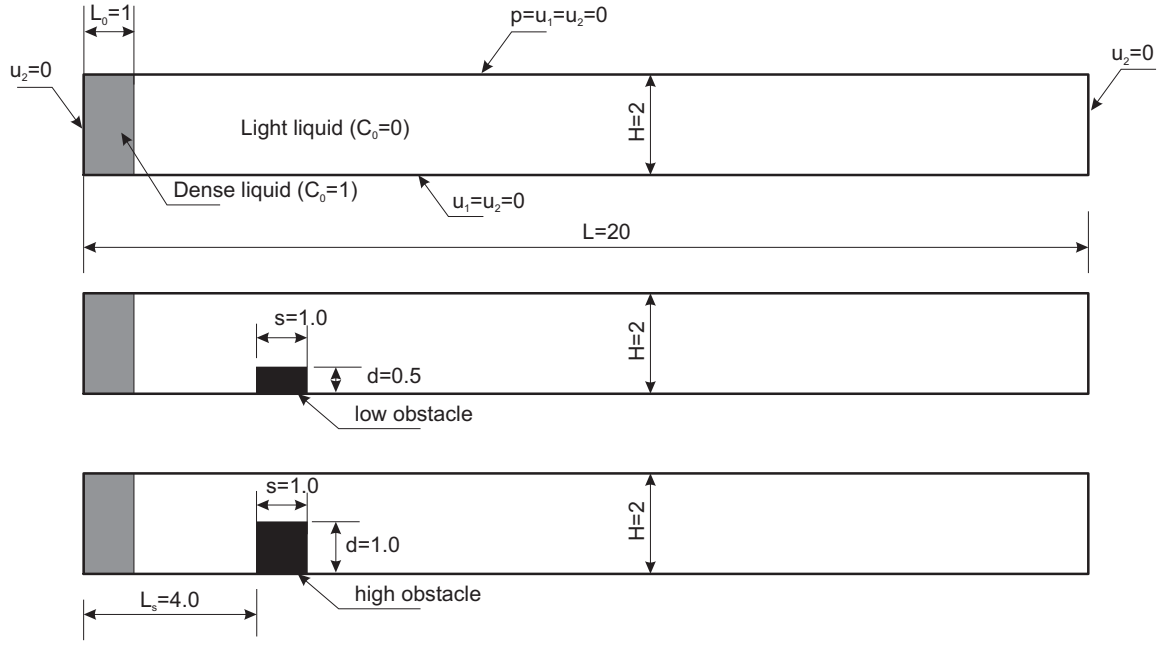


Figure 3: Numerical examples, geometry and boundary conditions

4.3 Flow through the high obstacle

The third example presents the flow over a high rectangular obstacle (see Fig.3c). The dimensionless height of the obstacle is now $d = 1.0$, i. e. $H/d = 2$. The location of the obstacle and parameters of flow remain the same as in the previous example. Fig. 6 presents contours of concentration. As can be seen in Fig.6 the deformation of the particle current is now a bit different than in the previous example. Almost all particles form now the backward wave and only the negligible amount of particles overcome the obstacle. The newly formed backward current is now more accentuated.

The shear stress distribution at the bottom of the channel is a very important part of the resuspension of particles. The dimensionless shear stress $\tilde{\tau}_w$ can be described as [1, 3, 6]

$$\tilde{\tau}_w = \frac{1}{\sqrt{Gr}} \sqrt{\left(\frac{\partial u_{p1}}{\partial x_2}\right)^2 + \left(\frac{\partial u_{p2}}{\partial x_2}\right)^2} \tag{29}$$

Fig. 7 presents the dimensionless shear stress at the bottom boundary at time intervals $\tilde{t}$ =4, 6, and 10. This shear stress shows several peaks that correspond to the front of the current and also to the Kelvin-Helmholtz vortices. The low values of the shear stress reflect mentioned decay of current in the case of low and high obstacles at time $\tilde{t} = 10$ (see Fig.7b,c).

The cumulative sedimentation of particles can be computed as (see [1])

$$D(\tilde{x}, \tilde{t}) = \frac{1}{L_0 H} \int_0^{\tilde{t}} \tilde{C}_w(\tilde{x}, \hat{t}) u_s d\hat{t} \tag{30}$$

Fig. 8 shows the value of this deposit as a function of the longitudinal coordinate x at the dimensionless time $\tilde{t}$=60. The cumulative sedimentation is concentrated in the upstream part of the channel in case of low and especially high obstacles.

Fig. 9 presents temporal variations of the mean dimensionless pressure $\tilde{p}$ (7) acting on the upstream and downstream faces of the obstacles. The maximum impact is localized at time $\tilde{t} = 6$.

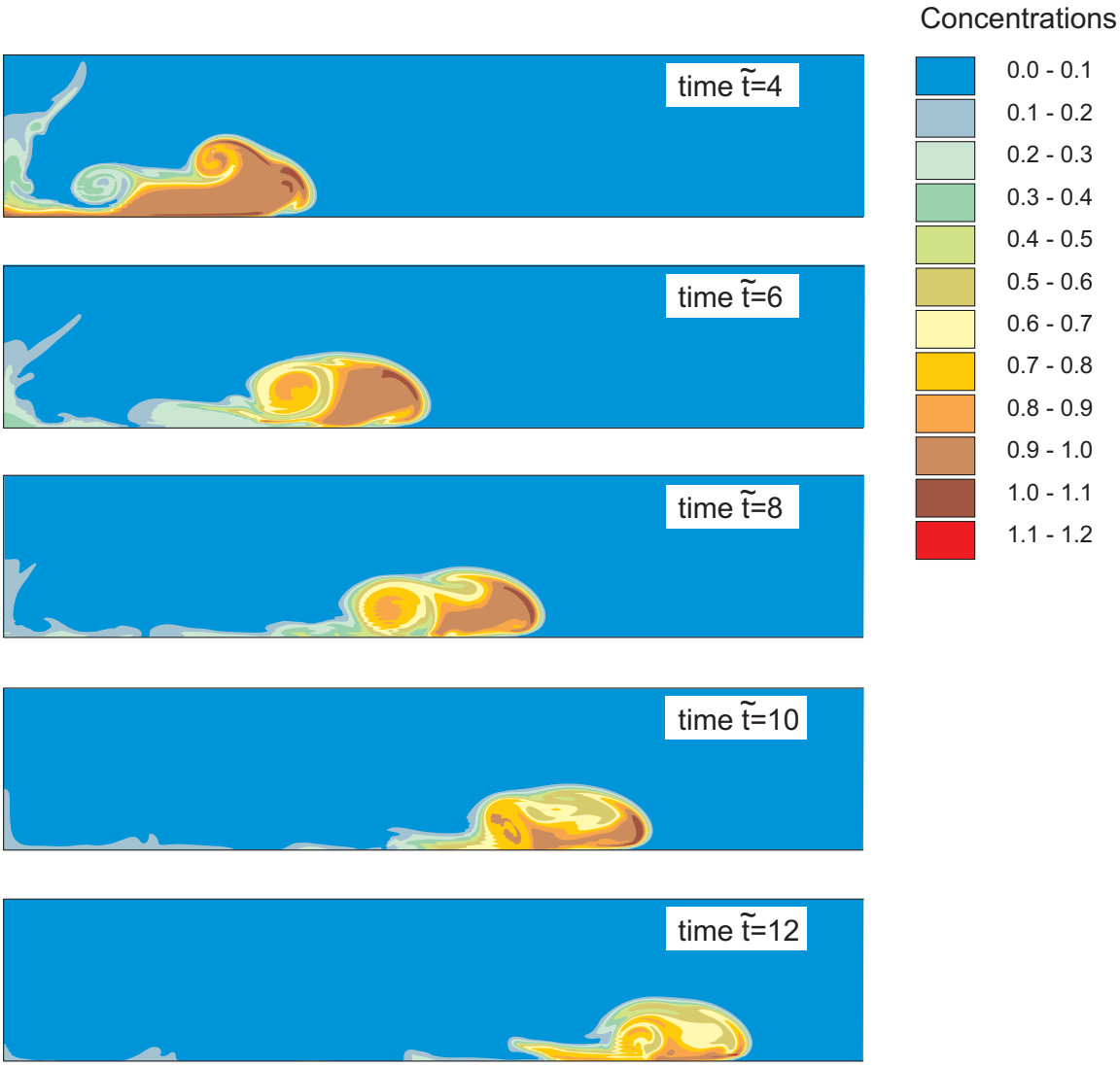


Figure 4: Dimensionless concentrations in the channel without any obstacles

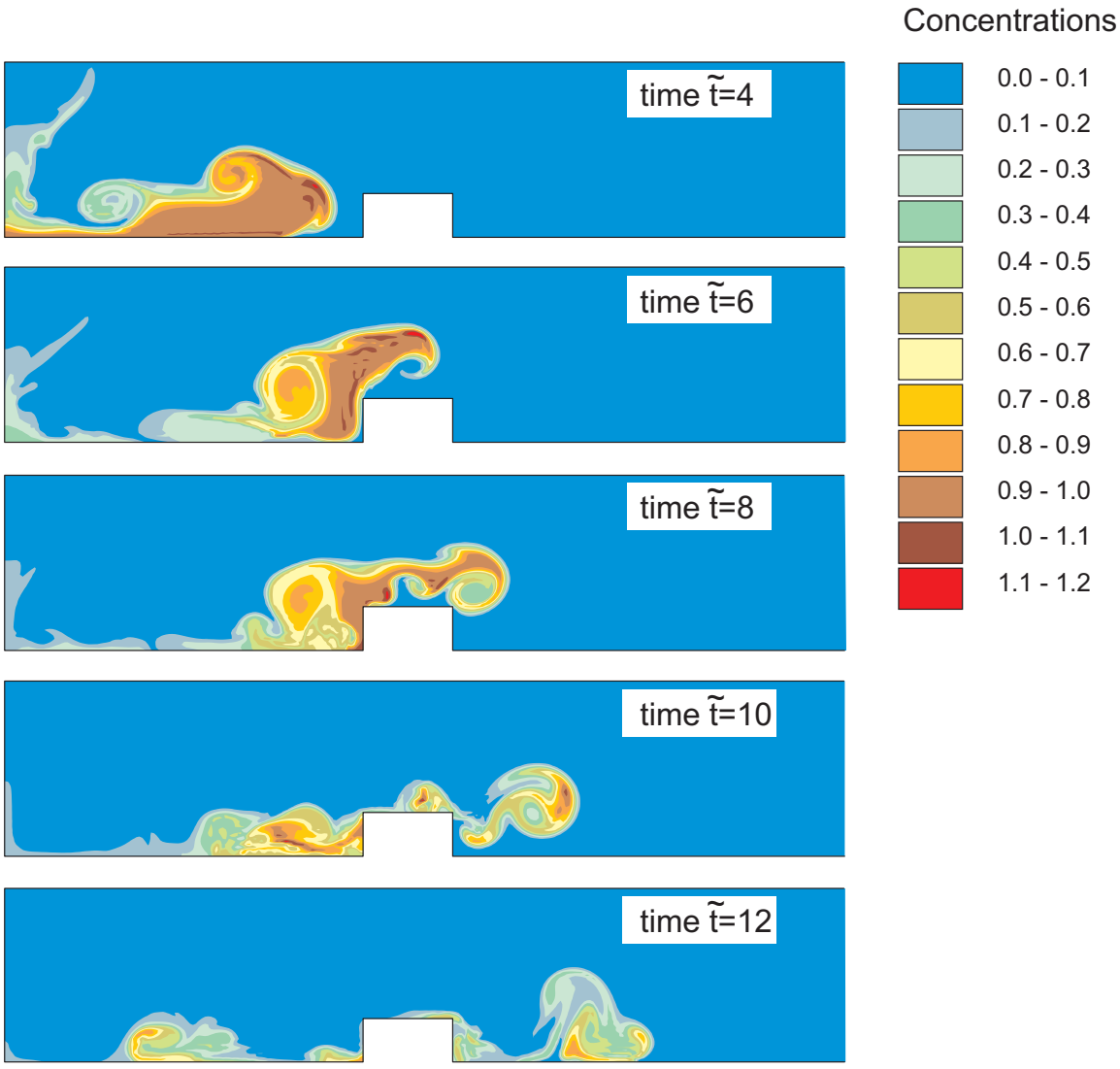


Figure 5: Dimensionless concentrations in the channel with the low obstacle

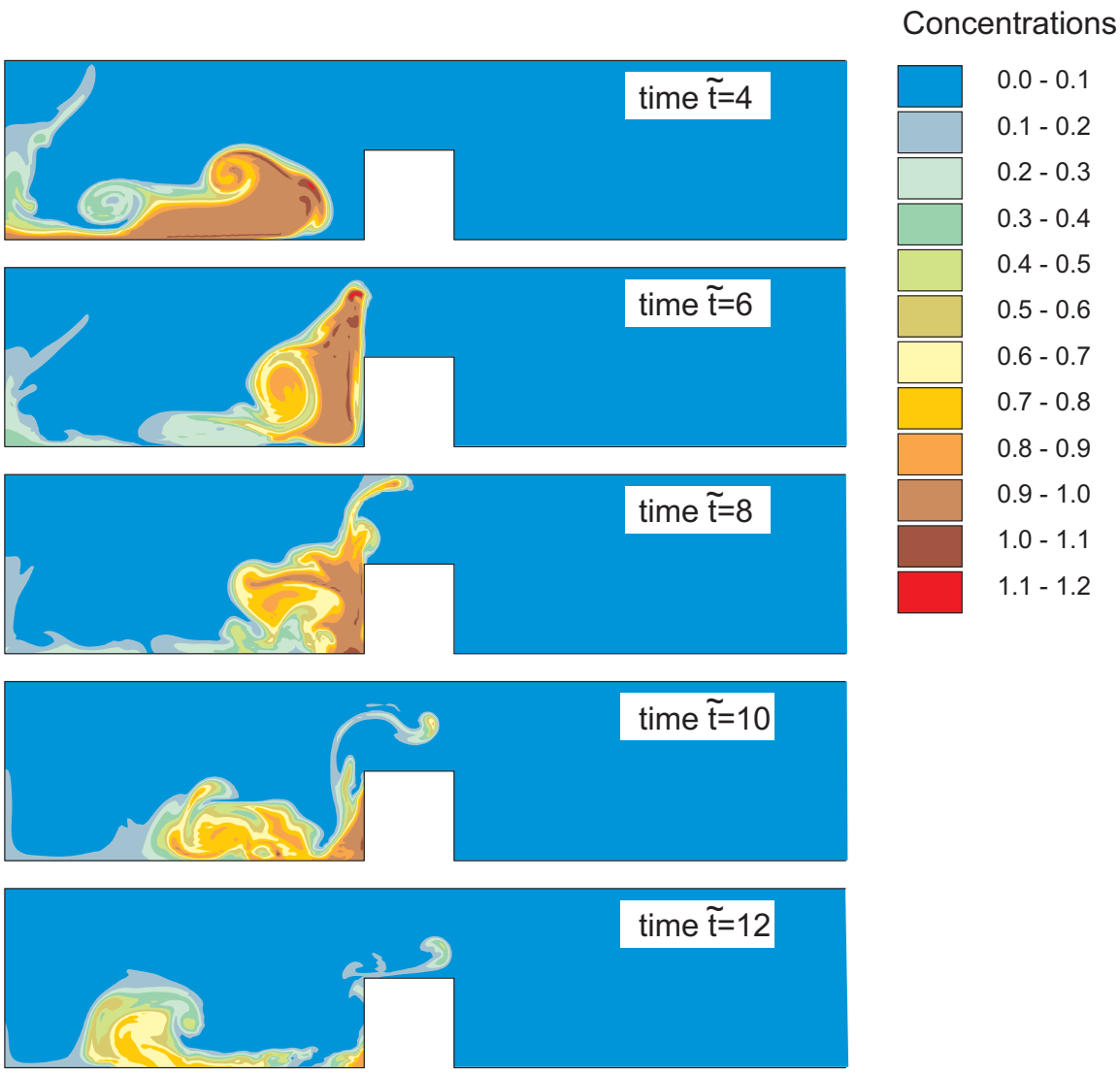


Figure 6: Dimensionless concentrations in the channel with the low obstacle

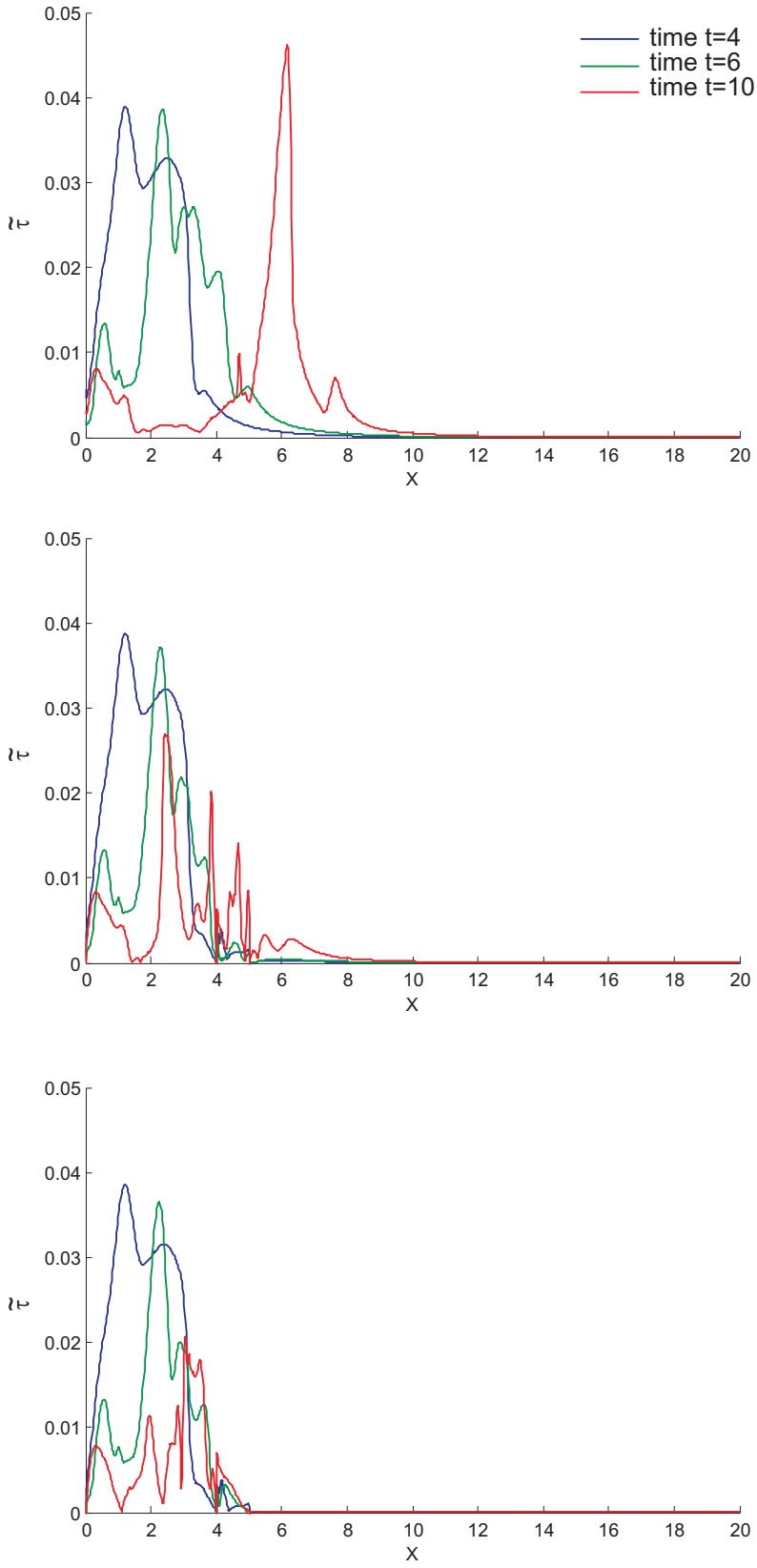


Figure 7: Dimensionless bed shear stress; a–without obstacle, b–low obstacle, c–high obstacle

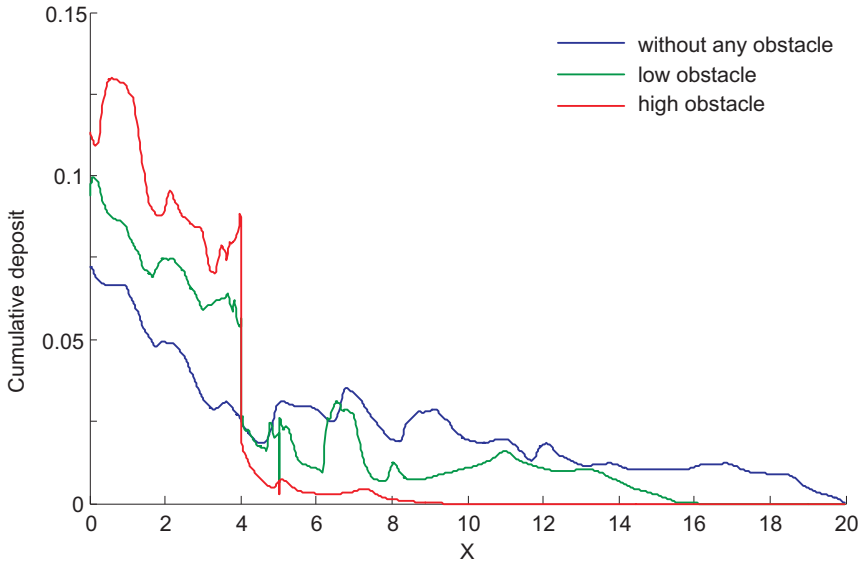


Figure 8: Dimensionless cumulative particle deposit at time $\tilde{t}$=60; a–without obstacle, b–low obstacle, c–high obstacle

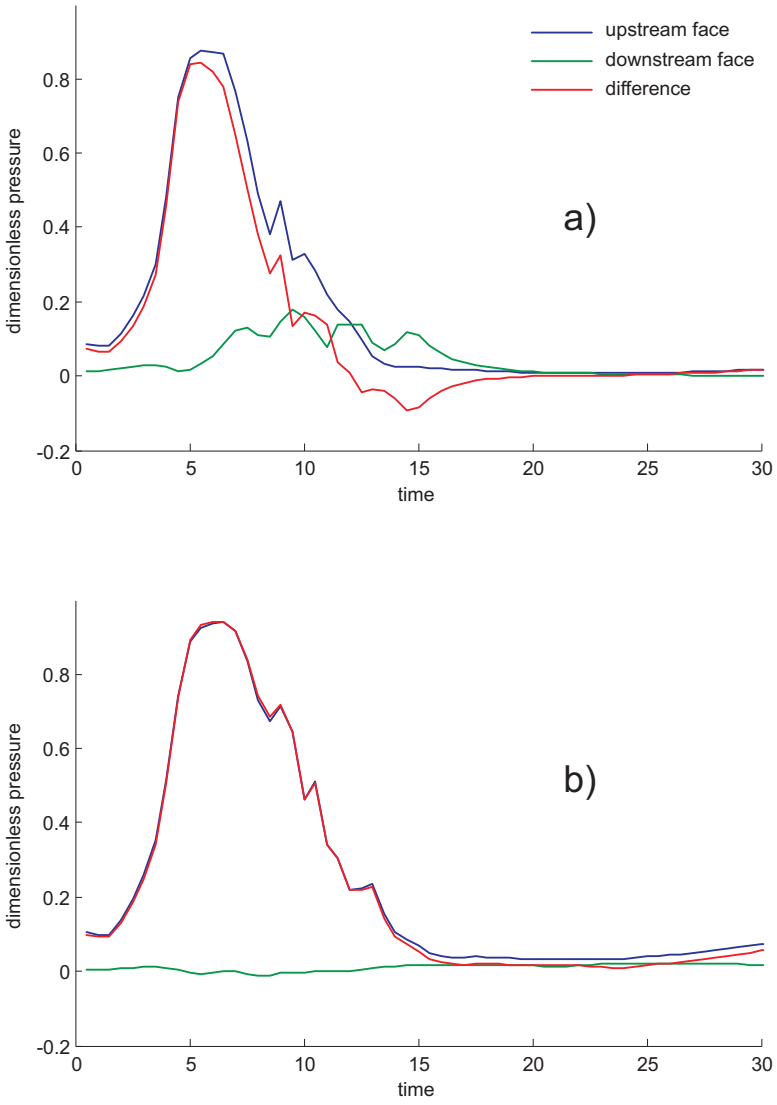


Figure 9: Drag acting on the obstacles; a–low obstacle, b–high obstacle

5 CONCLUSIONS

This paper presents a possible use of the LBIEM meshless method for solution of the gravity currents flow. We use the characteristic based split scheme to achieve stable and accurate results. The results obtained from test cases proved that the method is an effective and useful tool to solve this type of flow and can be used as an effective tool for modelling of several types of geohazards.

ACKNOWLEDGEMENT

This contribution was supported by PARAGEO project, co-funded by the EEA Grants and the state budget of the Slovak Republic from the EEA Scholarship Program Slovakia under the contract EHP-30/SK06-IV-02-007/2015.

REFERENCES

[1] M. I. Cantero, S. Balachandar, M. H. Garcia, An EulerianEulerian model for gravity currents driven by inertial particles, *International Journal of Multiphase Flow*, **34**, 484-501, 2008.

[2] C. Härtel, E. Meiburg, F. Necker, Analysis and direct numerical simulation of the flow at a gravity-current head. Part 1. Flow topology and front speed for slip and no-slip boundaries. *Journal of Fluid Mechanics*, **418**, 189-212, 2000.

[3] F. Necker, C. Härtel, L. Kleiser, E. Meiburg, High-resolution simulations of particle-driven gravity currents, *International Journal of Multiphase Flow* **28** 279–300, 2002.

[4] L. Espath, L. Pinto, S. Laizet, J. Silvestrini, Two- and three-dimensional Direct Numerical Simulation of particle-laden gravity currents, *Computers & Geosciences* **63** 9–16, 2014.

[5] K. Kovářík, J. Mužík, D. Sitányiová, A fractional step local boundary integral element method for unsteady two-dimensional incompressible flow, *Engineering Analysis with Boundary Elements* **44** 55–63, 2014.

[6] K. Kovářík, J. Mužík, S. Masarovičová, D. Sitányiová, A local boundary integral method for two-dimensional particle-driven gravity currents simulation, *Engineering Analysis with Boundary Elements* **56** 119–128, 2015.

[7] K. Kovářík, J. Mužík, M. S. Mahmood, A meshless solution of two dimensional unsteady flow, *Engineering Analysis with Boundary Elements* **36** 738–743, 2012.

[8] S. N. Atluri, H.-G. Kim, J. Y. Cho, A critical assessment of the truly Meshless Local Petrov-Galerkin (MLPG), and Local Boundary Integral Equation (LBIE) methods, *Computational Mechanics* **24** 348–372, 1999.

[9] V. Sladek, J. Sladek, S. N. Atluri, R. Van Keer, Numerical integration of singularities in meshless implementation of local boundary integral equations, *Computational Mechanics* **25** 394–403, 2000.

[10] E. J. Sellountos, A. Sequeira, An advanced meshless LBIE/RBF method for solving two-dimensional incompressible fluid flows, *Computational Mechanics* **41** (2007) 617–631.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

APPLICATION OF THE SHALLOW WATER EQUATIONS TO REAL FLOODING CASE

H. Ali[1], P-Y. Lagrée[2], and J-M. Fullana[2]

[1]AXA Global P&C, 75008 Paris, France
e-mail: hani.ali@axa.com

[2] Institut Jean le Rond d'Alembert, CNRS & UPMC Université Paris 06, UMR 7190, 75005 Paris, France
e-mail: {pierre-yves.lagree,jose.fullana}@upmc.fr

Keywords: Shallow-Water equations, Computational Methods, Engineering Sciences.

Abstract. *In this paper, we solve and validate the Shallow Water equations with a well-balanced finite volume method and semi-implicit treatment of the friction term. The method is validated by simulating the Seine river flooding (France 1910), which is served as a benchmark. The method is implemented in the open source software Basilisk which is freely available at* `http://www.basilisk.fr`*. Furthermore, we compare the obtained result with Basilisk to a simple propagation model developed by the actuary team at AXA Global P&C.*

1 INTRODUCTION

The Shallow Water equations (or Saint-Venant Equations) are commonly used to represent the flow of water under the assumption that the characteristic vertical length scale of the flow is very small with respect to the characteristic horizontal length scale. These equations can be obtained by depth averaging the incompressible three dimensional Navier-Stokes equations under the previous assumption in the vertical direction. This results in neglecting any acceleration in the direction of the gravity, which gives rise to a hydrostatic pressure distribution.

The numerical resolution of the Shallow Water equations is a good tool to predict flow properties in various situation such as Dam Breaks [15, 5], Tsunami propagation [10], Rivers flows [2] and flooding [3, 6], among others as there is a huge literature on the Shallow Water equations.

The Shallow Water equations are given by the following system:

$$\begin{aligned} \partial_t h + \partial_x(hu) + \partial_y(hv) &= 0, \\ \partial_t(hu) + \partial_x(hu^2 + gh^2/2) + \partial_y(huv) &= gh(S_{0_x} - S_{f_x}), \\ \partial_t(hv) + \partial_x(huv) + \partial_x(hv^2 + gh^2/2) &= gh(S_{0_y} - S_{f_y}). \end{aligned} \tag{1}$$

Here:

- $u(x, y, t)$ and $v(x, y, t)$ are the scalar components in the horizontal direction (x, y) of the depth-averaged velocity for a given time $t > 0$,
- $h(x, y, t)$ is the local water depth,
- $g = 9.81m/s^2$ is the gravity constant,
- $\mathbf{S}_0 = \left(S_{0_x}, S_{0_y}\right) = (-\partial_x z(x, y), -\partial_y z(x, y))$ is the linearized bed slope, where $z(x, y)$ is the topography,
- and $\mathbf{S}_f = \left(S_{f_x}, S_{f_y}\right) = \left(n^2 \frac{\sqrt{u^2+v^2}}{h^{4/3}} u, n^2 \frac{\sqrt{u^2+v^2}}{h^{4/3}} v\right)$, is the friction force, where n is the Manning's coefficient which depends on the considered type of soil (see [4]).

In this paper we aim to study an application of the Shallow Water equations to a real life flooding case. The application consists in simulating the flooding of the Seine river in Paris which occurred in January 1910. Predicting this flooding represents a challenge for public policy. In fact, the damage that could be caused by a potential flood having the same level as in 1910 is considerable, especially in Paris where the main activity pool of the French economy is concentrated.

The first difficulty before running the simulation with the SWE is the choice of the numerical scheme. Referring to the work of [1] see also [6], we consider a hydrostatic reconstruction scheme coupled with a semi implicit treatment of the friction term to avoid spurious flows due to topography. Second, we have to consider the appropriate topography that is relevant to the physical one. Given that we are limited by the computer power and by the availability of such topography, we choose to use a free available topography from the National Institute of Geographic and Forest Information (IGN) where a mesh of 75 m is used. First, we will use the topography in its original form. Second, we will modify the topography in order to take into account the so called barriers at the main bed of the river. We will use in all these simulations the open source code Basilisk which is a C code based on a finite volume method

with a structured mesh. This code uses also a quadtrees mesh to allow efficient adaptive grid refinement. Methods based on quadtrees mesh have been successfully used to compute adaptive solutions of many problems such as the 2D and the 3D Navier-Stokes equations [13, 12] or Shallow Water equations [14].

The rest of this paper is organized as follows: In section 2, we set up the formulation of the model equations and recall the numerical scheme. In section 3, we reproduce the 1910 flooding in Paris by using the historical registration of the water depth. We superimpose some obtained footprints with some maps from [8] in order to visualize clearly the inundated regions. Furthermore, we compare our result against the result that is furnished by the Regional and Interdepartmental Directorate of Environment and Energy (DRIEE) in order to validate our method. In section 4, we study the sensitivity of the model with respect to the Manning's coefficient. In section 5, we compare the result furnished by the resolution of the SWE with the result obtained by a simple propagation model developed by the actuary team at AXA Global P&C. Section 6 contains the conclusion of this paper.

2 Numerical method

2.1 General form

The system 1 can be written in the conservative form as

$$\partial_t \mathbf{U} + \nabla \left(\mathbf{F}, \mathbf{G}\right) \quad = \mathbf{R}, \tag{2}$$

where

$$\begin{aligned} \mathbf{U} = \left(h, hu, hv\right), \quad \mathbf{F} &= \left(hu, hu^2 + gh^2/2, huv\right), \\ \mathbf{G} &= \left(hv, huv, hv^2 + gh^2/2\right), \\ \text{and } \mathbf{R} &= \left(0, gh(S_{0_x} - S_{f_x}), gh(S_{0_y} - S_{f_y})\right). \end{aligned} \tag{3}$$

Here, $\mathbf{U}$ is the vector of conserved variables, $\mathbf{F}$ and $\mathbf{G}$ are the flux functions, and $\mathbf{R}$ stands for the sources which are formed from the bed slope $gh(\mathbf{S}_0)$ term and the friction term $gh(\mathbf{S}_f)$. We note that for $h > 0$ the above system is strictly hyperbolic.
As the 2D case can be treated by a complete analogy with the 1D case, for seek to simplicity, we will consider only the 1D case. The two-dimensional system (2) rewrite

$$\partial_t \mathbf{U} + \partial_x \mathbf{F} \quad = \mathbf{R}. \tag{4}$$

Hereafter,

$$\mathbf{U} = \left(h, hu\right), \quad \mathbf{F} = \left(hu, hu^2 + gh^2/2\right), \text{ and } \mathbf{R} = \left(0, gh(S_{0_x} - S_{f_x})\right). \tag{5}$$

We split in two problems, first an inviscid one, second a purely viscous one.

2.2 Inviscid part treatment

Following [1], we use the hydrostatic reconstruction scheme to the problem (4), without the friction term, which is given by

$$\mathbf{U}_i^* = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}_{i+1/2}^n - \mathbf{F}_{i-1/2}^n\right) + \frac{\Delta t}{\Delta x}{\mathbf{S}_0}_i^n, \tag{6}$$

where

$$\mathbf{F}^n_{i+1/2} = \mathcal{F}\left(\mathbf{U}^n_{i+1/2,L}, \mathbf{U}^n_{i+1/2,R},\right), \quad \mathbf{F}^n_{i-1/2} = \mathcal{F}\left(\mathbf{U}^n_{i-1/2,L}, \mathbf{U}^n_{i-1/2,R}\right), \tag{7}$$

and

$$\mathbf{S}_{0i}^{\ n} = \left(0, g(h^n_{i+1/2,L})^2/2 - g(h^n_{i-1/2,R})^2/2\right). \tag{8}$$

Here $\mathcal{F}$ stands to the numerical flux which is defined in the next section and the values of $\mathbf{U}^n_{i+1/2,L}$ and $\mathbf{U}^n_{i+1/2,R}$ are obtained by applying once again the hydrostatic reconstruction.

$$\mathbf{U}^n_{i+1/2,L} = \left(h^n_{i+1/2,L}, h^n_{i+1/2,L}u^n_i\right), \quad \mathbf{U}^n_{i+1/2,R} = \left(h^n_{i+1/2,R}, h^n_{i+1/2,R}u^n_{i+1}\right) \tag{9}$$

where

$$\begin{aligned} h^n_{i+1/2,L} &= max\left(0, h^n_i + z^n_i - max\left(z^n_i, z^n_{i+1}\right)\right), \\ h^n_{i+1/2,R} &= max\left(0, h^n_{i+1} + z^n_{i+1} - max\left(z^n_i, z^n_{i+1}\right)\right). \end{aligned} \tag{10}$$

We note that this scheme is designed to capture the dry regions where $h = 0$ and to ensure the non-negativity of the water weight, we refer the reader to [1] for more details.

2.3 Numerical flux

In Basilisk many consistent flows were implemented for the homogeneous part of the system 2, for example the Harten Lax Van Leer (HLL) flux, the Kurganov flux and the one obtained by the kinetic method. In this work we used the Kurganov flux. The Kurganov flux [11] is based on a central-upwind scheme which does not require exact or approximate Riemann solvers and a characteristic analysis that upwind Godunov-type methods require. Furthermore, they admit a less numerical dissipation than the other numerical flux.

The application of this scheme gives the following expression for the numerical flux:

$$\begin{aligned} \mathcal{F}\left(\mathbf{U}^n_{i+1/2,L}, \mathbf{U}^n_{i+1/2,R},\right) \quad &= \frac{a_{i+1/2,R}\mathbf{F}(\mathbf{U}^n_{i+1/2,L}) - a_{i+1/2,L}\mathbf{F}(\mathbf{U}^n_{i+1/2,R})}{a_{i+1/2,R} - a_{i+1/2,L}} \\ &+ \frac{a_{i+1/2,R}a_{i+1/2,L}}{a_{i+1/2,R} - a_{i+1/2,L}}\left(\mathbf{U}^n_{i+1/2,L} - \mathbf{U}^n_{i+1/2,R}\right), \end{aligned} \tag{11}$$

where

$$\begin{aligned} a_{i+1/2,R} &= max\left(0, u^n_{i+1/2,R} + \sqrt{gh^n_{i+1/2,R}}, u^n_{i+1/2,L} + \sqrt{gh^n_{i+1/2,L}}\right), \\ a_{i+1/2,L} &= min\left(0, u^n_{i+1/2,R} - \sqrt{gh^n_{i+1/2,R}}, u^n_{i+1/2,L} - \sqrt{gh^n_{i+1/2,L}}\right). \end{aligned} \tag{12}$$

The above scheme is stable under the following CFL condition on the time step Δt:

$$\Delta t \leq CFL \frac{\Delta x}{max(a_{i+1/2,R}, -a_{i+1/2,L})}, \tag{13}$$

where $CFL < 1$ is the Courant number and Δx is the space step.

2.4 Friction term treatment

The fractional-step method is a standard way to deal with the friction term. The complete solution of problem (2) is obtained by first solving the equations without the friction term as in section 2.1 to obtain an intermediate solution $\mathbf{U}^*$. The following problem is then solved using $\mathbf{U}^*$ to find the complete solution $\mathbf{U}^{n+1}$:

$$\partial_t \mathbf{U} = \mathbf{S}_f. \tag{14}$$

Thus, we obtain the following semi-implicit scheme for the friction term

$$h^{n+1} = h^* \quad \text{and } u^{n+1} = \frac{u^*}{1 + g\Delta t n^2 \frac{|u^n|}{(h^{n+1})^{4/3}}}. \tag{15}$$

We note that this semi-implicit scheme preserves the steady state at rest and also the stability under the CFL condition.

3 Numerical simulation: application to Seine inundation

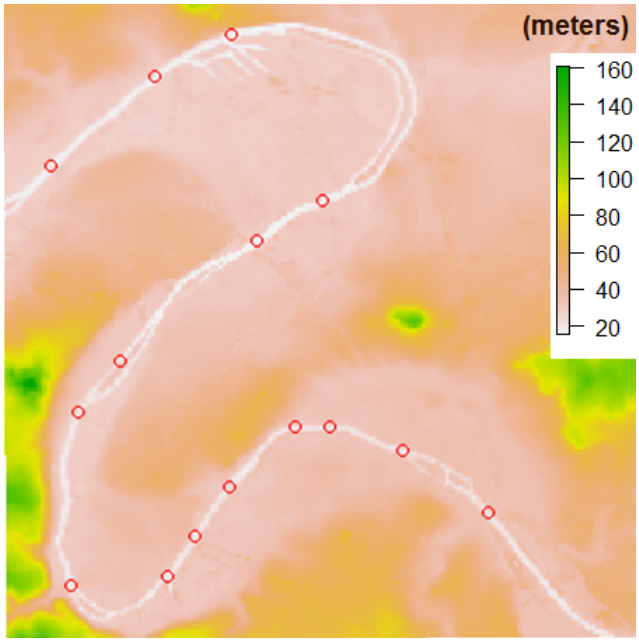


Figure 1: Topography generated from the IGN data showing the domain of the study with 75 m mesh resolution, 23.6 $m \leq z \leq 161$ m. Locations of different gauges are given.

The Seine river in Paris is our case study. Figure 1 represents the topography of the river in the landscape: the upstream end is located near Austerlitz train station and the downstream end is located near Chatou train station. The domain area is about 225km^2.

We will distinguish between two cases: First, we run the simulation by using the topography that is furnished by the IGN ([9]). Second, we modify the topography in order to take into account the barriers along the river which are constructed to prevent flooding (see figure 2).

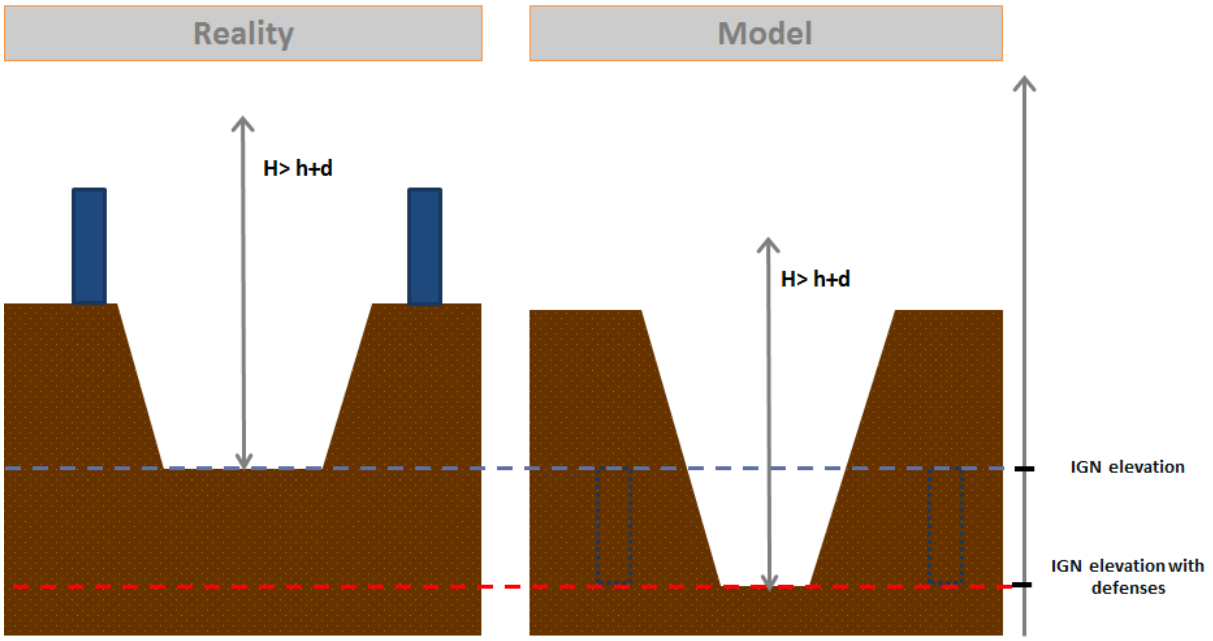


Figure 2: Figure showing the modified topography

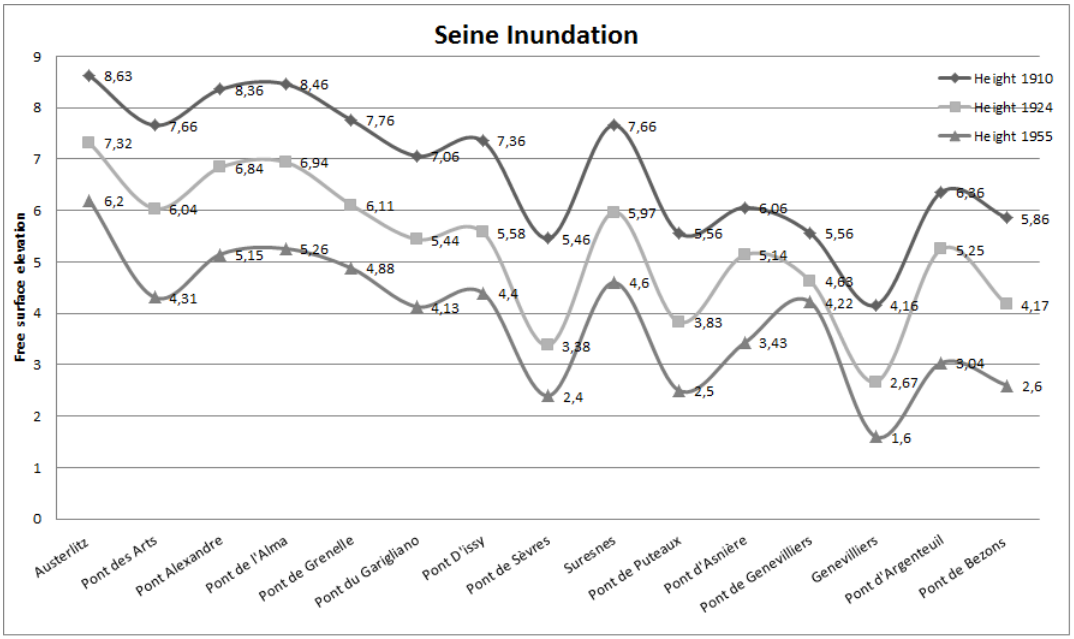


Figure 3: Free elevation final height at several places in Paris, computed values

3.1 Numerical resolution and the raw IGN data

The numerical solution has been computed by using the SWE described in section 2, where the topography is given by the IGN (see figure 1). The Manning's coefficient n is set to 0.01 $s/m^{\frac{1}{3}}$ for two reasons: First, in order to accelerate the propagation of the water. Second, in order to represent the minor and major bed of the Seine river at Paris as a firm soil.



Figure 4: A figure generated from simulation at $t = 120000s$ showing the inundated area in 1910

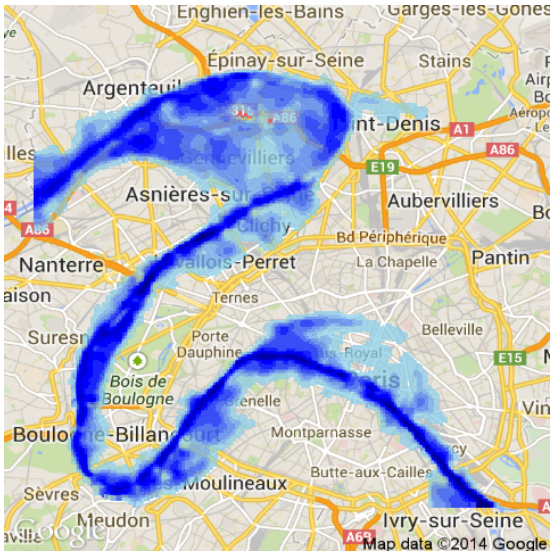


Figure 5: A figure generated from simulation at $t = 120000s$ showing the inundated area in 1924

Figure 3 shows the evolution of the water depth in meters between Austerlitz and the Bridge of Bezons. This figure gives the level of the free surface, i.e. the depth of the water, at each station.

We took as a reference the historical depth of the water in Austerlitz in 1910, 1924 and 1982. Within the used IGN topography, the altitude of Austerlitz is about 27 m from the zero level which is the NGF69 level. Thus, we take into account this information in order to reproduce the historical events. The numerical results at the final stage after 12h of propagation is given in figures 4, 5 and 6. These figures allow us to determine the inundated regions in Paris in 1910, 1924 and 1982.



Figure 6: A figure generated from simulation at $t = 120000s$ showing the inundated area in 1982



Figure 7: A figure generated from simulation at $t = 120000s$ showing a comparison between our simulation (blue) and the result of the DRIEE (red).

3.2 Simulation with defenses

Next, we will modify the topography in order to take into account the barriers along the river. These barriers reach the value of 7m in Austerlitz. Here, we compare our simulation with the modified topography that we call simulation with defenses against the result that are furnished by the DRIEE [7]. Figure 7 shows that our result is consistent with the DRIEE result. In order to calculate the prediction rate we calculate the ratio between the inundated area given by our simulation and the inundated area given by the DRIEE. We deduce that our result is similar to the result of the DRIEE with a prediction rate of 90 percent. The regions that are potentially inundated can be detected from this figure.

4 Sensitivity to Manning's coefficient

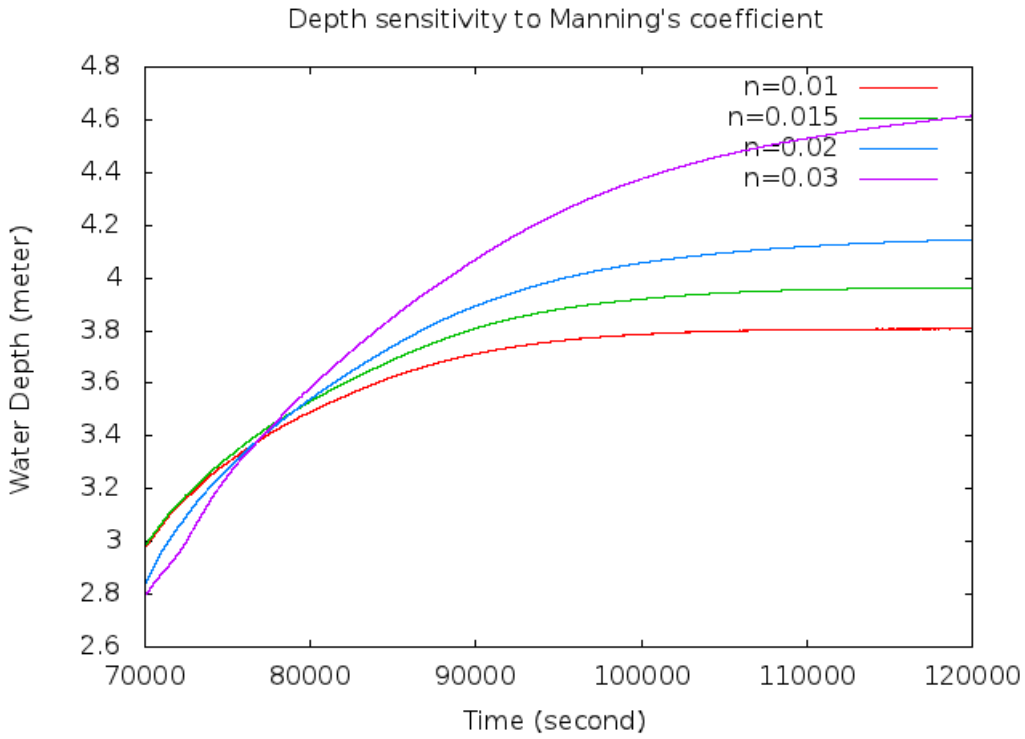


Figure 8: Variations of the water depth with respect to the manning's coefficient n. The gauge is located near Issy bridge in Paris

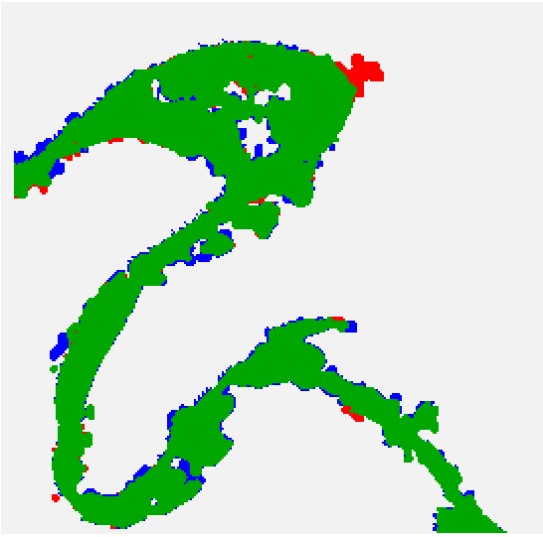


Figure 9: A figure generated from simulation at $t = 120000s$ showing a comparison between the inundated area in 1982 with $n = 0.01$ (blue) and with $n = 0.015$ (red) and their intersection (green)

The Manning's coefficient n is the most important parameter that should be determined for hydrodynamics models. In fact, selecting the value of n means to estimate the flow resistance

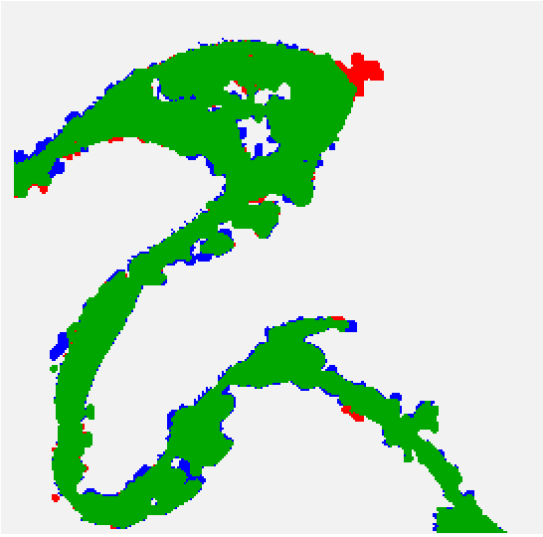


Figure 10: A figure generated from simulation at $t = 120000s$ showing a comparison between the inundated area in 1982 with $n = 0.01$ (blue) and with $n = 0.02$ (red) and their intersection (green)

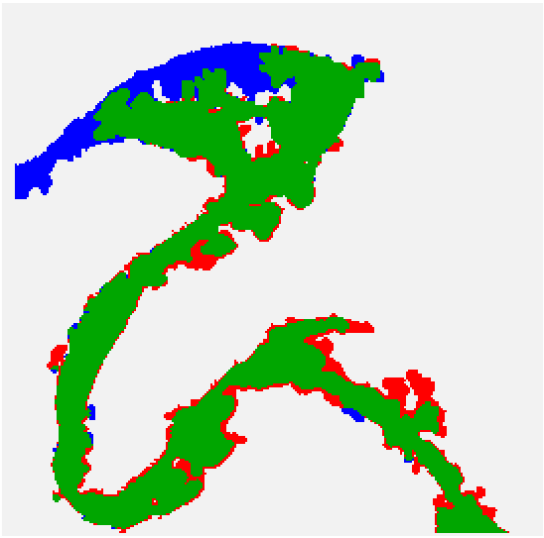


Figure 11: A figure generated from simulation at $t = 120000s$ showing a comparison between the inundated area in 1982 with $n = 0.01$ (blue) and with $n = 0.03$ (red) and their intersection (green)

to the movement. Moreover, this coefficient greatly impacts the flow speed, which then impacts the depth and the flood extent. In order to test the sensitivity of the model with respect to the Manning's coefficient, we run the simulation for four values of n 0.01, 0.015, 0.02, and 0.03. These values seem reasonable for a river channel with a firm soil bed and no vegetation.

Figure 8 shows that when n is bigger than 0.01 the flow is slower and the water can be more accumulated. A comparison between the footprint of the 1982 flood in Paris with $n = 0.01$ and $n = 0.015$, $n = 0.02$, and $n = 0.03$ respectively at the same time are given in figures 9, 10 and 10. These figures show that the new inundated region will be bigger than that given before for $n = 0.01$. However, the water propagation will be slower than that given before for $n = 0.01$ as we see that at the final time of our simulations the water does not attain yet the downstream

end of the domain for $n = 0.02$ and $n = 0.03$.

5 Comparaison between a simple propagation model and the SWE

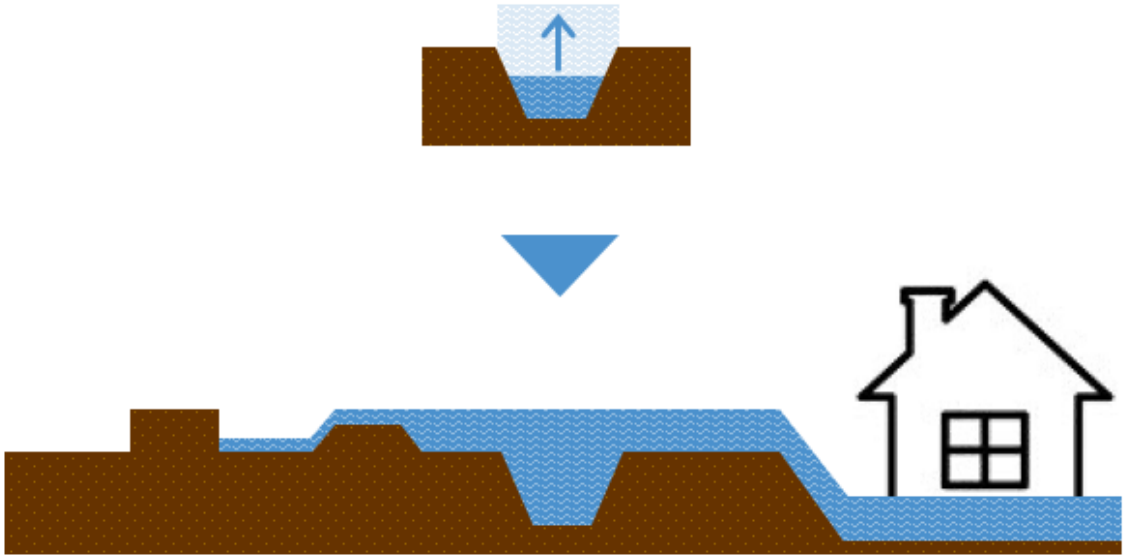


Figure 12: Figure showing the propagation algorithm with R which is developed at AXA Global P&C

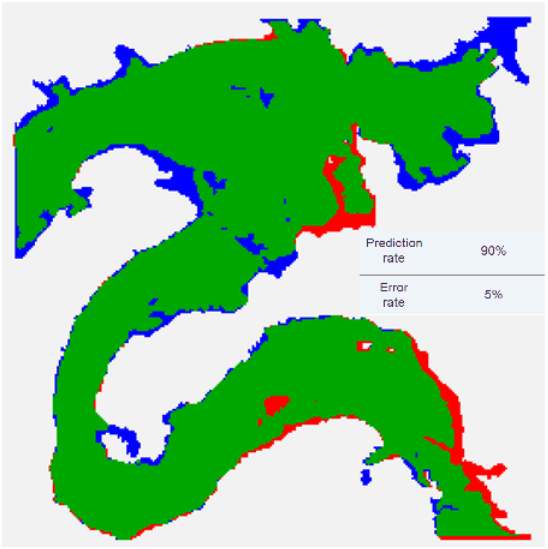


Figure 13: A figure generated from simulation showing a comparison between a simple filling method with Basilisk (blue) and the simulation with the simple propagation algorithm (red)

In this section, we give a comparative study between the model based on the resolution of the SWE (described above) and a propagation model which is developed by the actuary team at AXA Global P&C. This propagation model is simpler than that based on the resolution of

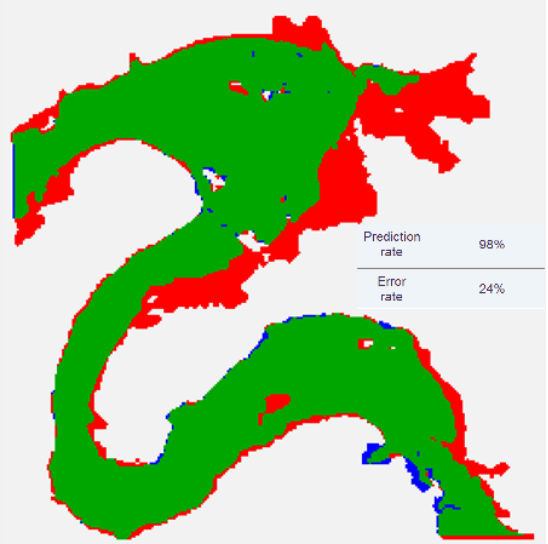


Figure 14: A figure generated from simulation showing a comparison between the inundated area in 1910 obtained with Basilisk at $t = 120000s$ (blue) and the simulation obtained with the propagation algorithm with non-uniform initial height (red) and their intersection (green)

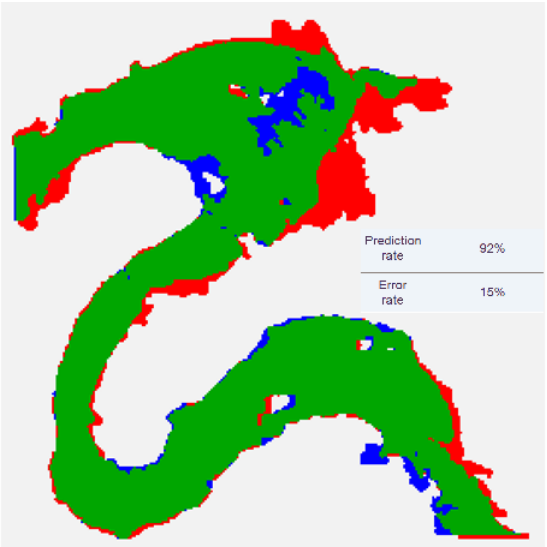


Figure 15: A figure generated from simulation showing a comparison between the inundated area in 1910 obtained with Basilisk at $t = 120000s$ (blue) and the simulation obtained with the propagation algorithm with non-uniform initial height and a dissipative term (red) and their intersection (green)

the SWE. It allows us to obtain footprint of the inundation. This model is programmed in R language and verifies the following properties (see figure 12):

- A cell is only flooded by one of its 8 direct neighbors
- The water level of a flooded cell is the maximum of its flooded neighbors
- The water depth of a cell cannot exceed the largest depth of its neighbors

In order to evaluate the performance of this simple propagation model, we will compare between the footprint obtained by this model for Paris flooding in 1910 and the footprint obtained

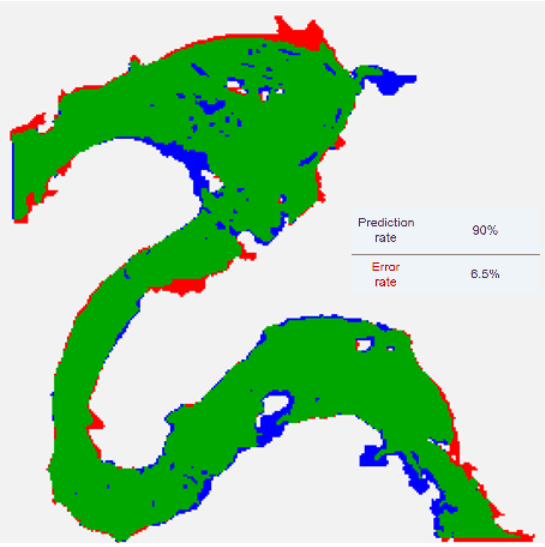


Figure 16: A figure generated from simulation showing a comparison between the inundated area in 1910 obtained with Basilisk at $t = 120000s$ (blue) and the simulation obtained with the propagation algorithm with non-uniform initial height and an interpolated topography (red) and their intersection (green)

above using the SWE. In both models, we use the IGN Digital elevation data with 75 m resolution. Figure 13 shows a comparison between the result of the flooding scenario with Basilisk, by filling in all the regions with a depth less than 35 m, and the result of the simple propagation algorithm. It can be seen that both models match well. The inundated region in figure 13 is clearly bigger than the one obtained in figure 4 with the hydrodynamic model. We deduce from this visual index that the simple model needs to be calibrated with respect to the hydrodynamic model in order to obtain a better approximation of the inundated region. For this reason, we consider first the simple propagation algorithm with a non-uniform height along the major bed and second by adding a dissipative term in the algorithm. These both modifications allow us to reduce the inundated region and thus ameliorate the error ratio (see figures 14 and 15), i.e. the over detected area, defined as the ratio of the area obtained with the simple algorithm which does not belong to the intersection area (the red area), and the area obtained with the same algorithm (error ratio $= \frac{\text{red area}}{\text{red area+green area}}$). Moreover, in both cases, the prediction ratio is greater than 90%, where the prediction ratio is defined as the ratio of the intersection area and the area obtained with the SWE model (prediction ratio $= \frac{\text{green area}}{\text{blue area+green area}}$).

Another way to calibrate the simple model written in R with respect to the simulation result obtained with Basilisk is to interpolate manually in R the furnished topography from IGN. In fact, Basilisk interpolates automatically the topography whenever the gradient of the height h is greater than 10 cm. Thus, we get in the final stage of the simulation with Basilisk an interpolated topography with approximately 25 m grid resolution. Figure 16 shows that the calibrated algorithm with the interpolated topography with (25 m grid resolution) approximates well the hydrodynamic model. Thus, a better resolution of the digital elevation model has a profound impact on the accuracy of the footprint extent.

6 Conclusion

From this case study, we can conclude that both models, the hydrodynamic model, based on the SWE, and the calibrated simple model, are able to simulate the inundated extent region of

Paris 1910 flood with a certain accuracy level. The hydrodynamic 2D model gave the best overall performances and a more precise information about the depth, the extent and the velocity of the flow. Basilisk code uses digital elevation models and generates the topography required by the hydrodynamic model. The mathematical model utilizes the SWE written in their conservative form without any simplification. Furthermore, the numerical method used is stable, which makes the Basilisk code suitable for practical applications.

However, the major obstacle of the simulation concerns the long runtime of the hydrodynamic model which is about 3 hours. This long runtime is needed to better calibrate the model with respect to the Manning coefficient. This is still the major disadvantage of the hydrodynamic model even with the adaptive meshes.

Furthermore, the Basilisk code should be tested with a better resolution grid for the DEM. Using a better resolution grid for the DEM has also the disadvantage of the calculation time which becomes again longer. Thus, a balance must be found between the available DEM's with different resolution, the runtime of the simulation, and the method accuracy.

The calibrated simple model worked well in the case of Paris flooding without defenses. Thus, one can use the calibrated simple model when we do not need a lot of precision. Further test cases should be undertaken to validate the general applicability of this conclusion. In particular, we need first to test this model to simulate Paris 1910 flood scenario with defenses and also simulate the flooding of other regions of the river where the morphology of the floodplain is different.

REFERENCES

[1] E., Audusse, F., Bouchut ,M.-o., Bristeau, R., Klein, B., Perthame, A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows, *SIAM J. Sci. Comp.,* **25**, 2050–2065, 2004.

[2] J., Burguete, P., Garcia-Navarro, Implicit schemes with large time step for non-linear equations : application to river flow hydraulics, *International Journal for Numerical Methods in Fluids*, **46**, 607–636, 2004.

[3] V., Caleffi, A. , Valiani, A., Zanni, Finite volume method for simulating extreme flood events in natural channels, *Journal of Hydraulic Research,* **41**, 167–177, 2003.

[4] V. T., Chow, *Open-Channel Hydraulics*, N. York, Ed. McGraw-Hill, 1959.

[5] S., Cordier, H., Coullon, O., Delestre, C., Lauerre, M.H. Le D., Pierre, G. Sadaka, Fullswofparal: Comparison of two parallelization strategies mpi and skelgis on a software designed for hydrology applications, *ESAIM Proceedings*, **43**, 59–79, 2013.

[6] O., Delestre, S., Cordier, F., James, F., Darboux, Simulation of rain-water overland-flow, *Proceeding of symposia in Applied Mathematics*, **67**, 537–546, 2009.

[7] DRIEE, *Les scenarios de crue en région île-de-france*, 1–22, 2012, available at `http://www.driee.ille-de-france.developpement-durable.gouv.fr/scenarios-de-crue-a1075.html`.

[8] GoogleMaps, 2014, available at `https://maps.google.fr`.

[9] IGN, Bd alti, 2014, available at `http://professionels.ign.fr/bdalti`.

[10] D.-H., Kim, Y.-S., Cho, Y.-K., Yi, Propagation and run-up of near shore tsunamis with HLLC approximate riemann solver, *Ocean Engineering*, **34**, 1164–1173, 2007.

[11] A., Kurganov, D., Levy, Central-upwind schemes for the saint-venant system, *Mathematical Modeling and Numerical Analysis*, **36**, 397–425, 2002.

[12] S., Popinet, An accurate adaptive solver for surface-tension-driven interfacial flows, *J Comput Phys*, **228**, 5838–5866, 2003.

[13] S., Popinet, Gerris: a tree-based adaptive solver for the incompressible Euler equations in complex geometries, *J Comput Phys*, **2**, 572–600, 2003.

[14] S., Popinet, G., Rickard, A tree-based solver for adaptive ocean modelling, *Ocean Model*, **16**, 224–249, 2007.

[15] A., Valiani, , V., Caleffi, A., Zanni, Case study : Malpasset dam-break simulation using a two-dimensional finite volume methods, *Journal of Hydraulic Engineering*, **128**, 440–472, 2002.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MULTIPHASE MODEL FOR THE NUMERICAL SIMULATION OF ICE-FORMATION IN SEA-WATER

Vanessa Covello[1], Antonella Abbà[1], Luca Bonaventura[2], Alessandro Della Rocca[2,3], and Lorenzo Valdettaro[2]

[1]Department of Aerospace Science and Technology, Politecnico di Milano
Via La Masa 34, 20156, Milano, Italy
e-mail: vanessa.covello, antonella.abba@polimi.it

[2] MOX-Modelling and Scientific Computing, Department of Mathematics, Politecnico di Milano
Via Bonardi 9, 20133, Milano, Italy
e-mail: luca.bonaventura, lorenzo.valdettaro@polimi.it

[3] Tenova S.p.A.
Via Albareto 31, 16135, Genova, Italy
e-mail: alessandro.dellarocca@tenova.com, alessandro.dellarocca@polimi.it

Keywords: ice modelling, multiphase flows, projection methods, finite volume method.

Abstract. *The first aim of this work is to improve the models currently available for the simulation of ice production in turbulent seawater, by means of the development of a multiphase model able to describe all the stages of ice production, overcoming the limitation of previous attempts, mainly based on Boussinesq approximation. We consider the mixture of ice and seawater as a dense compressible fluid, and we model the behaviour of seawater by an equation of state that links seawater density to temperature, salinity and pressure. The model is able to reproduce the interaction phenomena occurring between phases when the ice volume fraction exceeds the values allowed by the Boussinesq approximation, including in the momentum equations additional terms, related to the drag force between liquid and particles, and to the particle-particle interaction force. The second aim of our work is to implement and validate a numerical solver of our model. For this purpose, the model uses a sophisticated modelling approach, typically adopted for the numerical simulation of multiphase flows of industrial interest. The multiphase model can be coupled to the Large Eddy Simulation technique. The behaviour of the governing equations in the incompressible limit is investigated by means of a low-Mach number asymptotic analysis. The divergence constraint condition for the velocity field of continuous phase can then been imposed on the zero-Mach number equations by means of a projection method. The governing equations are discretized using the finite volume method, and the performance of the multiphase model has been assessed by solving a laminar Rayleigh-Bénard convection, for both large and small ice concentration regimes. In small concentration regime, the numerical solutions have been compared with the solutions obtained by a finite difference numerical code, based on the Boussinesq approximation.*

1 INTRODUCTION

Ice formation in the ocean is a complex phenomenon, whose correct description is essential in order to model adequately the atmosphere-ocean heat fluxes, the salt rejection phenomena and the damping effects related to the interactions among ice production, salt rejection and turbulent convection. The first stage of ice formation at the supercooled free surface of oceans, rivers and lakes is denoted as frazil-ice, a suspension of small individual randomly-oriented crystals, typically of 1 - 4 mm in diameter and 1 - 100 μ m in thickness [28, 31]. The frazil ice evolves rapidly into a thin slurry of ice platelets, known as grease ice. As cooling continues, the crystals freeze together into small disks called pancake-ice. Finally, during the last stage of sea ice evolution, the pankace-ice coalesces into a continuous ice sheet.

In this work, we focus our attention on the initial stages of ice production, frazil and grease ice. It is well known that ice formation greatly affects the geophysical and biological processes occurring in the polar oceans and that the subsequent salt rejection could play an important role in stimulating the convection processes in the oceans. Moreover, the presence of ice particles in rivers could produce serious damages to hydroelectric facilities, such as the blocking of turbine intakes, or the blockage of hydroelectric reservoirs. Therefore, the experimental and mathematical investigation of ice production has become, during the years, an active field of research.

The aim of this work is to improve the models currently available for the simulation of ice production in turbulent seawater, by means of the development of a multiphase model able to describe all the stages of ice production, overcoming the limitation of previous attempts, see e.g. [36, 42, 24, 40, 41, 21], mainly based on Boussinesq-like approximation. More specifically, in this work we consider the mixture ice-sea water as a dense compressible fluids. We model the behaviour of seawater by means of an equation of state derived from [7], that links seawater density to temperature, salinity and pressure, allowing us to go beyond the Boussinesq approximation. The model is able to reproduce the interaction phenomena occurring between phases when ice volume fraction is large, including in the momentum equations additional terms, related to the drag force between liquid and particles, and to the particle-particle interaction force. For this purpose, the model uses a sophisticated modelling approach, typically adopted for the numerical simulation of multiphase flows of industrial interest [16]. A low-Mach number asymptotic analysis has been performed to investigate the behaviour of the multiphase equations in the incompressible limit. The divergence constraint condition for the velocity field of continuous phase can then been imposed on the zero-Mach number equations by means of a projection method. As a starting point toward the development of an advanced numerical solver implementing our multiphase model, we develop a finite volume solver by means of a projection method based on [4, 19, 20], to be valid under the Boussinesq approximation, i.e. for small ice concentration regime, using the OpenFOAM library.

The paper is organized as follows: section 2 describes the multiphase model equations for the mixture ice-seawater, the non-dimensional form and the asymptotic analysis, whereas section 3 is devoted to the Boussinesq model equations. In section 4 we present the time discretization by projection method, for the Boussinesq model. Section 5 is devoted to the numerical results for a Boussinesq laminar Rayleigh-Bénard convection, for water and for the mixture ice-water, and ice-salt-water.

2 THE MULTIPHASE MODEL EQUATIONS

We now introduce the multiphase model equations according to the Eulerian-Eulerian approach, coupled with a sophisticated modelling approach for the interaction terms between phases, typically adopted for the numerical simulation of multiphase flows of industrial interest. The governing equations proposed here have been derived carrying out a volume averaging considering an averaging volume V such as $l^3 \ll V$, where l denotes the mean distance between the ice particles. We define the density of the mixture ice-seawater as $\rho = \phi_I \rho_I + (1 - \phi_I)\, \rho_w$, in which ϕ_I stands for the ice volume fraction, and ρ_I and ρ_w denote the ice and water density, respectively. In this work we consider the ice density as a constant value. A large number of models for the equation of state for seawater have been proposed over the years in the oceanographic literature. We have focused our attention on those formulations suitable for numerical modelling, see e.g.[6, 15, 14, 33, 32, 29, 23]. Here we propose a reduced form for the equation of state, derived from the equation proposed in [7], that combines the approaches proposed in [15, 23], providing a simplification of the UNESCO formula [33] that has been found to be valid for $-2°C < \theta < 40°C$, $0\ \ psu < S < 42\ \ psu$. Our state equation can be written as

$$\rho_w\,(T, S, p) = D_r p + E_r, \tag{1}$$

where the D_r and E_r coefficients stand for $D_r = b_5 TS + b_2 T + b_1$, and $E_r = a_5 TS + a_4 T^2 + a_3 S + a_2 T + a_1 + \rho_0$, $\rho_0 = 1000 kg/m^3$ is a reference density. The coefficients a_i and b_i are given in Table 1. According to the equation 1, the speed of sound is given by $c = [1/D_r]^{1/2}$.

n	a_n	b_n
1	-9.20601×10^{-2}	$+5.07043 \times 10^{-1}$
2	$+5.10768 \times 10^{-2}$	-3.69119×10^{-3}
3	$+8.05999 \times 10^{-1}$	-
4	-7.40849×10^{-3}	-
5	-3.01036×10^{-3}	$+1.75145 \times 10^{-5}$

Table 1: Coefficients of the equation of state

The continuity equations for the liquid and solid phases can be written as

$$\frac{\partial}{\partial t}[\,(1 - \phi_I)\,\rho_w] + \frac{\partial}{\partial x_i}[\,(1 - \phi_I)\,\rho_w u_i] = S_{mass}, \tag{2}$$

$$\frac{\partial}{\partial t}\,(\phi_I \rho_I) + \frac{\partial}{\partial x_i}\,(\phi_I \rho_I v_i) = -S_{mass}. \tag{3}$$

where S_{mass} is the source term, defined as

$$S_{mass} = \frac{2\phi_I h}{RL}\,(T_f - T_w)\,, \tag{4}$$

where h is the convective heat transfer coefficient, R stands for the average radius of the ice particles, L is the specific latent heat of fusion, T_w denotes the water temperature and T_f is the freezing temperature. The momentum equations for water and ice read

$$\frac{\partial}{\partial t}\Big[(1 - \phi_I)\,\rho_w u_i\Big] + \frac{\partial}{\partial x_j}\Big[(1 - \phi_I)\,\rho_w u_j u_i\Big] + f^{cor}_{i,j} u_j = -\,(1 - \phi_I)\,\frac{\partial \bar{p}}{\partial x_i} + (1 - \phi_I)\,\frac{\partial \tau_{ij}}{\partial x_j} + F^w_i + (1 - \phi_I)\,\bar{\rho}_w g_i + S_{mom}, \tag{5}$$

$$\frac{\partial}{\partial t}\left(\phi_I \rho_I v_i\right) + \frac{\partial}{\partial x_j}\left(\phi_I \rho_I v_j v_i\right) + f_{i,j}^{cor} v_j = -\phi_I \frac{\partial \bar{p}}{\partial x_i} + \phi_I \frac{\partial \tau_{ij}}{\partial x_j} + F_i^I + \phi_I \rho_I g_i - S_{mom}, \tag{6}$$

where the quasi pressure $\bar{p}$ is defined as $\bar{p} = p + \rho_0 g z$, and $\bar{\rho}_W$ stands for $\bar{\rho}_W = \rho - \rho_0$.
F_i^W and F_i^I can be split in two terms $F_i^W = f_i^{W,drag} + f_i^{W,I}$ and $F_i^I = f_i^{I,drag} + f_i^{I,I}$. Here, $f_i^{W,drag}$ and $f_i^{I,drag}$ denote the drag terms between seawater and particles, depending on the relative velocity $(u_i - v_i)$; $f_i^{W,I}$ denotes the shear stress term, analogous to the Reynolds stress in a single phase flow, that accounts for the velocity deviations that can occur on the seawater flow around a single particle; $f_i^{I,I}$ denotes the particle-particle interaction term. S_{mom} is the source term and $f_{i,j}^{cor}$ is an antisymmetric tensor such that $f_{i,j}^{cor} u_j$ and $f_{i,j}^{cor} v_j$ represent the Coriolis rotational forces on water and ice, respectively.
The momentum source term can be expressed as

$$S_{mom} = \frac{2\phi_I h}{RL}\left(T_f - T_W\right) v_i. \tag{7}$$

The liquid-solid drag term satisfies $f_i^{W,drag} = -f_i^{I,drag}$. Following the approach proposed in [17], for gas-solid fluidized bed, $f_i^{W,drag}$ can be written as $f_i^{W,drag} = K_D\left(u_i - v_i\right)$, where K_D stands for the two-phase drag coefficient. For the evaluation of K_D we refer to [17], in which K_D is defined as a function of a drag coefficient C_D, and the relative Reynolds number $Re_s = \left(\rho_W D \mid v_i - u_i \mid\right)/\mu_W$. A large number of models have been proposed during the years for the evaluation of the drag terms, see e.g. [43], for gas-solid multiphase flows, and [17, 43, 45], for the evaluation of the liquid-solid drag in the framework of pipeline slurry flows, fluidized bed reactors, as well as in other chemical reacting systems, see e.g. [27, 22, 39, 46, 11, 37], or in numerical modelling of volcanic eruptions, see e.g. [8, 9]. The term $f_i^{W,I}$ is analogous to the Reynolds stress in a single phase flow. Following the approach reported in [12], we can define $f_i^{W,I}$ by means of an effective stress τ_{ij}^e such that

$$f_i^{W,I} = \frac{\partial}{\partial x_j}\left(\left(1 - \phi_I\right)\tau_{ij}^e\right), \tag{8}$$

where the effective shear stress, τ_{ij}^e, is related to the turbulence parameters of the flow. The evaluation of τ_{ij}^e, is often carried out by means of eddy viscosity models. The particle-particle interaction term can be written as,

$$f_i^{I,I} = \phi_I \frac{\partial}{\partial x_j}\left(-p_s \delta_{ij} + \tau_{s,ij}\right). \tag{9}$$

The terms p_s and $\tau_{s,ij}$ stand for the solid pressure and the solid shear stress tensor, respectively. p_s and $\tau_{s,ij}$ can be evaluated as function of the granular temperature Θ, according to the kinetic theory, originally applied by Bagnold [2] to this purpose. Starting from [2], several efforts have been devoted over the years to the development of the kinetic theory, see e.g [35, 38, 30, 26]. Here we mainly refer to the formulations proposed in [16, 30, 25]. Following this theory, an additional transport equation for the kinetic energy of the fluctuating motion of the dispersed phase, also called granular temperature Θ, is required. The kinetic energy associated with the particle velocity fluctuation, can be expressed as $\Theta = \frac{1}{3}\langle C\rangle^2$, where C is the deviation of the particle velocity from the mean velocity. The operator $\langle\ \rangle$ denotes the mean value based on the Gaussian distribution. The equation for the granular temperature can be written, following [16], as

$$\frac{3}{2}\left[\frac{\partial}{\partial t}\left(\rho_I \phi_I \Theta\right) + \frac{\partial}{\partial x_i}\left(\rho_I \phi_I v_i \Theta\right)\right] = \left[-p_s \delta_{ij} + \tau_{s,ij}\right]\frac{\partial v_i}{\partial x_j} - \frac{\partial}{\partial x_i}\left(k_\Theta \frac{\partial \Theta}{\partial x_i}\right) - \Gamma_\Theta. \tag{10}$$

Here, p_s, $\tau_{s,ij}$, k_θ, Γ_θ stand for the solid pressure, the solid shear stress tensor, the coefficient of conductivity of granular temperature, and the dissipation term, respectively, and can be calculated as functions of Θ using different approaches, see e.g. [30, 16, 44, 25].

The transport equation for the water temperature can be written as follows

$$\frac{\partial}{\partial t}\Big[\left(1-\phi_I\right) T_W\Big] + \frac{\partial}{\partial x_i}\Big[\left(1-\phi_I\right) u_i T_W\Big] = \frac{\partial}{\partial x_i}\left(\left(1-\phi_I\right) k_t \frac{\partial T_W}{\partial x_i}\right) + \phi_I \lambda_t \left(T_I - T_W\right) + S_T. \tag{11}$$

The parameter λ_t is defined as $\lambda_t = \frac{3h}{c_p \rho_W R}$. S_T is the source term, defined as

$$S_T = \frac{2h\phi_I}{R c_p \rho_W}\left(T_f - T_W\right), \tag{12}$$

where c_p stands for the water specific heat at constant pressure.

The equation for the salinity S is formulated as

$$\frac{\partial}{\partial t}\Big[\left(1-\phi_I\right) S\Big] + \frac{\partial}{\partial x_i}\Big[\left(1-\phi_I\right) S u_i\Big] = \frac{\partial}{\partial x_i}\left(\left(1-\phi_I\right) k_s \frac{\partial S}{\partial x_i}\right) + S_{sal}, \tag{13}$$

where k_s is the salt diffusivity, and S_{sal} is the source term,

$$S_{sal} = \frac{2h S_v \phi_I \rho_I}{R L \rho_W}\left(T_f - T_W\right), \tag{14}$$

where S_v denotes the increase of salinity due to the freezing, that can be defined as $S_v = m_s / \Big[V\left(1 - \frac{\phi_I \rho_I}{\rho}\right)\Big]$, in which m_s stands for the salt mass and V is the volume of the water-ice mixture.

2.1 Non-dimensional form

In order to obtain the non-dimensional form of the governing equations, we consider the vertical length of the computational domain L as the reference length; the temperature gap between upper and lower boundary of the computational domain, ΔT, as temperature scale; the water velocity ,u_∞, as the velocity scale; the characteristic time t^* as temporal scale; the water density, ρ_{W_∞}, as density scale; and p_∞ and S_0, as reference pressure and reference salinity, respectively. We define the following dimensionless parameters,

$$\widetilde{\alpha} = \alpha \Delta T,\ \ \widetilde{S} = \frac{S}{S_0 \alpha \Delta T},\ \ \widetilde{\Theta} = \frac{\Theta}{u_\infty^2},\ \ \widetilde{p}_s = \frac{p_s}{\rho_{W_\infty} u_\infty^2},\ \ M_\infty = \frac{u_\infty}{c_\infty}, \tag{15}$$

and the f parameter as $f = u_\infty t^* / L$. For simplicity, in the following sections we denote $S_0 \alpha \Delta T$ as S_∞, and ΔT as T_∞. Other dimensionless parameters are the non-dimensional numbers, Reynolds, R_e, Prandtl, P_r, Rayleigh, R_a, and Schmidt, S_{c_s}

$$R_e = \frac{u \rho_W L}{\mu},\ \ P_r = \frac{\nu}{k_t},\ \ R_a = \frac{g \alpha L^3 \Delta T}{k_t \nu},\ \ S_{c_s} = \frac{\nu}{k_s}, \tag{16}$$

where μ and ν are the dynamic and kinematic viscosity, respectively, k_t is the water thermal diffusivity, α the thermal expansion coefficient, g the gravity acceleration and k_s is the salt diffusivity. The non-dimensional form of the equation of state, Equation 1, yields

$$\widetilde{\rho}_{\mathrm{w}} = D_r^* \widetilde{D}_r \widetilde{p} + E_r^* \widetilde{E}_r, \tag{17}$$

$$D_r^* = \frac{D_{r_\infty} p_\infty}{\rho_{\mathrm{w}_\infty}}, \;\; E_r^* = \frac{E_{r_\infty}}{\rho_{\mathrm{w}_\infty}}, \;\; \widetilde{D}_r = \frac{D_r}{D_{r_\infty}}, \;\; \widetilde{E}_r = \frac{E_r}{E_{r_\infty}}. \tag{18}$$

In the following we will work with the non dimensional equations, dropping the˜sign for simplicity. The non-dimensional mass conservation equations can be written as,

$$\frac{\partial}{\partial t}\Big[(1-\phi_I)\,\rho_{\mathrm{w}}\Big] + f\frac{\partial}{\partial x_i}\Big[(1-\phi_I)\,\rho_{\mathrm{w}} u_i\Big] = S_{mass}, \tag{19}$$

$$\frac{\partial}{\partial t}\left(\phi_I \rho_I\right) + f\frac{\partial}{\partial x_i}\left(\phi_I \rho_I v_i\right) = -S_{mass}, \tag{20}$$

where the source term yields,

$$S_{mass} = \frac{2\phi_I h t^*}{RL\rho_{\mathrm{w}_\infty}}\left(T_f - T_{\mathrm{w}}\right). \tag{21}$$

The non-dimensional momentum equations can be defined as

$$\frac{\partial}{\partial t}\Big[(1-\phi_I)\,\rho_{\mathrm{w}} u_i\Big] + f\frac{\partial}{\partial x_j}\Big[(1-\phi_I)\,\rho_{\mathrm{w}} u_j u_i\Big] + f_{Cor} u_i = -f\frac{p_\infty}{\rho_{\mathrm{w}_\infty} c_\infty^2}\frac{1}{M_\infty^2}(1-\phi_I)\frac{\partial \bar{p}}{\partial x_i} + f\frac{1}{Re_\infty}(1-\phi_I)\left(\frac{\partial^2 u_j}{\partial x_j \partial x_i} + \frac{1}{3}\frac{\partial}{\partial x_j}\left(\frac{\partial u_i}{\partial x_i}\right)\right) + F_i^{\mathrm{w}} + f\frac{R_a}{P_r Re_\infty^2 \alpha}(1-\phi_I)\,\bar{\rho}_{\mathrm{w}} - S_{mom}, \tag{22}$$

$$\frac{\partial}{\partial t}\left(\phi_I \rho_I v_i\right) + f\frac{\partial}{\partial x_j}\left(\phi_I \rho_I v_j v_i\right) + f_{Cor} v_i = -f\frac{p_\infty}{\rho_{\mathrm{w}_\infty} c_\infty^2}\frac{1}{M_\infty^2}\phi_I\frac{\partial \bar{p}}{\partial x_i} + f\frac{1}{Re_\infty}\phi_I\left(\frac{\partial^2 u_j}{\partial x_j \partial x_i} + \frac{1}{3}\frac{\partial}{\partial x_j}\left(\frac{\partial u_i}{\partial x_i}\right)\right) + F_i^I + f\frac{R_a}{P_r Re_\infty^2 \alpha}\phi_I \rho_I + S_{mom}, \tag{23}$$

where the source term stands for

$$S_{mom} = \frac{2\phi_I h t^*}{RL\rho_{\mathrm{w}_\infty}}\left(T_f - T_{\mathrm{w}}\right) v_i. \tag{24}$$

The temperature equation for water can be written as

$$\frac{\partial}{\partial t}\Big[(1-\phi_I)\,T_{\mathrm{w}}\Big] + f\frac{\partial}{\partial x\,i}\Big[(1-\phi_I)\,u_i T_{\mathrm{w}}\Big] = \frac{t^*\nu}{L^2}\frac{\partial}{\partial x_i}\left((1-\phi_I)\frac{1}{P_r}\frac{\partial T_{\mathrm{w}}}{\partial x_i}\right) + \phi_I t^* \lambda_t\left(T_I - T_{\mathrm{w}}\right) + S_T. \tag{25}$$

The source term is given by

$$S_T = \frac{2ht^*\phi_I}{Rc_p\rho_{\mathrm{w}}\rho_{\mathrm{w}_\infty}}\left(T_f - T_{\mathrm{w}}\right). \tag{26}$$

The non-dimensional transport equation for salinity can be written as,

$$\frac{\partial}{\partial t}\Big[(1-\phi_I)\,S\Big]+f\frac{\partial}{\partial x_i}\Big[(1-\phi_I)\,Su_i\Big]=f\frac{\partial}{\partial x_i}\left((1-\phi_I)\,\frac{1}{ScRe_\infty}\frac{\partial S}{\partial x_i}\right)+S_{sal}, \tag{27}$$

$$S_{sal}=\frac{2ht^*S_v\phi_I\rho_I\rho_{w_\infty}}{RLS_\infty\rho_w}\,(T_f-T_w)\,. \tag{28}$$

Finally, for granular temperature, the non dimensional form yields,

$$\frac{3}{2}\Big[\frac{\partial}{\partial t}\,(\rho_I\phi_I\Theta)+f\frac{\partial}{\partial x_i}\,(\rho_I\phi_Iv_i\Theta)\Big]=[-fp_s\delta_{ij}+\tau_{s,ij}]\,\frac{\partial v_i}{\partial x_j}-\frac{\partial}{\partial x_i}\left(k_\Theta\frac{\partial\Theta}{\partial x_i}\right)-\Gamma_\Theta. \tag{29}$$

2.2 Low-Mach number asymptotic analysis

To investigate the behaviour of the multiphase equations in the incompressible limit, we have expanded each flow variables as e.g., the pressure,

$$p\,(x,t,M_\infty)=p^{(0)}+M_\infty^2p^{(2)}\,(x,t,M_\infty)\,. \tag{30}$$

For simplicity, here we report only the zero-Mach number equation for seawater density, Equation 17, the zero-Mach number equation for mass concentration, Equation 19, and the leading and the zero-order equations for momentum, Equations 22 and . The other zero-Mach number equations follow the non-dimensional representation, with all the variables evaluated at the zero-order. The non-dimensional equation for seawater density, at order $O\left(M_\infty^0\right)$, is derived as:

$$\rho_w^{(0)}=D_r^*\left(D_r^{(0)}p^{(0)}\right)+E_r^*E_r^{(0)}, \tag{31}$$

where the coefficients $D_r^{(0)}$ and $E_r^{(0)}$ are functions of the zero-order temperature and salinity, see the Appendix A for the detailed expression of each coefficient. The water mass conservation at order $O\left(M_\infty^0\right)$ yields

$$\frac{\partial}{\partial t}\Big[\left(1-\phi_I^{(0)}\right)\rho_w^{(0)}\Big]+f\frac{\partial}{\partial x_i}\Big[\left(1-\phi_I^{(0)}\right)\rho_w^{(0)}u_i^{(0)}\Big]=S_{mass}^{(0)}. \tag{32}$$

The zero-order ice mass conservation is given similarly. The water and ice momentum equations at the leading order $O\left(M_\infty^{-2}\right)$ yield

$$\frac{\partial\bar{p}^{(0)}}{\partial x_i}=0, \tag{33}$$

therefore, according to the definition of quasi-pressure $\bar{p}^{(0)}$, we obtain

$$\frac{\partial p^{(0)}}{\partial x}=\frac{\partial p^{(0)}}{\partial y}=0, \tag{34}$$

$$\frac{\partial p^{(0)}}{\partial z}=-\rho_0g. \tag{35}$$

The water momentum equation at order $O\left(M_\infty^0\right)$ yields:

$$\frac{\partial}{\partial t}\left[\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}u_i^{(0)}\right]+f\frac{\partial}{\partial x_j}\left[\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}u_j^{(0)}u_i^{(0)}\right]+f_{Cor}^{(0)}u_i^{(0)}=$$
$$-f\frac{p_\infty}{\rho_{W_\infty}c_\infty^2}\left(1-\phi_I^{(0)}\right)\frac{\partial\bar{p}^{(2)}}{\partial x_i}+f\frac{1}{Re_\infty}\left(1-\phi_I^{(0)}\right)\left(\frac{\partial^2 u_j^{(0)}}{\partial x_j\partial x_i}+\frac{1}{3}\frac{\partial}{\partial x_j}\left(\frac{\partial u_i^{(0)}}{\partial x_i}\right)\right)+$$
$$F_i^{w(0)}+f\frac{R_a}{P_rR_{e_\infty}^2\alpha}\left(1-\phi_I^{(0)}\right)\bar{\rho}_W^{(0)}-S_{mom}^{(0)}. \tag{36}$$

The ice momentum equation at order $O\left(M_\infty^0\right)$ can be similarly derived. Using the continuity equation, Equation 32, and assuming $f=1$, the following constraint on the divergence of the water velocity is obtained

$$\nabla\cdot u_i^{(0)}=\frac{1}{\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}}\left[-\frac{D}{Dt}\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}+S_{mass}^{(0)}\right], \tag{37}$$

where $\frac{D}{Dt}\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}$ can be written as

$$\frac{D}{Dt}\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}=\left(1-\phi_I^{(0)}\right)\frac{D\rho_W^{(0)}}{Dt}-\rho_W^{(0)}\frac{D\phi_I^{(0)}}{Dt}. \tag{38}$$

Using the zero-oder ice mass conservation, $\frac{D\phi_I^{(0)}}{Dt}$ can be obtained as

$$\frac{D\phi_I^{(0)}}{Dt}=-\frac{S_{mass}^{(0)}}{\rho_I^{(0)}}-\phi_I^{(0)}\frac{\partial v_i^{(0)}}{\partial x_i}, \tag{39}$$

and $\frac{D\rho_W^{(0)}}{Dt}$ can be derived as

$$\frac{D\rho_W^{(0)}}{Dt}=\left.\frac{\partial\rho_W^{(0)}}{\partial S^{(0)}}\right|_{T_W^{(0)},p^{(0)}}\frac{DS^{(0)}}{Dt}+\left.\frac{\partial\rho_W^{(0)}}{\partial T_W^{(0)}}\right|_{S^{(0)},p^{(0)}}\frac{DT_W^{(0)}}{Dt}+\left.\frac{\partial\rho_W^{(0)}}{\partial p^{(0)}}\right|_{T_W^{(0)},S^{(0)}}\frac{Dp^{(0)}}{Dt}, \tag{40}$$

in which all the partial derivative can be evaluated by means of the equation of state, whereas $\frac{DS^{(0)}}{Dt}$ and $\frac{DT_W^{(0)}}{Dt}$ can be derived using the zero-order temperature and salinity equations, respectively, and $\frac{Dp^{(0)}}{Dt}$ can be evaluated considering the Equation 33, see the Appendix A for the complete derivation of these terms. The divergence constraint can then be rewritten as

$$\nabla\cdot u_i^{(0)}=\frac{1}{\left(1-\phi_I^{(0)}\right)\rho_W^{(0)}}\left(-\phi_I^{(0)}\frac{\partial v_i^{(0)}}{\partial x_i}+\frac{1}{\rho_W^{(0)}}\left(1-\frac{\rho_W^{(0)}}{\rho_I}\right)S_{mass}^{(0)}\right)-\frac{1}{\rho_W^{(0)}}\frac{D\rho_W^{(0)}}{Dt}. \tag{41}$$

The zero-order equations coupled with the divergence constraint, Equation 41, can then be discretized in time using a projection method, and the divergence constraint will be used to enforce the mass conservation and the equation of state.

3 BOUSSINESQ APPROXIMATION OF THE MULTIPHASE MODEL EQUATIONS

Here we present the Boussinesq approximation for the multiphase equations introduced in the previous sections. This is a starting point for the development of a finite volume multiphase solver, based on the aforementioned multiphase model. The Boussinesq approximation, indeed, allows us to model the initial stage of the ice production phenomena, frazil-ice, in which the ice volume fraction typically does not exceed 10^{-3}. The formulation reported here follows the approach presented in [1]. According to the Boussinesq approximation, the seawater density can be expressed by means of the following linearized equation of state, see e.g. [21]

$$\rho_w = \rho_0 \left(1 - \beta \left(T - T_0\right) + \beta_s \left(S - S_0\right)\right), \tag{42}$$

in which ρ_0, T_0, S_0, denote the reference density, temperature, and salinity, respectively, whereas β and β_s are the expansion coefficients. Using the Boussinesq approximation, density variations are sufficiently small such that they can be retained only in the buoyancy term of the momentum equation, see e.g [10]. According to such approximation, our multiphase model reduces to a single-phase model, then all the interaction terms and the mass and momentum source terms vanish. Retaining only the equations for water, and assuming the water density to be constant, except in the buoyancy term, the governing equations reduce to

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{43}$$

$$\frac{\partial u_i}{\partial t} + \frac{\partial \left(u_i u_j\right)}{\partial x_j} = -\frac{\partial \pi}{\partial x_i} + \frac{1}{Re}\frac{\partial^2 u_i}{\partial x_j^2} - g_i \left(\beta \left(T - T_0\right) - \beta_s \left(S - S_0\right)\right), \tag{44}$$

$$\frac{\partial T_w}{\partial t} + \frac{\partial \left(u_i T_w\right)}{\partial x_i} = \frac{\partial}{\partial x_i}\left(k_t \frac{\partial T_w}{\partial x_i}\right) + S_T, \tag{45}$$

where π denotes the quasi-pressure, defined as $\pi = \left(p - \rho_0 g\right)/\rho_0$, according to the treatment proposed in [13]. The source term for temperature, S_T, is given by

$$S_T = \frac{2h}{R c_p \rho_w \rho_I} \rho C_I \left(T_f - T_w\right), \tag{46}$$

where C_I denotes the ice mass concentration, and ρ stands for the density of the mixture, given by

$$\rho = \left(\frac{C_I}{\rho_I} + \frac{\left(1 - C_I\right)}{\rho_w}\right)^{-1}. \tag{47}$$

We can describe the scalar transport phenomena for the ice mass concentration C_I, by means of the following equation

$$\frac{\partial}{\partial t}\left(\rho C_I\right) + \frac{\partial}{\partial x_i}\left(\rho C_I u_i\right) = -\omega_{r_i}\frac{\partial \left(\rho C_I\right)}{\partial x_i} + k_I \frac{\partial}{\partial x_i}\left(\rho \frac{\partial C_I}{\partial x_i}\right) + S_{C_I}, \tag{48}$$

where k_I is the ice thermal diffusivity, and S_{C_I} denotes the source term, that can be expressed as

$$S_{C_I} = \frac{2h}{R L \rho_I} \rho C_I \left(T_f - T_w\right). \tag{49}$$

The Equation 48 is derived assuming the ice velocity as $u_I = u + \omega_r$, where u is the velocity of the mixture, and $\omega_r = (0, v_r, 0)$ stands for the rise velocity, related to buoyancy forces. In this

work we left ω_r as a free parameter, to be evaluated according to experimental investigations, see e.g. [34]. The transport equation for salinity S can be written as

$$\frac{\partial S}{\partial t} + \frac{\partial}{\partial x_i}\left(S u_i\right) = \frac{\partial}{\partial x_i}\left(k_s \frac{\partial S}{\partial x_i}\right) + S_{sal}, \tag{50}$$

where k_s is the salt diffusivity, and S_{sal} is the source term, defined as follows

$$S_{sal} = \frac{2h}{RL\rho_w} S_v \rho C_I \left(T_f - T_w\right), \tag{51}$$

where the increase of salinity due to the freezing, S_v, stands for $S_v = m_s / \left[V\left(1 - C_I\right)\right]$, in which m_s is the salt mass and V is the volume of the water-ice mixture.

4 TIME DISCRETIZATION BY PROJECTION METHOD FOR BOUSSINESQ FLOWS

The Boussinesq simplification of the Navier-Stokes equations is integrated in time by a second order projection method constructed in order to have good kinetic energy conservation. Space discretization is performed by a cell-centered second order finite volume approach, while time evolution is computed by using the trapezoidal rule in a colocated non-incremental approximate projection method. The overall method is characterized by a limited amount of numerical dissipation, similarly to [19], and thus it forms an appropriate framework for developing future variants to be applied in direct and large-eddy simulations. An alternative projection approach for density variable flows is reported in [5]. Here we present only a brief overview of the numerical methodology, thus postponing a complete discussion of the projection method elsewhere.

The discretization of the evolution equations under the Boussinesq approximation involves several steps. All the primitive quantities are defined as cell averages at cell center of mass and at integer time steps, while only the quasi-pressure π is defined at cell centers and at half-times, similarly to [3]. Here we denote with A_f the area of cell faces, with $n_{f,i}$ the face normal unit vector and with $\overline{(\cdot)}^f$ the interpolation operator acting on cell centered quantities to construct face centered quantities.

Step 1. – The face normal velocities $U^n = \overline{u_i^n}^f n_{f,i}$ are extrapolated in time using second order Adams-Bashforth

$$U^{n+1,\star} = \frac{3}{2}U^n - \frac{1}{2}U^{n-1}, \tag{52}$$

which will be consistent with the incompressibility condition if the face normal velocities from the previous time steps are divergence-free. For the construction of the evolution equation for the ice concentration the face normal mass fluxes $m^n = \overline{\rho^n}^f \, \overline{\left(u_i^n + \omega_{r,i}\right)}^f n_{f,i}$ are constructed in a similar manner

$$m^{n+1,\star} = \frac{3}{2}m^n - \frac{1}{2}m^{n-1}, \tag{53}$$

accordingly:

$$\rho^{n+1} = \frac{3}{2}\rho^n - \frac{1}{2}\rho^{n-1}. \tag{54}$$

Step 2. – The temperature equation is integrated implicitly by the trapezoidal rule

$$\begin{aligned} \frac{T^{n+1} - T^n}{\Delta t} = \; & -\tfrac{1}{2}\left[D\left(U^{n+1,\star}\,\overline{T^{n+1}}^f\right) + D\left(U^{n+1,\star}\,\overline{T^n}^f\right)\right] \\ & + \tfrac{1}{2}\left[L_{k_t}(T^{n+1}) + L_{k_t}(T^n)\right] + S_T^{n+1/2}, \end{aligned} \tag{55}$$

where we introduced the discrete second-order divergence operator D, approximating the divergence at cell centers from face centered quantities, and the discrete second order operator L_{k_t}, approximating $\frac{\partial}{\partial x_j}\left(k_t \frac{\partial T}{\partial x_j}\right)$ at cell centers from cell centered quantities, in which we silently introduce the interpolation $\overline{k_t}^f$. This allows to construct the buoyancy terms in the other equations.

Step 3. – The salinity equation is integrated implicitly by the trapezoidal rule as well

$$\begin{aligned}\frac{S^{n+1}-S^n}{\Delta t} = & -\tfrac{1}{2}\left[D\left(U^{n+1,\star}\,\overline{S^{n+1}}^f\right)+D\left(U^{n+1,\star}\,\overline{S^n}^f\right)\right] \\ & +\tfrac{1}{2}\left[L_{k_S}(S^{n+1})+L_{k_S}(S^n)\right]+S_S^{n+1/2},\end{aligned} \tag{56}$$

where the discrete operator L_{k_S} approximates $\frac{\partial}{\partial x_j}\left(k_S \frac{\partial S}{\partial x_j}\right)$. Successively the ice concentration equation is integrated by

$$\begin{aligned}\frac{\rho^{n+1}C^{n+1}-\rho^n C^n}{\Delta t} = & -\tfrac{1}{2}\left[D\left(m^{n+1,\star}\,\overline{C^{n+1}}^f\right)+D\left(m^{n+1,\star}\,\overline{C^n}^f\right)\right] \\ & +\tfrac{1}{2}\left[L_{k_I\rho^{n+1}}(C^{n+1})+L_{k_I\rho^n}(C^n)\right]+S_I^{n+1/2},\end{aligned} \tag{57}$$

where the discrete operator $L_{k_I\rho^n}$ approximates $\frac{\partial}{\partial x_j}\left(k_I\rho^n \frac{\partial C}{\partial x_j}\right)$ to second order accuracy.

Step 4. – The momentum equation is integrated by the trapezoidal rule, thus obtaining a predicted velocity field $u_i^{n+1,\star}$ which is not generally divergence-free

$$\begin{aligned}\frac{u_i^{n+1,\star}-u_i^n}{\Delta t} = & -\frac{1}{2}\left[D\left(U^{n+1,\star}\,\overline{u_i^{n+1,\star}}^f\right)+D\left(U^{n+1,\star}\,\overline{u_i^n}^f\right)\right] \\ & +\frac{1}{2}\left[L_\nu\left(u_i^{n+1,\star}\right)+L_\nu\left(u_i^n\right)\right]-G_i\left(\pi^{n-1/2}\right) \\ & -\frac{1}{2}\,g_i\left[\beta\left(\left(T^{n+1}-T_0\right)+\left(T^n-T_0\right)\right)-\beta_s\left(\left(S^{n+1}-S_0\right)+\left(S^n-S_0\right)\right)\right]\end{aligned} \tag{58}$$

where we introduced the discrete second-order gradient operator G_i approximating the gradient at cell centers from cell centered quantities by the discrete Gauss theorem applied on $\overline{\pi^{n-1/2}}^f$.

Step 5. – The term containing $\pi^{n-1/2}$ is removed from the predicted velocity field

$$u_i^{n+1,\star\star} = u_i^{n+1,\star} + \Delta t\; G_i\left(\pi^{n-1/2}\right) \tag{59}$$

in order to construct a non-incremental projection step (i.e., a Poisson equation on $\pi^{n+1/2}$ rather than on $\pi^{n+1/2}-\pi^{n-1/2}$).

Step 6. – The Poisson equation is solved for the unknown $\pi^{n+1/2}$ required to update the divergence-free face normal velocities

$$L\left(\pi^{n+1/2}\right) = \frac{1}{\Delta t} D\left(\overline{u_i^{n+1,\star\star}}^f\right). \tag{60}$$

Here homogeneous Neumann boundary conditions are applied to the discrete second order laplacian operator L, which are responsible for the preservation of the initial normal gradient of π^0 at the boundaries (see, e.g., [18]).

Step 7. – The predicted and face normal velocities are respectively updated according to

$$u_i^{n+1} = u_i^{n+1,\star\star} - \Delta t \; G_i\left(\pi^{n+1/2}\right), \tag{61}$$

$$U^{n+1} = \left[\overline{u_i^{n+1,\star\star}}^{f} n_{f,i} - \Delta t \; G_{n_f}\left(\pi^{n+1/2}\right)\right], \tag{62}$$

$$m^{n+1} = \left[\overline{\rho^{n+1}}^{f} \; \overline{\left(u_i^{n+1,\star\star} + \omega_{r,i}\right)}^{f} n_{f,i} - \Delta t \; \overline{\rho^{n+1}}^{f} G_{n_f}\left(\pi^{n+1/2}\right)\right], \tag{63}$$

where the discrete second order face normal gradient operator G_{n_f} is introduced, approximating the normal gradient at cell faces from cell centered quantities. The resulting face normal velocities U^{n+1} are divergence-free as required by the convective terms. Tangential boundary conditions on u_i^{n+1} are now directly enforced, since they cannot result naturally from the projection step.

The initial value problem requires initial values of u_i^0 at time $t = 0$ and relative boundary conditions. In our calculations we always started from uniformly zero fields on u_i^0 and π^0 whenever possible, or we started from a velocity field resulting after an initial projection step. As a consequence we implicitly assumed at boundaries the artificial condition $\frac{\partial \pi}{\partial x_i} n_{b,i} = 0$, where $n_{b,i}$ represents the unit vector normal to the boundary.

A numerical solvers called `fsBMA` has been implemented according to this formulation, by means of the OpenFOAM library [47].

5 NUMERICAL RESULTS

In order to validate the full multiphase model we shall proceed step by step. As a first step we consider here a simplified version of the equations, namely the Boussinesq approximation. The time discretization is the same described in the previous section, except for source terms, that have been integrated at first order for simplicity, and for the density of the mixture ρ that has been extrapolated at first order. Our analysis has to be intended as a sensitivity study of the influence of the ice formation process on the behaviour of the velocity and temperature fields, for unsalted and salted water, therefore does not completely reproduce the realistic condition of ice production in seawater. The results we present here are the numerical solutions of a three dimensional laminar Rayleigh-Bénard convection of a Boussinesq fluid. Numerical simulations have been performed using the developed `fsBMA` solver, for water, section 5.1, for the water-ice and water-ice-salt mixture, section 5.2.

5.1 Laminar Rayleigh-Bénard convection for water

Numerical simulations have been performed for $Ra = 2.3 \times 10^4$, using a three dimensional $6D$xDx$6D$ domain, where D denotes the height, discretized by means of a uniform 60x10x60 computational grid and a non-uniform 60x50x60 grid, with the smaller mesh size near the top and bottom walls. The boundary conditions at top and bottom walls are free-slip for velocity. The zero-gradient condition has been imposed at top and bottom for pressure, whereas for the temperature we impose T_t, and T_b, respectively. The boundary conditions for all the lateral boundaries are periodic. The simulations have been carried out using $T_t = 272.15\,[K]$, $T_b = 274.15\,[K]$, and $\beta = 3.6476 \times 10^{-3}\,[1/K]$, $\nu = 1.6438 \times 10^{-6}\,[m^2/s]$, $k_t = 1.3905 \times 10^{-7}\,[m^2/s]$ for the thermal expansion coefficient, kinematic viscosity coefficient, and thermal diffusivity, respectively. The simulations have been performed by imposing an initial random perturbation for the temperature and velocity fields. Numerical results have been compared with those obtained by means of a finite difference numerical code presented in [1], on a non-uniform 60x50x60 grid, locally refined near top and bottom walls.

Figure 1 shows the contours and the isolines for the temperature fields, and the U_y velocity component, in the vertical x, y midplane at $z = 3D$ and in the horizontal x, z midplane, at $y = D/2$, respectively, obtained by means of the non-uniform 60x50x60 grid.

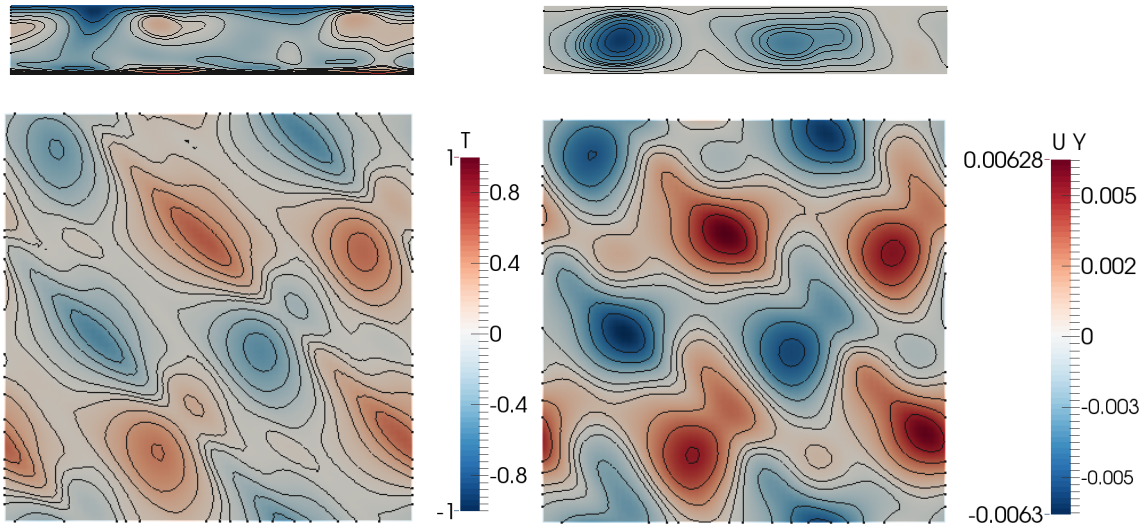


Figure 1: Contours and isolines of temperature, T, left column, and of the U_y velocity component, right column, in the x, y midplane, top row, and in the x, z midplane, bottom row, for water at $Ra = 2.3 \times 10^4$.

Figure 2 shows the comparison between the finite difference code, `FD`, and the developed finite volume solver `fsBMA`, in terms of the time average of the temperature profile, spatially averaged over the x, z plane, and the root mean square of U_y velocity profile, along the y axis.

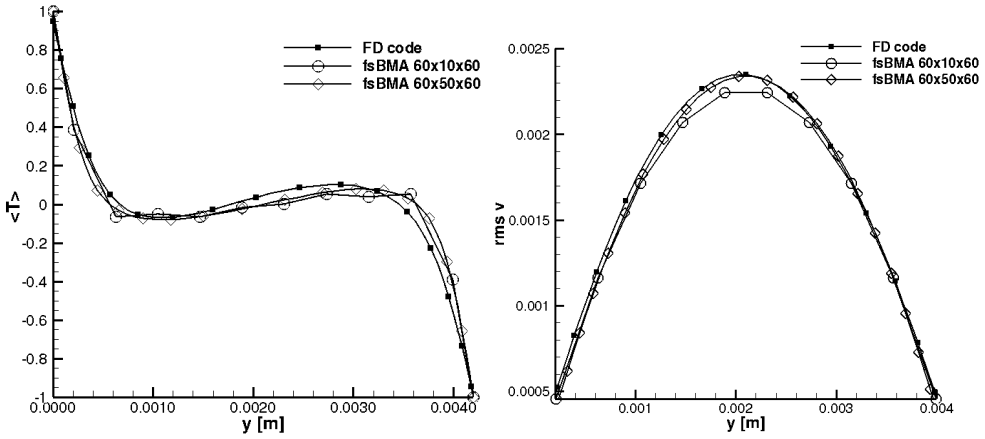


Figure 2: Comparison between the space-time average temperature profiles $< T >$, left plot, and between the root mean square of the U_y component, right plot, along the y axis, obtained with the `FD` code and the `fsBMA` solver.

The profiles in Figure 2 are referred to the 60x50x60 non-uniform mesh for the `FD` solver, and to the 60x10x60 and 60x50x60, uniform and non-uniform mesh, for the `fsBMA` solver. We observe that both the `fsBMA` space-time average profile $< T >$, and the root mean square of the U_y component are in good agreement with the `FD` results. Moreover, the comparison highlights the effects of the grid resolution. The finer non-uniform grid, indeed, allows us to obtain an

average temperature profile and a root mean square of U_y velocity that better reproduce the reference FD profile, suggesting a convergent trend of the curves increasing the grid resolution.

5.2 Laminar Rayleigh-Bénard convection for the ice-water and the ice-salt-water mixture

The numerical simulations have been carried out following the configuration adopted in the previous section, but now simulating the ice-water and the ice-water-salt mixture. Computations have been performed for both configurations on uniform 60x10x60 grid. We begin our analysis presenting the numerical results of the ice-water simulation. The simulation has been performed imposing on the top and bottom walls a zero gradient condition for the ice mass conservation. The rise velocity has been set $w_r = 0.01m/s$, according to the experimental investigation reported in [34], moreover the ice thermal diffusivity, the ice density and the particle average radius are, $k_I = 1.18 \times 10^{-6}\,[m^2/s]$, $\rho_I = 917\,[kg/m^3]$, $R = 0.5 \times 10^{-3}\,[m]$, respectively, assuming the frazil-ice particles of disk-like shape. We perform the numerical simulation imposing an initial random perturbation for the ice concentration. Figure 3 shows the behaviour of temperature and ice mass concentration on the x, y and x, z midplanes.

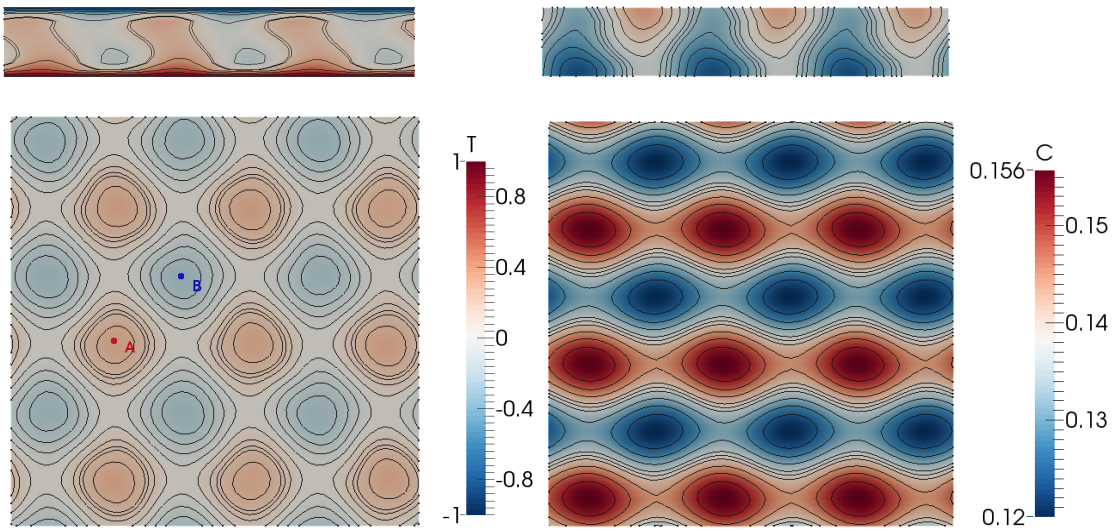


Figure 3: Contours and isolines of temperature, T, left column, and of the ice mass concentration, C, right column, in the x, y midplane, top row, and in the x, z midplane, bottom row, for the ice-water mixture at $Ra = 2.3 \times 10^4$.

Note that on the x, y midplane the distribution of frazil ice concentration follows the trend of the temperature field, showing its maximum values near the zones at lower temperature, whereas the behaviour of frazil ice concentration on the x, z midplane highlights the effects of the transport phenomena due to the U_x and U_y velocity components. The velocity fields on the x, y and x, z midplanes are shown in Figure 4, for each components. In Figure 5 we present the temperature and ice concentration profiles, evaluated along two different vertical lines, A, and B. The A, B lines are located on the x, z plane as shown in Figure 3, left plot.

Starting from the numerical solution obtained for the water-ice mixture, an initial uniform salinity distribution has then been imposed on the entire computational domain. The initial salinity field has been set equal to the reference value $S_0 = 34.5\ psu$, whereas the boundary condition for salinity at top and bottom walls are zero-gradient. The salt diffusivity and the expansion coefficient are $\alpha_s = 0.7 \times 10^{-9}\,[m^2/s]$ and $\beta_s = 7.86 \times 10^{-4}\,[1/psu]$, respectively.

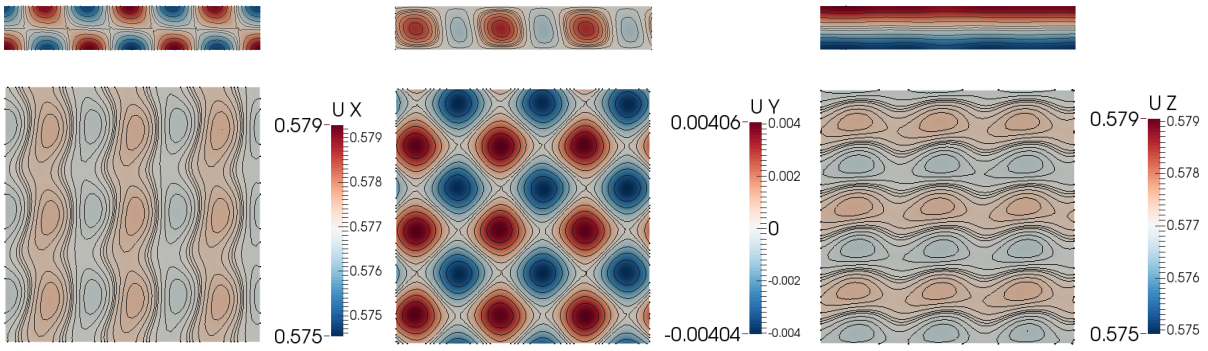


Figure 4: Contours and isolines of U_x, left column, U_y, middle column, and U_z, right column, in the x, y midplane, top row, and in the x, z midplane, bottom row, for the ice-water mixture at $Ra = 2.3 \times 10^4$.

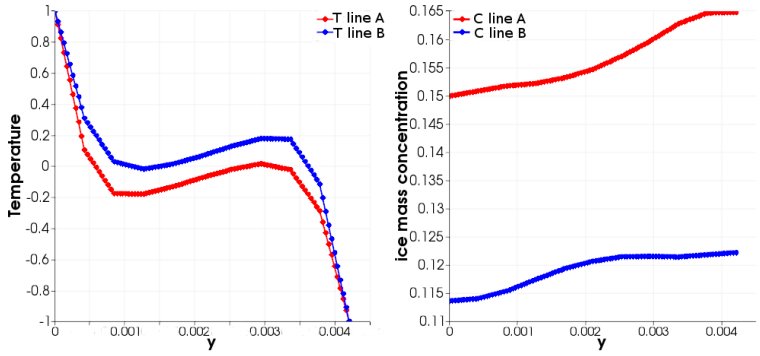


Figure 5: Temperature, left column, and ice mass concentration, right column, profiles along two different vertical lines A, and B, located on the x, z plane as in Figure 3, left plot.

In order to take into account the variation of the freezing temperature due to salinity and width, we define $T_f = T_{f_0} + ah + a_s\,(S - S_0)$, where a and a_s are the coefficients describing the decrease of T_f with the width, h, and salinity, respectively, and T_{f_0} denotes the freezing temperature of the salted water. The simulation has been performed using $a = 7.61 \times 10^{-4}\,[K/m]$, $a_s = -0.0573\,[K]$, and $T_{f_0} = 271.17\,[K]$. Here we report the numerical results obtained after about 30 hours of simulated time. Figure 6 shows the behaviour of temperature, ice mass concentration and salinity, on the x, y and x, z midplanes, whereas Figure 7 shows the velocity fields on the x, y and x, z midplanes, for each component. We observe from Figures 6 and 7, that on the x, y midplane the larger values of salinity and frazil ice concentration occur close to the top boundary, where temperature shows its minimum values, conversely the salinity and ice concentration reach their minimum values near to the bottom boundary, where the temperature increases. The behaviour of the salinity and the frazil-ice concentration on the x, z midplane highlights the dependence of the salinity on the temperature distribution, and the effects of the velocity field on the transport of ice particles. In Figure 8 we present the temperature, ice mass concentration, and salinity profiles, evaluated along two different vertical lines, C and D. The C, D lines are located on the x, z plane as shown in Figure 6, middle plot.

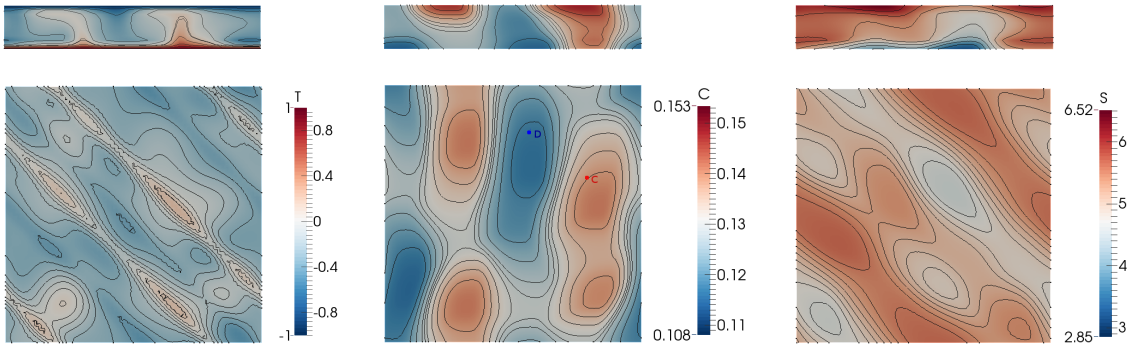


Figure 6: Contours and isolines of temperature, T, left column, ice mass concentration, C, middle column, and salinity, S, right column, in the x, y midplane, top row, and in the x, z midplane, bottom row, for the ice-water-salt mixture at $Ra = 2.3 \times 10^4$.

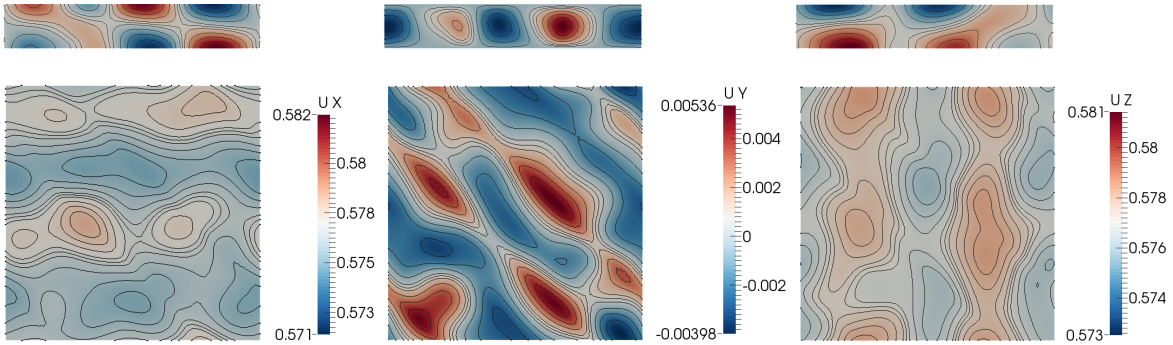


Figure 7: Contours and isolines of U_x, left column, U_y, middle column, and U_z, right column, in the x, y midplane, top row, and in the x, z midplane, bottom row, for the ice-water-salt mixture at $Ra = 2.3 \times 10^4$.

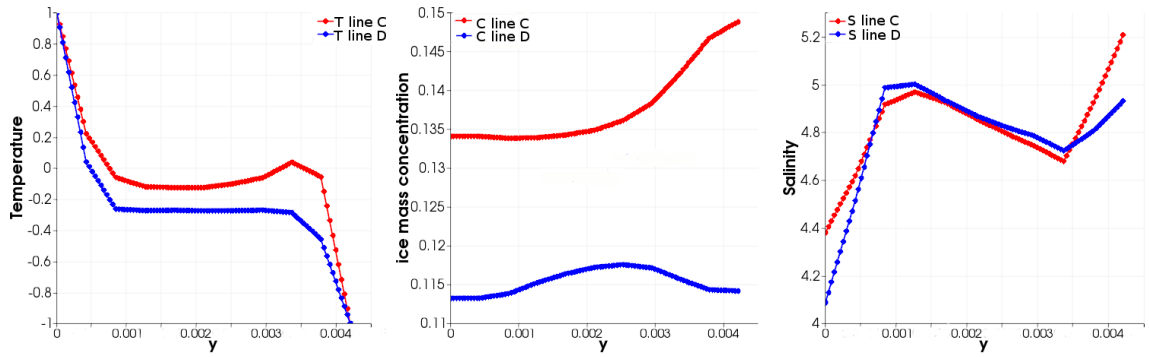
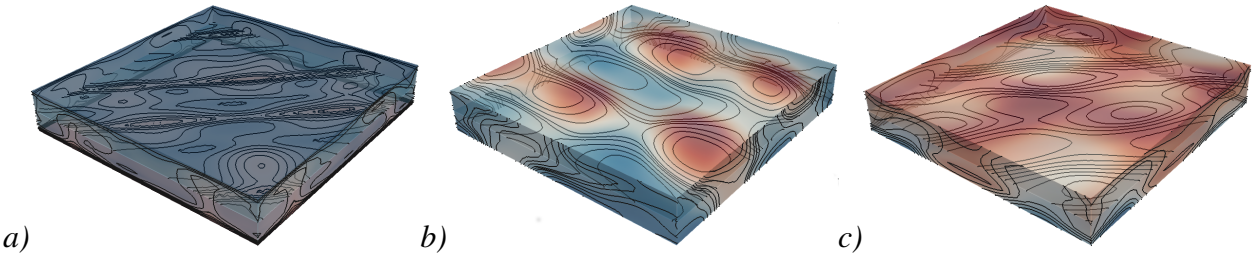


Figure 8: Temperature, left column, ice mass concentration, middle column, and salinity, right column, profiles along two different vertical lines C, and D, located on the x, z plane as in Figure 6, middle plot.

A more comprehensive representation of the convective structures is given in Figure 9, in which we present the three-dimensional contours and isolines of temperature, ice mass concentration and salinity. The isolines are reported on each lateral boundary and on the x, z midplane for $y = D/2$.



a) *b)* *c)*

Figure 9: Contours and isolines of temperature, T, a), ice mass concentration, C, b), and salinity, S, c), for the ice-water-salt mixture at $Ra = 2.3 \times 10^4$.

6 CONCLUSION AND FUTURE DEVELOPMENTS

The aim of this work was to improve the predictive capability of the numerical models currently available for the simulation of ice production in seawater, by means of the development of a sophisticated multiphase numerical model able to describe all the stages of ice production, overcoming the limitation of previous attempts, mainly based on the Boussinesq approximation.

We proposed a multiphase model that consider the mixture ice-seawater as a dense compressible fluid, and we modelled the behaviour of seawater density by means of an equation of state, that links density to temperature, pressure and salinity. In order to reproduce the interaction phenomena occurring between phases, we have included in the momentum equations additional terms, derived from the comprehensive modelling approach adopted for the simulation of multiphase flows of industrial interest. A low-Mach number asymptotic analysis has been performed to investigate the behaviour of the multiphase equations in the incompressible limit.

As an initial step of our study, we developed a finite volume numerical solver, by means of the OpenFOAM library, able to reproduce the behaviour of the initial stage of ice formation, according to the Boussinesq approximation. The solver is based on an approximate implicit projection method, according to the colocated version proposed in [4]. The performance of the solver has been assessed by means of the resolution of a Rayleigh-Bénard convection problem, for water and for the mixture ice-seawater. The numerical results obtained for water have been compared with those obtained with a finite difference numerical code.

Ongoing work is focused on the development of an advanced finite volume multiphase solver, for incompressible flows with variable density effects, according to the proposed multiphase approach. We expect that our multiphase model, coupled with the finite volume discretization, could provide an efficient and accurate numerical prediction of all the complex phenomena characterizing the ice production.

Acknowledgements

The present research has been founded by the Italian Ministry of Research and Education in the framework of the PNRA (Italian Program for Antarctic Research) project ”PANACEA The role of frazil and PANcake ice in the mass and energy budgets of the AntartiC sEA ice cover”. The authors gratefully acknowledge the support of Tenova S.p.A.

Appendix A Lagrangian derivative of the zero-Mach number equation of state

Here we present the details of the evaluation of the Lagrangian derivative of the seawater equation of state at order $O\left(M_\infty^0\right)$, used for the derivation of the divergence constraint for the water velocity, Equation 41.
The zero-Mach number equation for the seawater density is given by

$$\rho_w^{(0)} = D_r^* \left(D_r^{(0)} p^{(0)}\right) + E_r^* E_r^{(0)}, \tag{64}$$

where the coefficients are defined as

$$D_r^* = \frac{D_{r_\infty p_\infty}}{\rho_w^{(0)}}, \qquad E_r^* = \frac{E_{r_\infty}}{\rho_w^{(0)}}, \tag{65}$$

$$D_r = \left[b_5 S_\infty T_\infty S^{(0)} T_w^{(0)} + b_2 T_w^{(0)} T_\infty + b_1\right] / D_{r_\infty} \tag{66}$$

$$E_r = \left[a_5 S_\infty T_\infty S^{(0)} T_w^{(0)} + a_4 T_w^2 T_\infty^2 + a_3 S^{(0)} S_\infty + a_2 T_w^{(0)} T_\infty + a_1 + \rho_0\right] / E_{r_\infty} \tag{67}$$

As reported in section 2.2, the Lagrangian derivative of $\rho_w^{(0)}$ can be expressed as

$$\frac{D\rho_w^{(0)}}{Dt} = \left.\frac{\partial \rho_w^{(0)}}{\partial S^{(0)}}\right|_{T_W^{(0)}, p^{(0)}} \frac{DS^{(0)}}{Dt} + \left.\frac{\partial \rho_w^{(0)}}{\partial T_w^{(0)}}\right|_{S^{(0)}, p^{(0)}} \frac{DT_w^{(0)}}{Dt} + \left.\frac{\partial \rho_w^{(0)}}{\partial p^{(0)}}\right|_{T_W^{(0)}, S^{(0)}} \frac{Dp^{(0)}}{Dt}. \tag{68}$$

The partial derivative can be written as

$$\begin{aligned}
\left.\frac{\partial \rho_w^{(0)}}{\partial S^{(0)}}\right|_{T_W^{(0)}, p^{(0)}} &= \frac{1}{\rho_w^{(0)}} \left(p_\infty p^{(0)} b_5 S_\infty T_\infty T_w^{(0)} + a_5 S_\infty T_\infty T_w^{(0)} + a_3 S_\infty\right), \\
\left.\frac{\partial \rho_w^{(0)}}{\partial T_w^{(0)}}\right|_{S^{(0)}, p^{(0)}} &= \frac{1}{\rho_w^{(0)}} \left[p_\infty p^{(0)} \left(b_5 S_\infty T_\infty S^{(0)} + b_2 T_\infty\right) + a_5 S_\infty T_\infty S^{(0)} + 2 a_4 T_w T_\infty^2 + a_2 T_\infty\right], \\
\left.\frac{\partial \rho_w^{(0)}}{\partial p^{(0)}}\right|_{T_W^{(0)}, S^{(0)}} &= \frac{p_\infty}{\rho_w^{(0)}} \left(b_5 S_\infty T_\infty S^{(0)} T_w^{(0)} - b_2 T_w^{(0)} T_\infty + b_1\right).
\end{aligned}$$

and the Lagrangian derivative $\frac{DS^{(0)}}{Dt}$ and $\frac{DT_W^{(0)}}{Dt}$, can be derived from the zero-Mach number temperature and salinity equations, respectively, as

$$\begin{aligned}
\frac{DS^{(0)}}{Dt} = \frac{1}{\left(1 - \phi_I^{(0)}\right)} \Bigg[& \frac{\partial}{\partial x_i} \left(\left(1 - \phi_I^{(0)}\right) \frac{1}{ScRe_\infty} \frac{\partial S^{(0)}}{\partial x_i}\right) + \\
& S_{sal}^{(0)} - S^{(0)} \left(\frac{\partial}{\partial t}\left(1 - \phi_I^{(0)}\right) + \frac{\partial}{\partial x_i}\left(1 - \phi_I^{(0)}\right) u_i^{(0)}\right)\Bigg]
\end{aligned} \tag{69}$$

$$\begin{aligned}
\frac{DT_w^{(0)}}{Dt} = \frac{1}{\left(1 - \phi_I^{(0)}\right)} \Bigg[& \frac{t^* \nu}{L^2} \frac{\partial}{\partial x_i} \left[\left(1 - \phi_I\right) \frac{1}{Pr} \frac{\partial T_w^{(0)}}{\partial x_i}\right] + \phi_I^{(0)} t^* \lambda_t^{(0)} \left(T_f - T_w^{(0)}\right) + \\
& S_T^{(0)} - T_w^{(0)} \left(\frac{\partial}{\partial t}\left(1 - \phi_I^{(0)}\right) + \frac{\partial}{\partial x_i}\left(1 - \phi_I^{(0)} u_i^{(0)}\right)\right)\Bigg]
\end{aligned} \tag{70}$$

In order to evaluate the Lagrangian derivative of pressure at zero-order, $Dp^{(0)}/Dt$, we consider the momentum equation at the leading order $O\left(M_\infty^{-2}\right)$. Since $p^{(0)}$ represent the hydrostatic pressure, for which $\partial p^{(0)}/\partial z = -\rho_0 g$, see Equation 35, and considering $\partial p^{(0)}/\partial t = 0$, the Lagrangian derivative of the seawater density at order $O\left(M_\infty^0\right)$, can then be rewritten as

$$\frac{D\rho_w^{(0)}}{Dt} = \frac{1}{\rho_w^{(0)}\left(1-\phi_I^{(0)}\right)}\left(p_\infty p^{(0)} b_5 S_\infty T_\infty T_w^{(0)} + a_5 S_\infty T_\infty T_w^{(0)} + a_3 S_\infty\right)\left[\frac{\partial}{\partial x_i}\left((1-\phi_I^{(0)}\right)\frac{1}{ScRe_\infty}\frac{\partial S^{(0)}}{\partial x_i}\right) + S_{sal}^{(0)} - S^{(0)}\left(\frac{\partial}{\partial t}\left(1-\phi_I^{(0)}\right) + \frac{\partial}{\partial x_i}\left(1-\phi_I^{(0)}\right)u_i^{(0)}\right)\right] + \left(b_5 S_\infty T_\infty S^{(0)} T_w^{(0)} - b_2 T_w^{(0)} T_\infty + b_1\right)\left[\frac{t^*\nu}{L^2}\frac{\partial}{\partial x_i}\left[\left(1-\phi_I\right)\frac{1}{Pr}\frac{\partial T_w^{(0)}}{\partial x_i}\right] + \phi_I^{(0)} t^* \lambda_t^{(0)}\left(T_f - T_w^{(0)}\right) + S_T^{(0)} - T_w^{(0)}\left(\frac{\partial}{\partial t}\left(1-\phi_I^{(0)}\right) + \frac{\partial}{\partial x_i}\left(1-\phi_I^{(0)}\right)u_i^{(0)}\right)\right] - \frac{p_\infty}{\rho_w^{(0)}}\left(b_5 S_\infty T_\infty S^{(0)} T_w^{(0)} - b_2 T_w^{(0)} T_\infty + b_1\right) u_i^{(0)} \rho_0 g. \quad (71)$$

REFERENCES

[1] A. Abbà, P. Olla, L. Valdettaro, Numerical Analysis of Frazil Ice Formation in Turbulent Convection, *Progress in Turbulence VI, Proceedings of the iTi Conference on Turbulence, Springer Proceedings in Physics*, 165, 2014.

[2] R.A. Bagnold, Experiments on a gravity-free dispersion of large solid spheres in a Newtonian-fluid under shear, *Proceedings of the Royal Society of London A*, A225, pp. 49-63, 1954.

[3] J.B. Bell, P. Colella, H.M. Glaz, A second-order projection method for the incompressible Navier-Stokes equations, *J. Comput. Phys.*, 85, pp. 257-283, 1989.

[4] J.B. Bell, D.L. Marcus, A second-order projection method for variable-density flows, *J. Comput. Phys.*, 101, pp. 334-348, 1992.

[5] T.Benacchio, W.P. O'Neill, R. Klein, A Blended Soundproof-to-Compressible Numerical Model for Small- to Mesoscale Atmospheric Dynamics, *Mon. Wea. Rev.*, 142, pp. 4416-4438, 2014.

[6] K. Bryan, M.D. Cox, An approximate equation of state for numerical models of ocean circulation., *J. Phys. Oceanogr.*, 2.4, pp. 510-514, 1972.

[7] D. Brydon, S. Sun, R. Bleck, A new approximation of the equation of state for seawater, suitable for numerical ocean models, *JGR-Oceans (1978-2012)*, 104.C1, pp. 1537-1540, 1999.

[8] S. Carcano, L. Bonaventura, A. Neri, T. Esposti Ongaro, A second order accurate numerical model for multiphase underexpanded volcanic jets, *Geophysical Model Development*, 6, pp. 1905- 1924, 2013.

[9] S. Carcano, T. Esposti Ongaro, L. Bonaventura, A. Neri, Influence of grain-size distribution on the dynamics of underexpanded volcanic jets, *J. Volcanol. Geotherm. Res.*, 285, pp. 60-80, 2014.

[10] S. Chandrasekhar, Hydrodynamic and hydromagnetic stability, *Clarendon Press Oxford* 196.1, 1970.

[11] J.T. Cornelissen, F. Taghipour, R. Escudié, N. Ellis, J.R. Grace, CFD modelling of a liquid–solid fluidized bed, *Chem. Eng. Sci.*, 62.22, 6334-6348, 2007.

[12] C.T. Crowe, *Multiphase flows with Droplets and Particles*, CRC Press, Boca Raton, FL., 1998.

[13] J.H. Ferziger, M. Peric, Computational methods for fluid dynamics, *Springer*, 2002.

[14] N.P. Fofonoff, Physical properties of seawater: A new salinity scale and equation of state for seawater, *JGR-Oceans (1978-2012)*, 90.C2, pp. 3332-3342, 1985.

[15] H. Friedrich, S. Levitus, An approximation to the equation of state for sea water, suitable for numerical ocean models., *J. Phys. Oceanogr.*, 2.4, pp. 514-517, 1972.

[16] D. Gidaspow, *Multiphase Flow and Fluidization*, Academic Press, San Diego, CA, 1994.

[17] D. Gidaspow, R. Bezburuah, J. Ding, *Hydrodynamics of circulating fluidized beds, kinetic theory approach*, Potter, O.E., Nicklin, D.J. (Eds.), Fluidization VII. Engineering Foundation, New York, 1992.

[18] J.L. Guermond, P. Minev, J. Shen, An overview of projection methods for incompressible flows, *Computer Methods in Applied Mechanics and Engineering*, 195, pp. 6011-6045, 2006.

[19] F. Ham, G. Iaccarino, Energy conservation in colocated discretization schemes on unstructured meshes, *Ann. Rev. Brief 2004, CTR, NASA, Ames/Stanford Un.*, 2004.

[20] F. Ham, G. Iaccarino, Automatic mesh generation for LES in complex geometries, *Ann. Rev. Brief 2004, CTR, NASA, Ames/Stanford Un.*, 2005.

[21] P.R. Holland, D.L. Feltham, Frazil dynamics and precipitation in a water column with depth-dependent supercooling, *J. Fluid Mech.*, 530, pp. 101-124, 2005.

[22] S. Hosseini, D. Patel, F. Ein-Mozaffari, M. Mehrvar, Study of Solid-Liquid Mixing in Agitated Tanks through Computational Fluid Dynamics Modeling, *Ind. Eng. Chem. Res*, 49.9, pp. 4426-4435, 2010.

[23] D. Jackett, R. Mcdougall, J. Trevor, Minimal adjustment of hydrographic profiles to achieve static stability, *J. Atmos. Oceanic Technol.*, 12.2, pp. 381-389, 1995.

[24] A. Jenkins, A. Bombosch, Modeling the effects of frazil ice crystals on the dynamics and thermodynamics of ice shelf water plumes, *JGR-Oceans (1978-2012)*, 100.C4, pp. 6967-6981, 1995.

[25] J.J. Jenkins, S.B. Savage, A theory for the rapid flow of identical smooth nearly elastic spherical particles, *J. Fluid Mech.*, 130, pp. 187-202, 1983.

[26] P.C. Johnson, R. Jackson, Frictional-collisional constitutive relations for granular materials, with application to plane shearing, *J. Fluid Mech*, 176, pp. 67-93, 1987.

[27] D.R. Kaushal, T. Thinglas, Y. Tomita, S. Kuchii, H. Tsukamoto, CFD modeling for pipeline flow of fine particles at high concentration, *Int. J. Multiph.*, 43, pp. 85-100, 2012.

[28] H.R. Kivisild, *River and lake ice terminology*, International Association for Hydraulic Research, paper 1, 1970.

[29] S. Levitus, G. Isayev, Polynomial approximation to the international equation of state for seawater, *J. Atmos. Ocean. Tech.*, 9.5, pp. 705-708, 1992.

[30] C.K.K. Lun, S.B. Savage, N. Chepurning, Kinetic theories for granular flow: Inelastic particles in Couette flow and singly inelastic particles in a general flow field, *J. Fluid Mech.*, 140, pp. 223-256, 1984.

[31] S. Martin, Frazil ice in rivers and oceans, *Annu. Rev. Fluid Mech.*, 13, pp. 379-397, 1981.

[32] G.L. Mellor, An equation of state for numerical models of oceans and estuaries., *J. Atmos. Oceanic Technol.*, 8.4, pp. 609-611, 1991.

[33] F.J. Millero, A. Poisson, International one-atmosphere equation of state of seawater, *Deep Sea Res. Part A. Oceanogr. Res. Pap.*, 28.6, pp. 625-629, 1981.

[34] B. Morse, M. Richard, A field study of suspented frazil ice particles. *Cold Regions Science and Technology*, 2009.

[35] S.A. Ogawa, A. Umemura, N. Oshima, On the equations of fully fluidized granular materials, *J. Applied Math. Phys. ZAMP*, 31.4, 483-493, 1980.

[36] A. Omstedt, U. Svensson, Modeling supercooling and ice formation in a turbulent Ekman layer, *JGR-Oceans (1978-2012)*, 89.C1, pp. 735-744, 1984.

[37] S. Roy, M.P. Dudukovic, Flow mapping and modeling of liquid-solid risers, *Ind. Eng. Chem. Res.*, 40.23, 5440-5454, 2001.

[38] S.B. Savage, D.J. Jeffrey, The stress tensor in a granular flow at high shear rates, *J. Fluid Mech*, 110, pp. 255-272, 1981.

[39] D. Shi, Z. Luo, Z. Zheng, Numerical simulation of liquid-solid two-phase flow in a tubular loop polymerization reactor, *Powder Technol.*, 198.1, pp. 135-143, 2010.

[40] E.D. Skyllingstad, D.W. Denbo, Turbulence beneath sea ice and leads: A coupled sea ice/large-eddy simulation study, *JGR-Oceans (1978–2012)*, 106.C2, pp. 2477-2497, 2001.

[41] L.H. Smedsrud, A. Jenkins, Frazil ice formation in an ice shelf water plume, *JGR-Oceans (1978–2012)*, 109.C3, 2004.

[42] U. Svensson, A. Omstedt, Numerical simulations of frazil ice dynamics in the upper layers of the ocean, *Cold Reg. Sci. Technol.*, 28.1, pp. 29-44, 1998.

[43] M. Syamlal, W. Rogers, T.J. OBrien, *MFIX documentation: Theory guide*, National Energy Technology Laboratory, Department of Energy, Technical Note DOE/METC-95/1013 and NTIS/DE95000031, 1993.

[44] B.G.M. van Wachem, J.C. Schouten, C.M. van den Bleek, R. Krishna, J.L. Sinclair, Comparative analysis of CFD models of dense gas solid systems, *Journal. Amer. Inst. Chem. Engrs.*, 47, pp. 1035-1051, 2001.

[45] C.Y. Wen, Y.H. Yu, Mechanics of fluidization, *Chem. Eng. Prog. Symp.*, 62, pp. 100-111, 1966.

[46] W. Yan, D. Shi, Z. Luo, Y. Lu, Three-dimensional CFD study of liquid–solid flow behaviors in tubular loop polymerization reactors: The effect of guide vane, *Chem. Eng. Sci.*, 66.18, pp. 4127-4137, 2011.

[47] H.G. Weller, G. Tabor, H. Jasak, C. Fureby, A tensorial approach to computational continuum mechanics using object orientated techniques, *Computers in Physics*, 12.6, pp. 620–631, 1998.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A 3D CFD NUMERICAL STUDY OF THE BUBBLE GENERATION PROCESS INTO A BUBBLE T-JUNCTION GENERATOR AND ITS COMPARISON WITH EXPERIMENTAL DATA: PART I

S. Arias[1] **and A. Montlaur**[1,2]

[1]Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels, Universitat Politècnica de Catalunya - BarcelonaTech
c/ Esteve Terradas 5, 08860, Castelldefels (Barcelona), Spain
e-mail: santiago.arias@upc.edu, adeline.de. montlaur@upc.edu

[2] Laboratori de Càlcul Numèric, Universitat Politècnica de Catalunya, www-lacan.upc.edu

Keywords: Two-phase flows, Bubble generation, T-junction, Numerical simulation, Computational Fluid Dynamics (CFD), Volume of Fluid

Abstract. *This work presents a 3D numerical study of the bubble generation process into a bubble generator obtained with the commercial Computational Fluid Dynamics solver ANSYS Fluent v15.0.7, and its comparison with experimental data reproducing the same conditions [1]. The bubble generator is formed by two perpendicular capillaries in which liquid and gas are injected at perpendicular directions into a 1 mm internal diameter capillary T-junction with a total length of 10 mm. The fluids used in experiments and CFD simulations are air and water, both of them considered incompressible and isothermal, at a room temperature of 25°. A total of 23 different cases are studied for different injection conditions, and results between numerical simulations and experiments are compared.*

In this first part of the analysis, we focus on the flow pattern regimes and the dynamics of the bubble generation process. In addition to the new numerical simulations presented here, a new model has been used to predict the bubble generation frequency and tested with both experimental and numerical data. Results on bubble generation frequency are also presented by means of the non-dimensional Strouhal number. Same types of patterns, bubble and slug flow regimes, are obtained in simulations and experiments. In order to perform an exhaustive validation and comparison of numerical simulations with experimental data, several parameters have been selected: bubble velocity, volumetric void fraction, bubble generation frequency, Strouhal number and bubble equivalent diameter. Numerical simulations agree qualitatively, but not always quantitatively, with experimental results.

1 INTRODUCTION

Capillary gas-liquid two-phase flow occurs in increasingly more modern industrial applications and in particular in space-based systems. Two-phase systems present an improvement in performance as well as significant reductions in weight in different fields, such as power generation and life support systems, with respect to one-phase systems. A full understanding of the behaviour of the gas-liquid interfaces and flow characteristics, such as flow patterns and flow patterns transitions, is thus a major necessity in the development of these technologies.

Numerous experimental studies have been made on the generation of bubbles and droplets in T-junction bubble generators as the one studied in this paper [2, 3]. In the recent past, different Computational Fluid Dynamics (CFD) methods have also been used in order to study similar devices. A commercial CFD package was used to simulate the squeezing regime during the bubble formation in a T-junction microchannel [4], focusing on studying the effects of pressure, surface tension and shear stress action on the gas thread. CFD modelling aspects of internal circulation and slug flow generation have been discussed, focusing on the slug flow formation in a 120° Y-junction and the velocity profiles inside the slug [5]. A numerical investigation by means of a phase-field model of the breakup dynamics of streams of immiscible fluids in a microfluidic T-junction was carried out [6], where three regimes of formation of droplets (squeezing, dripping and jetting) were identified and studied.

More recently, a fluid dynamics numerical study of the formation of mini-bubbles in a 2D T-junction was presented [1], and results on the behaviour of bubble velocity, void fraction, bubble generation frequency and characteristic lengths were obtained and compared to experimental data. Also using a two-dimensional numerical model with Volume of Fluid method (VOF), alternating droplet formation, with applications such as nanoparticle synthesis, hydrogel bead generation, and cell transplantation in biomedical therapy, was studied [7]. The two-dimensional bubble formation process was analysed in detail by focusing on the effects of three main parameters, capillary number, contact angle and the gas-liquid flow rate ratio [8]. As for 3D simulations, droplet formation in T-junction microchannels was studied using the Lattice Boltzmann method [9] and simulations in heat exchangers using the VOF method were performed [10].

The paper presented here is a continuation of a previous study [1], completing it and extending it to 3D. Some parameters are now studied in more depth, such as the bubble generation frequency, a new model of frequency being developed here, and new values of interest are presented, such as the bubble velocity, volume void fraction, Strouhal number, bubble volume and the equivalent diameter. Section 2 presents the T-junction problem statement. Section 3 summarises the methodology used in the experimental study, following the one used in [1], as well as the hypotheses used in the numerical simulations and their validations. Section 4 exposes the results of the numerical simulations and compares them to the experimental ones. Finally, conclusions are drawn in section 5.

2 PROBLEM STATEMENT

The presented study aims to simulate a bubble generator in which liquid and gas are injected at perpendicular directions into a 1 mm internal diameter capillary T-junction with a total length of 10 mm, see figure 1. In order to compare numerical simulations with realistic results, CFD simulations have been carried out in the most similar conditions possible to the experiments. The fluids used in experiments and CFD simulations are air and water (incompressible, isothermal, at constant room temperature of 25°C). Standard physical values are assumed for

air (
visc

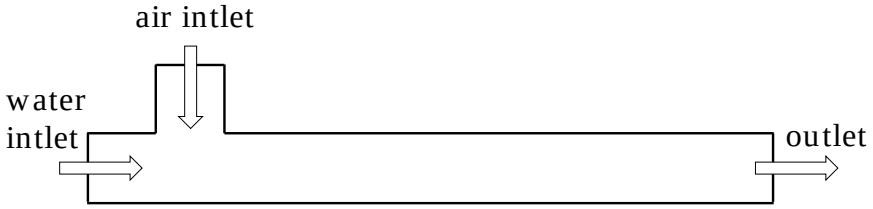


Figure 1: Sketch of T-junction.

The experimental setup used in this study [11, 12] can control the gas and liquid volumetric flow rates (Q_G and Q_L, respectively) in a very accurate way. Results are presented as a function of the gas and liquid superficial velocities defined as $U_{SG} = Q_G/A$ and $U_{SL} = Q_L/A$, respectively, where A is the capillary cross-sectional area.

Constant gas and liquid superficial velocities ($Q_G = 0.062 - 0.473$ m/s and three different $Q_L = 0.106, 0.318, 0.531$ m/s, respectively) were considered in each experimental run. A total of 23 experiments were compared with the results obtained in numerical simulations. The same injection conditions were used in the simulations. Additionally, three more simulations corresponding to the value $U_{SG} = 0.025$ m/s (and $Q_L = 0.106, 0.318, 0.531$ m/s) were conducted to obtain higher precision at very low superficial gas velocity than provided by experiments.

According to the Bond number, gravity played a smaller role than surface tension in the bubble generation process ($\rho_L g \phi_c^2/\sigma$=0.139, where g is the gravitational acceleration and ϕ_c the capillary internal diameter), resulting in the irrelevance of the channel orientation [13]. Therefore, the present analysis can be considered gravity independent and assuming $g = 0$ m/s^2 becomes a valid approach for numerical simulations.

In the considered experimental conditions, the formation and detachment of bubbles is mainly dominated by the competition between the capillary and the liquid drag forces, which can be evaluated by the Weber number ($We_G = \rho_G \phi_c {U_G}^2/\sigma$, where U_G is the gas (bubble) velocity). Capillary forces overcome the inertial forces when the criterion $We_G < 2$ is accomplished [14], and only bubble and slug flow regimes are expected to be observed in those circumstances. In our experiments, We_G ranges from 4.3×10^{-4} to 2.2×10^{-2} ($U_G = 0.159 - 1.143$ m/s, see section 4.1), corresponding to the surface tension controlled region.

The Reynolds number, based on the average mixture superficial velocity ($U_M = U_G + U_L$) and defined as $Re_M = \rho_L \phi_c U_M/\mu_L$, ranged from 167 to 1000 ($U_M = 0.167 - 1$ m/s in experiments), and hence we assume laminar conditions.

Having used the same fluids physical properties as well as the same values of gas and liquid superficial velocities in experiments and simulations, the bubble generation behaviour in numerical simulations is expected to be similar to the one observed in experiments.

3 METHODOLOGY

3.1 Experiments

Experiments were performed at constant liquid superficial velocity while increasing the gas superficial velocity. The generation of bubbles was observed to be regular and periodic for all gas and liquid superficial velocities considered in this work. Images were taken with a high-speed camera at 4000 f.p.s. or each pair of U_{SG} and U_{SL} values. These images were classified

into bubble or slug flow regimes [15]. Bubble generation frequency was measured by counting the total number of bubbles over a period of time larger than the time required to form a single bubble (typically three orders of magnitude smaller). The bubble velocity was measured directly over the calibrated images, by taking into account the displacement of the foremost part of a single bubble and the time taken to carry out such a movement.

3.2 Numerical simulations

3.2.1 Flow solver

All numerical simulations were carried out with a commercial CFD solver, ANSYS Fluent v15.0.7. The VOF model was used to model the two immiscible fluids by solving a single set of momentum equations and tracking the volume fraction of each fluid throughout the domain. In each control volume, the volume fractions of both phases sum to unity.

The tracking of the interface between the phases is accomplished by the solution of a continuity equation for the volume fraction of one of the phases [16]. For the q^{th} phase, this equation has the following form:

$$\frac{1}{\rho_q}\left[\frac{\partial}{\partial t}(\alpha_q \rho_q) + \nabla \cdot (\alpha_q \rho_q \mathbf{v_q}) = \sum_{\mathbf{p=1}}^{\mathbf{n}} (\dot{\mathbf{m}}_{\mathbf{pq}} - \dot{\mathbf{m}}_{\mathbf{qp}})\right] \tag{1}$$

where ρ_q, α_q and $\mathbf{v_q}$ are the density, volume void fraction and velocity of phase q, respectively, and $\dot{m}_{qp}$ is the mass transfer from phase q to phase p and $\dot{m}_{pq}$ is the mass transfer from phase p to phase q. The volume fraction equation is not solved for the primary phase; the primary-phase volume fraction is computed based on the following constraint: $\sum_{q=1}^{n} \alpha_q = 1$.

The fields for variables and properties are shared by the phases and represent volume-averaged values, as long as the volume fraction of each of the phases is known at each location. Thus the variables and properties in any given cell are either purely representative of one of the phases, or representative of a mixture of the phases, depending on the volume fraction values. That is, if the water volume fraction in the cell is denoted as a_w, then three conditions are possible. If $a_w = 0$, the cell is empty of water, if $a_w = 1$, the cell is full of water, and if $0 < a_w < 1$, the cell contains the interface between water and air [16]. Based on the local value of a_w, the appropriate properties and variables are assigned to each control volume within the domain.

An explicit formulation is used for the VOF model, with a Courant number of 0.25, and the PRESTO! (PREssure STaggering Option) scheme is used for pressure interpolation. This scheme uses the discrete continuity balance for a staggered control volume about the face to compute the staggered pressure. This procedure is similar in spirit to the staggered-grid schemes used with structured meshes [16]. For the volume fraction spatial discretisation scheme, the geometric reconstruction scheme, which represents the interface between fluids using a piecewise-linear approach is used. In Fluent this scheme is the most accurate and is applicable for general unstructured meshes. The geometric reconstruction scheme is generalised for unstructured meshes from the work of [17].

The VOF model also includes here the effects of surface tension along the interface between each pair of phases. The surface tension model in Fluent is the continuum surface force model proposed by [18]. With this model, the addition of surface tension to the VOF calculation results in a source term in the momentum equation. As commented in section 2, a surface tension of 0.072 N/m is applied.

The following boundary conditions were set for the numerical simulations. All walls are

number of elements	U_G [m/s]	ϵ_{U_G} [%]	f [1/s]	ϵ_f [%]	V_B [$\times 10^{-10}$m^3]	ϵ_{V_B} [%]
220000	0.649	7.1	266.7	14.7	7.31	8.1
413000	0.625	3.1	238.1	2.4	7.82	1.6
879000	0.606	–	232.6	–	7.95	–

Table 1: Study of mesh convergence for three meshes and three independent parameters of study.

treated as no-slip smooth walls. A wall adhesion angle in conjunction with the surface tension model is also applied in the VOF model [18]. Rather than imposing this boundary condition at the wall itself, the contact angle that the fluid is assumed to make with the wall is used to adjust the surface normal in cells near the wall [16]. This so-called dynamic boundary condition results in the adjustment of the curvature of the surface near the wall. The contact angle is the angle between the wall and the tangent to the interface at the wall. A contact angle of 0° was set up as wall adhesion (hydrophilic condition) on the horizontal walls, and of 25° on the vertical walls, based on experimental data [1].

Water and air inlets are considered as velocity inlets, with the corresponding fluid velocity in each case of study, namely U_{SL} and U_{SG}. At the water inlet, the fraction of water is considered as 1 and at the air inlet, the fraction of air is the one set to 1. The outlet of the T-junction is set up as a pressure outlet.

An operating pressure of 101325 Pa is set up at a point close to the air inlet, that is, at a point where there will always be only air. Finally the gravity value is set to 0, as explained in section 2.

3.2.2 Mesh

The computational mesh was composed of 413000 elements. An inflation mesh comprising 15 layers was used along all the walls, with a first layer height of 0.005 mm. A body sizing of 0.045 mm was used in all the domain. The resulting mesh is a mix of hexahedral (close to the walls) and tetrahedral (in the core of the domain) elements. The hex dominant option has been chosen, resulting in a fewer number of elements than if mainly tetrahedral elements had been considered. Mesh independence is presented in section 3.2.3.

A first-order implicit scheme is used for the transient formulation. Time step independence is also presented in section 3.2.3. In order to obtain a converged solution, the calculation was always continued until at least 7 or 8 bubbles of air were generated.

3.2.3 CFD validation

The mesh convergence is first checked. Three meshes are considered: the coarsest one of 220000 elements, the second one obtained by multiplying the body sizing and first layer height of the boundary layer by 0.75, of 413000 elements, and the third one, obtained by multiplying again the mesh dimensions by 0.75, of 879000 elements. Three independent parameters of study are shown, which will be later used in section 4: the bubble velocity U_G, the bubble frequency f and the bubble volume V_B. Absolute values are shown, as well as the percentages of error with respect to the finest mesh, ϵ, in table 1, for $U_{SL} = 0.318$ m/s and $U_{SG} = 0.242$ m/s.

These results show that for the 413000 element mesh, errors due to the mesh remain within a relatively small margin ($\leq 3.1\%$); this is thus the mesh that has been chosen for the numerical study.

Δt [s]	U_G [m/s]	ϵ_{U_G} [%]	f [1/s]	ϵ_f [%]	V_B [$\times 10^{-10}$m^3]	ϵ_{V_B} [%]
10×10^{-6}	0.633	1.9	243.9	3.5	8.04	5.0
5×10^{-6}	0.625	0.6	238.1	1.1	7.82	2.1
2.5×10^{-6}	0.621	–	235.6	–	7.66	–

Table 2: Study of time convergence for three Δt and three independent parameters of study.

Next, the influence of the time step has been studied considering the chosen mesh of 413000 elements. Three values of Δt are considered: $\Delta t = 2.5 \times 10^{-6}$ s, $\Delta t = 5 \times 10^{-6}$ s and $\Delta t = 10 \times 10^{-6}$ s. Again absolute values of velocity, bubble frequency and volume are shown in table 2, as well as the percentages of error with respect to the smallest Δt, for $U_{SL} = 0.318$ m/s and $U_{SG} = 0.242$ m/s.

Table 2 shows that the three values of Δt lead to similar values (all errors $\leq 5.0\%$), smaller for the value of $\Delta t = 5 \times 10^{-6}$ s ($\leq 2.1\%$) than for the value $\Delta t = 10 \times 10^{-6}$ s. Therefore, the middle value, $\Delta t = 5 \times 10^{-6}$ s, has been chosen for the rest of the CFD study.

3.2.4 CFD metrics

Figure 2 shows an example of data obtained as a post-process of the CFD simulations. It shows the propagation of the bubble, through the fraction of air at two cross-sections of the domain located at 7 and 8 mm from the beginning of the tube, as a function of time. This allows to calculate the bubble frequency f, as the inverse of the time T_f between two bubbles at the cross-section, and its velocity U_G, as the distance between two cross-sections (here 7 and 8 mm) divided by the time T_S needed for the bubble to travel this distance. Its volume V_B is also calculated from this figure, integrating the total fraction of air of a bubble over time A_B and multiplying by its velocity and by the area of a 1 mm diameter circle. The volume void fraction, α, was estimated for both the experiments and the CFD simulations as the ratio U_{SG}/U_G [19].

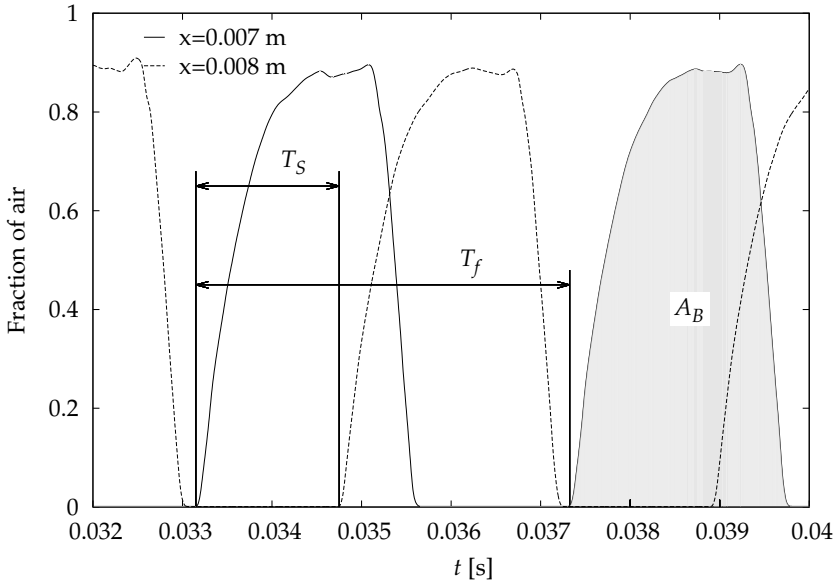


Figure 2: Example of CFD post-processed results from the graph of fraction of air as a function of time, for $U_{SL} = 0.318$ m/s and $U_{SG} = 0.242$ m/s.

4 RESULTS AND DISCUSSION

A set of numerical simulations are compared with experimental data. Numerical simulations satisfactorily reproduce trains of bubbles as the ones obtained in experiments. As expected, only two flow patterns, the bubble and the slug flow regimes, were obtained in the simulations and in the experiments [13, 14, 20]. In both flow regimes the gas phase is dispersed into the liquid phase. Spherical or almost spherical bubbles with a diameter smaller than the capillary diameter were classified as bubble flow. Larger bubbles were classified as slug flow [15].

Trains of bubbles are shown in figs. 3-5 at the same stage of formation, which do not correspond to the same times as detailed in section 4.2. In contrast with a previous study [1], no artificial pinch-off mechanism has been used in numerical simulations to detach bubbles. Bubbles were generated, both in experiments and in numerical simulations, with high regularity and small size dispersion. Fig. 3 corresponds to the formation of a bubble belonging to the bubble flow regime. In this figure, bubbles are not exactly spherical. This is a consequence of the liquid drag force, which noticeably deforms bubbles longitudinally. Figs. 4 and 5 show two examples of slug flow regime, the former being known as short Taylor bubble and the latter as middle Taylor bubble [21]. Bubbles from the slug flow regime are characterised by having a bullet-shaped body, rounded at the front part and mainly flattened at the rear of it. In every case, bubbles have an elongated rear part just after the detachment, this elongated shape being a remnant of the squeezed gas thread that was connecting the new bubble with the single-phase gas. From that moment, the effect of the surface tension tries to minimise the total surface of the bubble, which results in the vibration of the bubble surface until it stabilises. It can be seen that numerical simulations correctly reproduce the elasticity and fluctuation of the gas-liquid interface during the bubble detachment stage described here.

Next sections focus in a more exhaustive validation and comparison of numerical simulations with experimental data. Several parameters are selected to perform such a validation, namely, bubble velocity, volumetric void fraction, bubble generation frequency and bubble equivalent diameter.

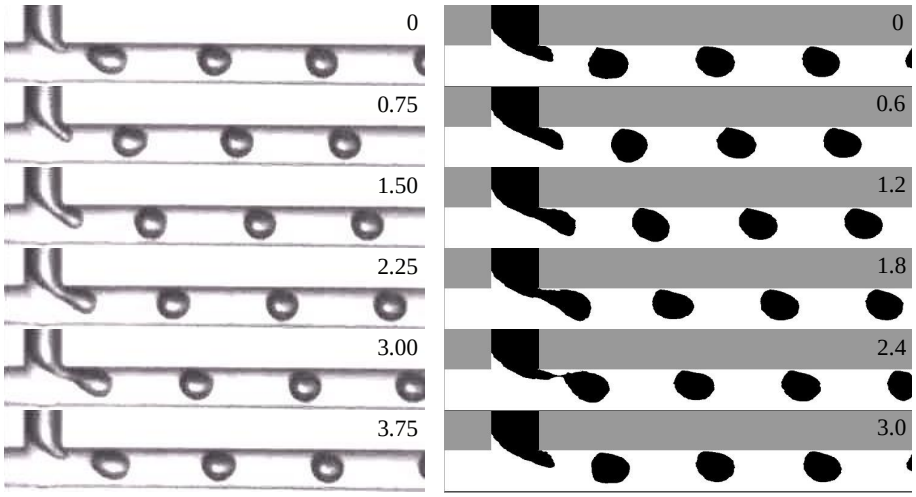


Figure 3: Bubble flow in (*left*) experiments and (*right*) numerical simulations. $U_{SL} = 0.531$ m/s and $U_{SG} = 0.110$ m/s. Time (ms) is indicated in the upper right corner.

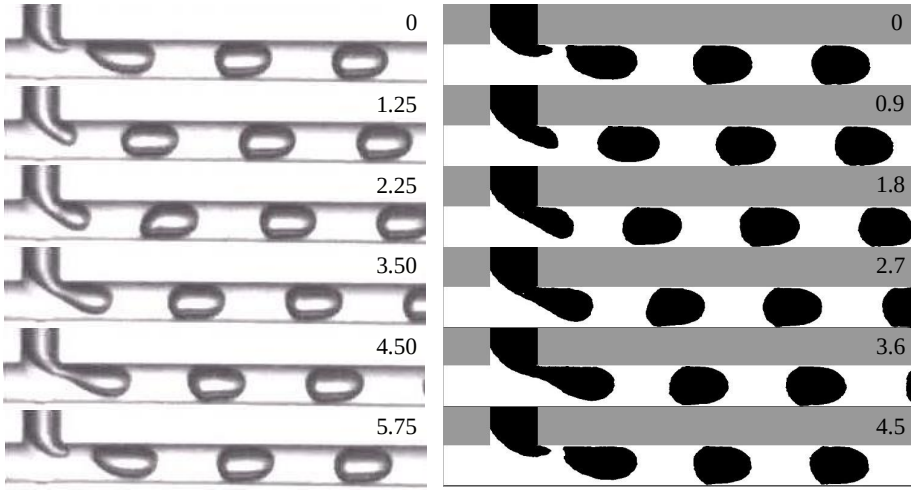


Figure 4: Slug flow in (*left*) experiments and (*right*) numerical simulations. $U_{SL} = 0.318$ m/s and $U_{SG} = 0.182$ m/s. Time (ms) is indicated in the upper right corner.

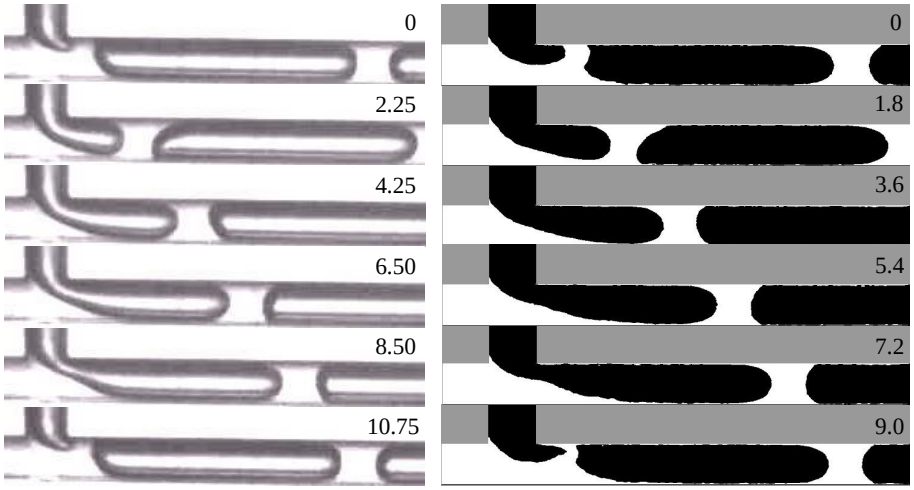


Figure 5: Slug flow close to the slug-churn transition in (*left*) experiments and (*right*) numerical simulations. $U_{SL} = 0.106$ m/s and $U_{SG} = 0.471$ m/s. Time (ms) is indicated in the upper right corner.

4.1 Bubble velocity and void fraction

Fig. 6 shows the bubble velocity as a function of the mixture superficial velocity. Data follow a linear tendency, and U_G increases when increasing U_M, as expected [19]. As a consequence, the bubble velocity was fitted with the drift-flux model [24], which under the hypothesis of negligible gravity levels can be written as [1, 19, 25]:

$$U_G = C_0 U_M = C_0(U_{SG} + U_{SL}), \tag{2}$$

where C_0 is the void fraction distribution coefficient, which considers both the effect of the non-uniform velocities of the gas and liquid phases as well as the void profiles through the flow capillary. The value of C_0 has been calculated as the slope of the straight line of best fit plotted in figure 6. A value of 1.10 was found for the experimental void fraction distribution coefficient. Numerical simulations correctly reproduced the same behaviour than the one observed in the experiments. C_0 was found to be 1.15 in numerical simulation, 4.5% greater than in the experiments. The percentage of error with respect to the experimental data range from 0.8% to 19.3%,

for the smallest value of U_{SG} and U_{SL}, with an average in difference of 5.4%. Both values of C_0 agree with previous results of gas and liquid mixtures flowing in minichannel reported in the related literature [1, 25]. Moreover, being greater than 1, both values of C_0 prove that bubbles moves faster than the mixture, which is due to the fact that bubbles move mainly through the centreline of the capillary.

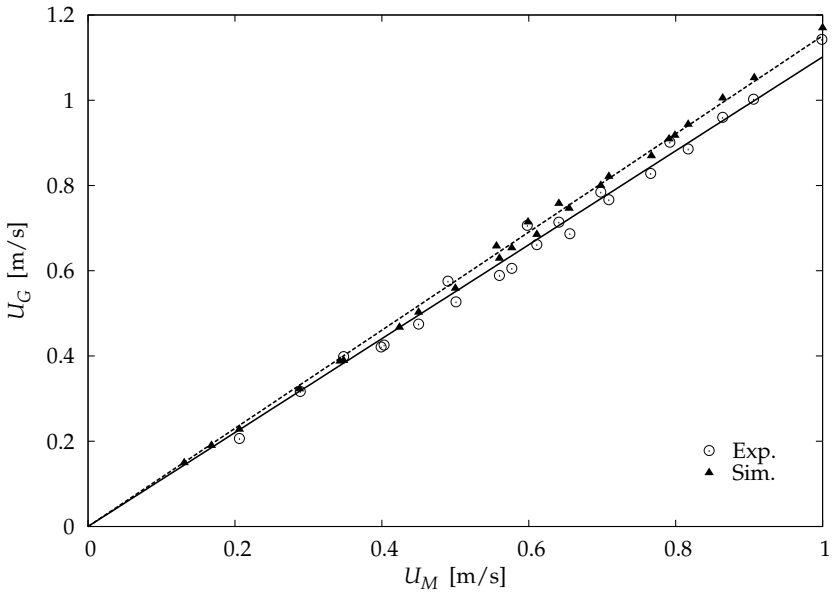


Figure 6: Bubble velocity as a function of the mixture superficial velocity. *Empty symbols*: experimental data, *solid symbols*: numerical data. *Lines*: fitting of the data by using Eq. 2, with $C_0 = 1.10$ for experimental data (solid line) and $C_0 = 1.15$ for numerical simulation (dashed line).

Combining Eq. 2 and the expression $\alpha = U_{SG}/U_G$, an analytical prediction of the volume void fraction as a function of the gas and liquid superficial velocities can be found:

$$\alpha = \frac{1}{C_0 \left(1 + \frac{U_{SL}}{U_{SG}}\right)} \tag{3}$$

Figures 7 shows the volume void fraction as a function of the ratio of the gas and liquid superficial velocities. The prediction of α provided by Eq. 3 has also been plotted using the corresponding value of C_0 in each case. Both experimental and numerical data fit well to their respective predictions. The values of void fraction were found to be slightly smaller in the numerical simulations, coherently with their previously observed greater values of velocity. A minimum and maximum relative error of 0.8% and 15.5%, respectively, and an average value of 5.0% were found. 3D numerical simulations proved to be useful by providing information that was not available from the experiments (two points below the value $U_{SG}/U_{SL} = 0.1$).

4.2 Bubble generation frequency

Figure 8 shows the bubble generation frequency, f, versus the superficial gas velocity for both experiments and numerical simulations. Data corresponding to three different superficial liquid velocity has been plotted. Three additional points corresponding to the superficial gas velocity of 0.025 m/s, which are not present in the experimental data, were also simulated in order to enrich the knowledge about the behaviour of bubble generation frequency

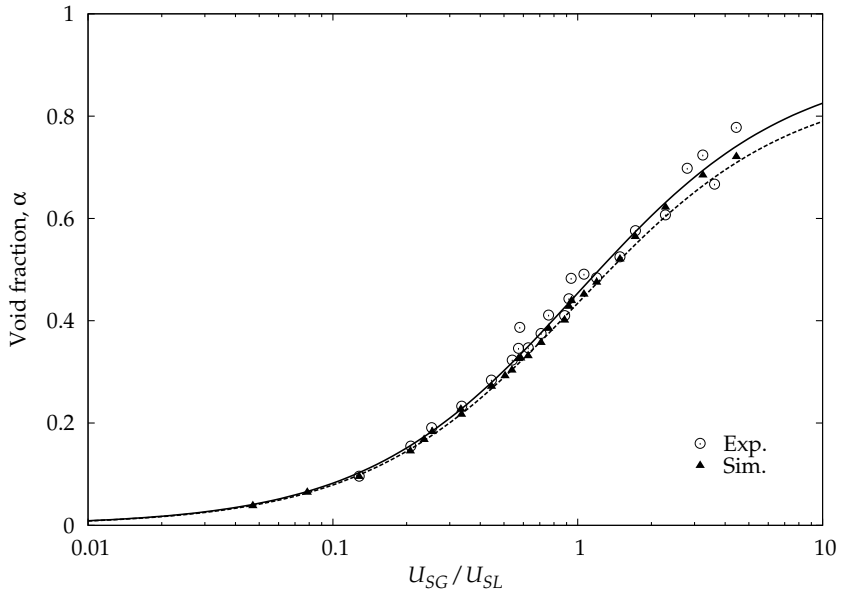


Figure 7: Void fraction as a function of the ratio of the gas and liquid superficial velocities. *Empty symbols*: experimental data, *solid symbols*: numerical data. *Lines*: theoretical prediction given by Eq. 3, with $C_0 = 1.10$ for experimental data (solid line) and $C_0 = 1.15$ for numerical simulation (dashed line).

At very low gas flow rate, f follows a close-to-linear tendency (the linear regime) that progressively curves when increasing U_{SG} until reaching a saturation value, f_{sat}, that remains constant thereafter (the saturation regime). Both behaviours (and regimes) were previously reported [11, 12].

Figure 8 shows how f approaches its corresponding saturated value for the three different values of U_{SL}. The saturation frequency increases when increasing U_{SL}, which is a consequence of the corresponding increment of the liquid drag force, the main force in charged of detaching bubbles [22, 23]. Therefore, larger liquid drag forces are expected to produce smaller bubbles resulting in larger frequency values [11, 12]. This effect also explains why the linear slope at the origin, a_0, increases when increasing the liquid superficial velocity (see table 3).

	Experiments		Simulations	
$U_{SL}[m/s]$	a_0 [m^{-1}]	f_{sat} [1/s]	a_0 [m^{-1}]	f_{sat} [1/s]
0.106	616.4	99.7	856.6	114.5
0.318	2033.9	228.4	2211.6	282.4
0.531	4483.6	376.3	6426.8	392.0

Table 3: Values of a_0 and f_{sat} obtained when fitting the experimental and numerical data by using Eq. 4.

As shown in figure 8, numerical simulations satisfactorily describe the behaviour of f observed in experiments. Numerical simulations agree qualitatively, but not quite quantitatively, with experimental results, providing values in accordance but always larger than the experimental ones. These quantitative differences, calculated as the percentages of error with respect to the experimental data, range from 3.5% (within the margin of error of the mesh) to 34% ($U_{SG} = 0.025$ m/s and $U_{SL} = 0.106$ m/s) with an average value of 17%. One possible explanation of this disagreement, could be that numerical simulations do not adequately resolve the necking region during the bubble pinch-off, leading to an inaccurate calculation of the surface

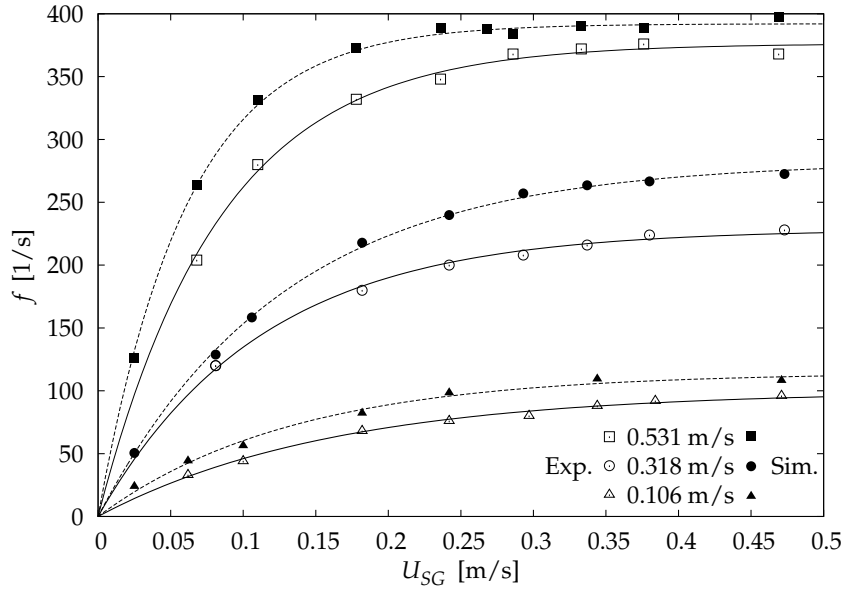


Figure 8: Bubble frequency as a function of the superficial gas velocity for three different superficial liquid velocities. *Empty symbols*: experimental data, *solid symbols*: numerical data. *Solid lines*: fitting of the experimental data by using Eq. 4, *dashed lines*: fitting of the numerical data by using Eq. 4.

tension forces and to a premature pinch-off of the bubble. Future work should focus especially on this point with a more detailed pinch-off study.

A simple exponential model, which fits well and smoothly both experimental and numerical data, is used to fit the values of the bubble generation frequency in figure 8:

$$f = f_{sat}\left(1 - e^{\left(-\frac{a_0}{f_{sat}}U_{SG}\right)}\right) \tag{4}$$

Only two parameters are considered in this model, the initial slope a_0 of the linear regime (i.e. slope when $U_{SG} = 0$), and the value of the saturation frequency f_{sat}. Eq. 4 is consistent with the experimental observations regarding the existence of a linear and a saturation regime. Assuming a regular generation of bubbles in size, the gas flow rate injected into the capillary must be equal to the bubble volume times the frequency, $Q_G = V_B f$, which leads to $f = AU_{SG}/V_B$. Matching this last expression with the first-order Taylor series expansion of Eq. 4, $f_0 = a_0 U_{SG}$, shows that $a_0 = A/V_B$. The initial slope of the linear regime corresponds to the maximum value of the ratio f/U_{SG} for each given U_{SL}. Then, a_0 is related to the minimum bubble volume that can be generated for each given U_{SL}, that is, $V_B|_{min} = A/a_0$. As a consequence, the greater the initial slope at the linear regime, the greater the liquid drag force and the smaller the minimum volume of bubble that can be generated for any U_{SL}. The saturation frequency is related in turn to the minimum time required to form a bubble, marking a limiting scale for the bubble generation process [11].

Table 3 summarises the values obtained for a_0 and f_{sat} when fitting both the experimental and numerical data by using Eq. 4. In every case, these values of both a_0 and f_{sat} were found to be of the same order of magnitude but larger in the numerical simulations. This result shows again that the bubble generation process was faster in the simulations than in the experiments, resulting in larger frequencies and smaller bubbles size in the numerical simulations.

The non-dimensional Strouhal number, St, is used as a normalisation of the bubble generation frequency. St is defined here by means of the gas velocity as $St = f\phi_c/U_G$. Two

linear and independent tendencies with respect to the volume void fraction can be found for St at small and larger values of α. For small values of α (linear regime), St can be written as $St = \bar{a}_0\alpha$ (a_0 being normalised with ϕ_c), by taking into account the first-order Taylor series expansion of eq. 4. For larger values of α (saturation regime), St can be written in turn as $St = f_{sat}\phi_c(1 - C_0\alpha)/(C_0U_{SL})$, by using eq. 3 and the definition of α. Being the ratio $f_{sat}\phi_c/(C_0U_{SL})$ a constant value for each given liquid superficial velocity, St becomes proportional to $(1 - C_0\alpha)$ in the saturation regime.

Fig. 9 shows the Strouhal number as a function of the volume void fraction. Eq. 3 has been rewritten to provide the change between variables (U_{SG} and α) as $U_{SG} = C_0U_{SL}\alpha/(1 - C_0\alpha)$. The same values of C_0 obtained in section 4.1 has been used in this figure. Eq. 4 is also used here to provide a prediction in fig. 9. Both linear tendencies, $St \propto \alpha$ and $St \propto (1 - C_0\alpha)$ (in the linear and saturation regime, respectively), can be observed in the figure. The location of a crossover point between the linear and saturation regimes (for each U_{SL}) can be defined as the maximum value in this figure. Numerical simulations behave similarly to experimens. The percentage of error in St range from 0.3% to 32.6% with an average in difference of 11.9%. Values of St are greater in simulations than in experiments, correspondingly to their bigger values of f. The increase in U_{SL} also increases St, as a consecuence again of the increase in f.

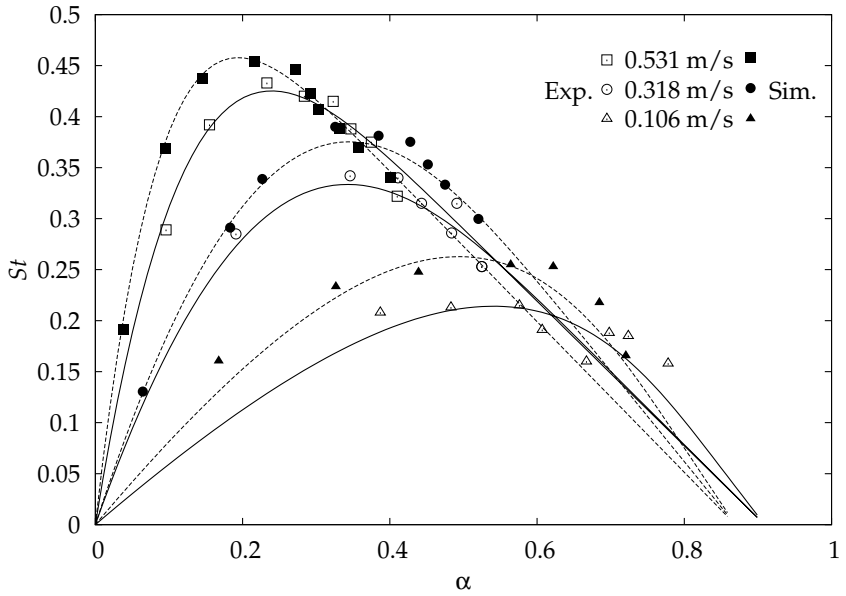


Figure 9: Strouhal number, St, as a function of the volume void fraction, α, for three different superficial liquid velocities. *Empty symbols*: experimental data, *solid symbols*: numerical data. *Solid lines*: prediction of the experimental data by using Eq. 4 and 2, *dashed lines*: prediction of the numerical data by using Eq. 4 and 2.

4.3 Bubble volume and equivalent diameter

Normalising the relationship $Q_G = V_Bf$ with the cross section area times the capillary diameter, a non-dimensional expression of the bubble volume can be obtained:

$$\bar{V}_B = \frac{U_{SG}}{f\phi_c}. \tag{5}$$

The bubble volume was measured in the 3D numerical simulations. No direct measures of V_B could be performed in the experiments. For that reason, figure 10 shows results only

from numerical simulations. In that figure, the non-dimensional bubble volume is plotted as a function of the ratio $U_{SG}/(f\phi_c)$. Data fit well to the prediction of Eq. 5, which confirms that bubbles were generated with high regularity.

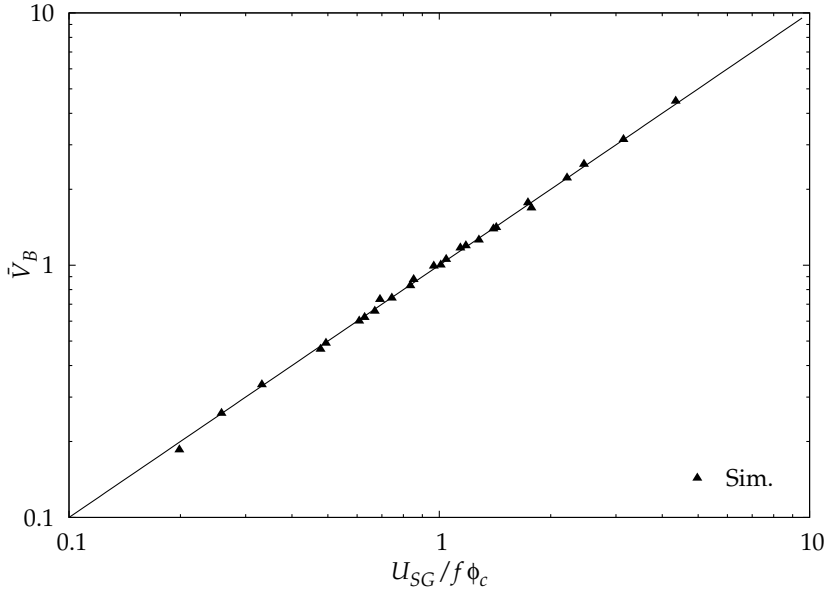


Figure 10: Normalised bubble volume as a function of $U_{SG}/(f\phi_c)$. *Symbols*: simulation data. *Line*: theoretical prediction given by Eq. 5.

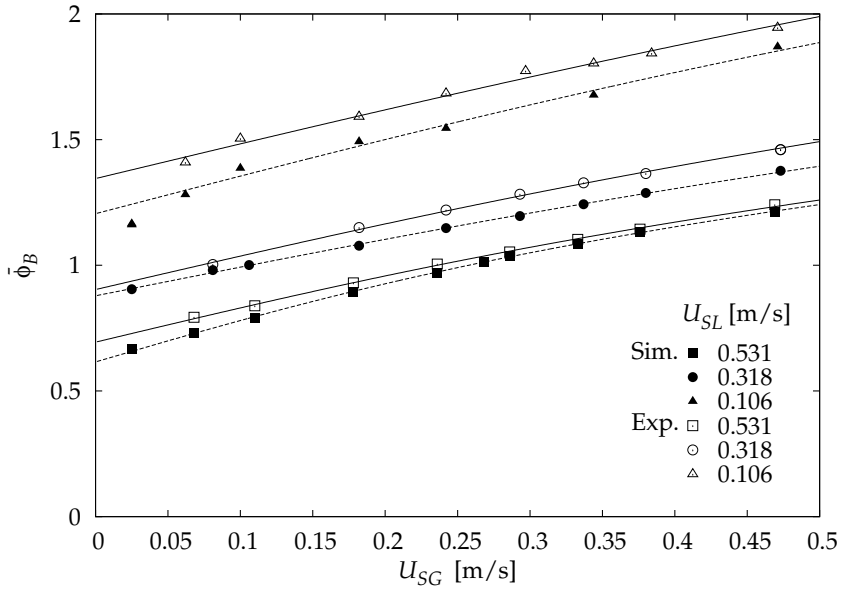


Figure 11: Normalised equivalent diameter $\bar{\phi}_B$ as a function of the gas superficial velocity U_{SG}. *Empty symbols*: experimental data, *solid symbols*: numerical data. *Lines*: solid and dashed lines, given by Eq. 6, correspond to experimental and numerical data, respectively.

In order to provide a quantitative comparison of the bubbles size, an equivalent diameter ϕ_B has been considered for each bubble. ϕ_B is estimated by matching the bubble volume predicted by Eq. 5 to the volume of a sphere $V_S = \pi\phi_B^3/6$. Scaling V_S with $A\phi_c$, the non-dimensionless

volume of the sphere can be written as $\bar{V}_S = 2\bar{\phi}_B^3/3$. The equivalent diameter is normalised with the capillary diameter and is then computed as the result of matching $\bar{V}_B$ and $\bar{V}_S$:

$$\bar{\phi}_B = \left(\frac{3}{2}\frac{U_{SG}}{f\phi_c}\right)^{1/3} \tag{6}$$

Figure 11 shows the non-dimensional equivalent diameter as a function of the gas superficial velocity. Differences in equivalent diameter between experimental and numerical results range from 1.1% to 9.1% with an average of 5.2% difference. The equivalent diameter is estimated using Eq. 6 in both cases. Lines correspond to Eq. 6, in which the prediction provided by Eq. 4 was included. Both experimental and numerical data correctly agree with the theoretical prediction. The equivalent diameter is found to be slightly smaller in numerical simulations than those from experiments, which coherently corresponds to the bigger values in U_G and f values from the simulations. It can be observed that increasing the liquid superficial velocity causes the reduction of the equivalent diameter. The value $\bar{\phi}_B = 1$ allows for the distinction between the bubble ($\bar{\phi}_B < 1$) and the slug flow regime ($\bar{\phi}_B > 1$) in Fig. 11.

5 CONCLUSIONS

- The first part of this work has presented a study on the numerical simulation of bubbles generated in a 3D T-junction (1 mm internal diameter capillaries), with an air/water mixture, using the commercial CFD solver ANSYS Fluent v15.0.7.
- Numerical results were compared with experimental data, both cases following the same conditions.
- Several trains of bubbles were simulated with high regularity and small dispersion in size during the bubble formation process.
- No artificial pinch-off mechanism was required for the detachment of bubbles.
- Bubble or slug flow regimes were obtained in all cases, both in experiments and numerical simulations. Only short and middle Taylor bubbles were generated in the slug flow pattern.
- Numerical simulations satisfactorily reproduced the formation of bubbles, both in the final shape as well as in the evolution of shape during the formation of bubbles.
- Several parameters were compared between 3D-CFD results and experimental data, such as the bubble velocity, void fraction, bubble generation frequency, Strouhal number and the bubble equivalent diameter.
- Numerical results were found to be of the same order of magnitude, though CFD simulations predict larger frequency, resulting in faster and smaller sized bubbles.
- Future work should focus in studying into more details the pinch-off process in the numerical simulation, in order to try to further improve these results.
- 3D-CFD simulations allowed to study new zones of interest, especially at low gas flow rate, and new parameters such as the bubble volume, which were not available from experimental results, showing that this is a useful tool to improve our current experimental knowledge on two-phase flows in minichannels.

Acknowledgments

This work has been financially supported by the Spanish *Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación, Desarrollo e Innovación* (Project numbers AYA2012-34131 and MTM2013-46313-R) and the *Generalitat de Catalunya* (Grant number 2014-SGR-1471).

REFERENCES

[1] S. Arias, D. Legendre D., R. González-Cinca, Numerical simulation of bubble generation in a T-junction. *Computers & Fluids*, **56**, 49–60, 2012.

[2] K. Yamamotoa, S. Ogata, Effects of T-junction size on bubble generation and flow instability for two-phase flows in circular microchannels. *Int. J. Multiphase Flow*, **49**, 24-30, 2013.

[3] K. Yamamotoa, S. Ogata, Drag reduction of slug flows in microchannels by modifying the size of T-junctions. *Int. J. Multiphase Flow*, **62**, 67-72, 2014.

[4] D. Qian, A. Lawal, Numerical study on gas and liquid slugs for Taylor flow in a Tjunction microchannel. *Chem Eng Sci*, **61(23)**, 7609–7625, 2006.

[5] M.N. Kashid, F. Platte, D.W. Agar, S. Turek, Computational modelling of slug flow in a capillary microreactor. *J Comput Appl Math*, **203(2)**, 487–497, 2007.

[6] M. De Menech, P. Garstecki, F. Jousse, H.A. Stone, Transition from squeezing to dripping in a microfluidic T-shaped junction. *J Fluid Mech*, **595**, 141–161, 2008.

[7] I.L. Ngo, T.D. Dang, C. Byon, S.W. Joo, A numerical study on the dynamics of droplet formation in a microfluidic double T-junction. *Biomicrofluidics*, **9**, 024107, 2015.

[8] S. Malekzadeh, E. Roohi, Investigation of Different Droplet Formation Regimes in a T-junction Microchannel Using the VOF Technique in OpenFOAM. *Microgravity Sci. Technol.*, **27**, 231–243, 2015.

[9] Y. Shi, G.H. Tang, H.H. Xia, Lattice Boltzmann simulation of droplet formation in T-junction and flow focusing devices. *Computers & Fluids*, **90**, 155–163, 2014

[10] S. Ben Saad, C. Gentric, J.F. Fourmigué, P. Clément, J.P. Leclerc, CFD and experimental investigation of the gas-liquid flow in the distributor of a compact heat exchanger. *Chemical Engineering Research & Design*, **92(11)**, 2361–2370, 2014.

[11] S. Arias, X. Ruiz, L. Ramírez-Piscina, J. Casademunt, R. González-Cinca R., Experimental study of a microchannel bubble injector for microgravity applications. *Microgravity Sci. Technol*, **21**, 107–111, 2009.

[12] S. Arias, R. González-Cinca, X. Ruiz, L. Ramírez-Piscina, J. Casademunt, Characterization of the performance of a minibubble generator in conditions relevant to microgravity. *Colloids and Surfaces A: Physicochem. Eng. Aspects*, **365**, 52–55, 2010.

[13] M. Suo, P. Griffith, Two-phase flow in capillary tubes. *J. Basic Eng.*, **86**, 576–582, 1964.

[14] K.S. Rezkallah, Weber number based flow-pattern maps for liquid-gas flows at microgravity. *Int. J. Multiphase Flow*, **22(6)**, 1265–1270, 1996.

[15] A.E. Dukler, J.A. Fabré, J.B. McQuillen, R. Vernon, Gas-liquid flow at microgravity conditions: flow patterns and their transitions. *Int. J. Multiphase Flow*, **14**, 389–400, 1988.

[16] ANSYS® Academic Research, Release 15.0,7, Help System, ANSYS, Inc, 2014.

[17] D.L. Youngs, Time-Dependent Multi-Material Flow with Large Fluid Distortion. *K. W. Morton and M. J. Baines, editors, Numerical Methods for Fluid Dynamics*, Academic Press, 1982

[18] J.U. Brackbill, D.B. Kothe, C. Zemach, A Continuum Method for Modeling Surface Tension. *J. Comput. Phys.*, **100**, 335–354, 1992.

[19] C. Colin, J. Fabré, A.E. Dukler, Gas-liquid flow at microgravity conditions - I Dispersed bubble and slug flow. *Int. J. Multiphase Flow*, **17**, 533–544, 1991.

[20] C. Colin, J. Fabré, Gas-liquid pipe flow under microgravity conditions: influence of tube diameter on flow pattern and pressure drops. *J. Adv. Space Research*, **16**, 137–142, 1995.

[21] G. Puccetti, M. Tosi, B. Pulvirenti, G.L. Morini, Flow patterns of an air-water mixture at the exit of a micro T-junction. *Experimental Thermal and Fluid Science*, **67**, 62–69, 2015.

[22] J. Carrera, X. Ruiz, L. Ramírez-Piscina, J. Casademunt, M. Dreyer, Generation of a monodisperse microbubble jet in microgravity. *AIAA Journal*, *46(8)*, 2010–2019, 2008.

[23] T. Fu, Y. Ma, D. Funfschilling, C. Zhu, H.Z. Li, Squeezing-to-dripping transition forbubble formation in a microfluidic T-junction. *Chem Eng Sci*, **65(12)**, 3739–3748, 2010.

[24] N. Zuber, J. Findlay, Average Volumetric Concentration in Two-Phase Systems. *Trans ASME Jul Ht Transfer*, **87**, 453–468, 1969.

[25] J. McQuillen, C. Colin, J. Fabre, Ground-based gas-liquid flow research in microgravity conditions: state of knowledge. *Space Forums*, **3**, 165–203, 1998.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

COUPLING OF FLUID STRUCTURE INTERACTION SOLVER WITH A VOF METHOD FOR MULTIPHASE STRUCTURE INTERACTION

D. Cerroni[1*], R. Da Vià[1], S. Manservisi[1] and F. Menghini[1]

[1]University of Bologna
Via dei Colli 16, Laboratory of Montecuccolino - DIN

[*] e-mail: daniele.cerroni2@unibo.it

Keywords: Fluid Structure Interaction, Volume of Fluid method, Multiphase flows

Abstract. *In this work we propose a preliminary model to study the deformation of solid structures induced by the interaction with a two-phase flow. The study of Fluid Structure Interaction and multiphase problems is of great interest because of many potential applications ranging from the biomedical field to the pressure tank design. We use a monolithic approach for the FSI problem while a Volume Of Fluid method (VOF) is considered for the reconstruction and the advection of the multiphase interface. An unstructured, time dependent computational grid and a fine Cartesian mesh are used for the FSI and the VOF problem, respectively. The interaction between the two different grids is obtained by projecting the velocity and the displacement field into the Cartesian grid and the color function into the unstructured mesh. This projection is performed with the MEDmem libraries included in the Salome platform. Concerning the VOF method, for an accurate reconstruction of the interface a huge number of computational elements are required and a multilevel algorithm coupled to an efficient compression-expansion technique is developed to reduce computational costs and memory requirements. After the mathematical description of the problems we test the proposed algorithm with different cases where the solid domain undergoes to both small and large deformation.*

1 INTRODUCTION

In recent years Fluid Structure Interaction (FSI) problems have gained attention in the computational field because of the improvement in computer performance and an increasing interest in more realistic modeling [17, 12]. Among all the applications, these computational techniques can be employed in biomedical field where the interaction between a fluid and a solid such as the blood and the artery wall, is of great importance. In a classic FSI problem a solid structure interacts with a single phase fluid. Few works can be found in literature where Multiphase Fluid Structure problems are considered. Two-phase flows are of great interest in several physical and industrial applications. In recent years such interest inspired the development of several numerical algorithms for the efficient and accurate simulation of multiphase flows. These algorithms include efficient solvers of the incompressible Navier-Stokes equations, stable and accurate techniques for the advection and reconstruction of the fluid interface and methods for the correct computation of surface tension forces. These algorithms can be classified in two categories considering how the computational grid is handled. In the first group moving grids are used [14]. The interface between the two phases always lies over the cell boundaries, which are continuously advected by following the flow motion. This representation allows a precise modeling of the fluid property discontinuities, and also a correct location of the capillary force, which is a singular term in the Navier-Stokes equations. Despite these good properties, moving grid methods can be used only when the general motion does not affect the topology of the computational grid, since severe grid stretching brings loss of accuracy in the discretization of the Navier-Stokes equations. Furthermore, since breaking or merging of drops and bubbles are almost impossible, a continuous re-meshing of the computational domain is required.

In the second group fixed grids are used. The interface points cross the cell boundaries and a numerical algorithm is then required to reconstruct its motion in the computational domain. The fluid properties are evaluated in the interface cells as an average between the properties of each phase. The capillary force, which is physically located only on the interface, is redistributed inside the cells cut by the interface. With this representation, severe deformations of the interface can be tracked without the need for re-meshing. For these reasons fixed grid methods are widely used. Fixed-grid methods for two-phase flows can be either Eulerian or Lagrangian. Eulerian front capturing schemes include volume-of-fluid (VOF), level set and phase field methods [16]. All these methods are based on a single scalar function defined in every grid cell. The phase motion can be followed during the simulation with a good conservation of mass, however the interface and all its geometrical information are usually poorly represented, especially in regions with high curvature. Moreover, filaments are usually artificially broken when their thickness becomes comparable to the grid spacing.

Lagrangian techniques, either front-tracking schemes with surface markers or with volume particles, maintain filamentary structures better than Eulerian methods and conserve mass adequately, even without explicit volume conservation constraints [1, 9]. Both the surface markers and the volume particles are moved along the flow characteristics, resulting in a more accurate description of the interface and its geometrical properties, such as the normal vector and the curvature. This representation allows for a very accurate volume conservation and interface advection even in situations where strong deformations of the interface are present. On the other hand, these schemes can be rather expensive,

as the number of markers or volume particles increases, or are associated to practical difficulties. These includes the need to keep logical connections among surface elements and the careful handling of drop breaking and merging. Fixed grid methods provide good solutions to the problem of interface advection. However the accurate representation of surface forces remains a problem since these forces are not located on the cell boundaries. The multiphase flow motion satisfies the Navier-Stokes equations, and it should be determined while taking into account the discontinuities of the fluid properties across the interface. Such discontinuities, together with high Reynolds numbers, may induce several numerical instabilities.

The coupling between a FSI and a VOF solver allows to increase dramatically the range of problems that can be studied. For example, deformation of offshore platforms or offshore wind farm due to water waves or the interaction between water and ship hulls can be simulated in a very realistic way. In order to couple the two problems many strategies can be employed, as an example, the first code output field can be written on a file and the other code can be run with this file as input and vice versa. This technique is simple and fast to implement but unfortunately it can be very slow because of the large time needed by writing and reading files. A more efficient and fast coupling can be obtained through dedicated libraries for the data exchange, like the MEDmem libraries implemented in the SALOME computational platform [2, 18].

In this work we consider Fluid Structure Interaction problems in which the fluid part is composed of two phases [11, 6]. For the reconstruction and advection of the interface we employ a front-tracking Volume of Fluid method [19, 13]. The coupling between the two solvers is performed using the MEDmem libraries for data exchange. In the next Section the mathematical model for a Multiphase Fluid Structure Interaction problem is presented. In the last Section some numerical results are show.

2 MATHEMATICAL MODEL

2.1 FSI problem

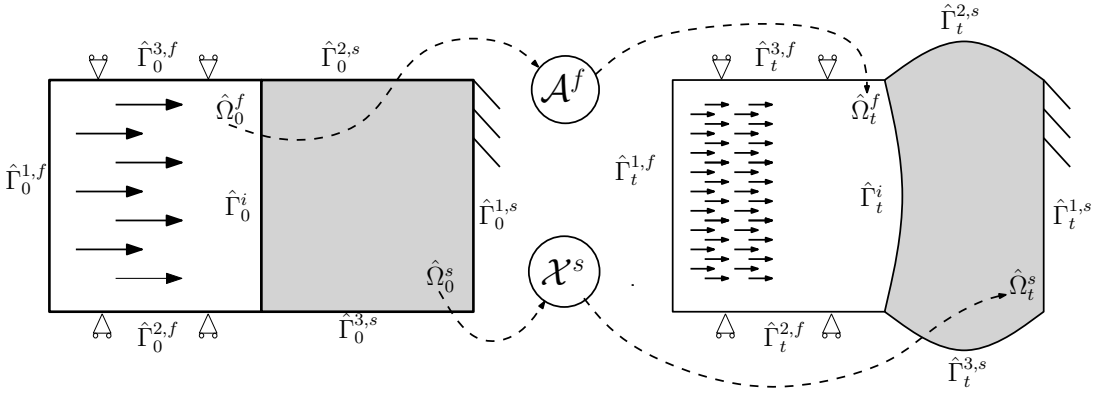


Figure 1: Reference and current configuration where a vessel wall interacts with a fluid.

In this Section we introduce the mathematical model for a generic Fluid Structure Interaction problem. In an ordinary FSI problem we consider a mechanical system composed by a laminar Newtonian fluid region and a solid one which defines a moving domain Ω_t. A schematic geometry of the problem is shown in Figure 1. Let Ω_t^f and Ω_t^s be the fluid and the solid region at $t \in (0, T]$, respectively. At $t = 0$ the fluid and solid region

are defined by $\hat{\Omega}_0^f$ and $\hat{\Omega}_0^s$. Let $\Gamma_t^i = \bar{\Omega}_t^f \cap \bar{\Omega}_t^s$ and $\hat{\Gamma}_0^i = \bar{\Omega}_0^f \cap \bar{\Omega}_0^s$ be the interface where solid and fluid interact. $\Gamma_t^k, k = 1, 2, 3$ and $\hat{\Gamma}_0^k, k = 1, 2, 3$ are defined to be the remaining external boundaries at $t \in (0, T]$ and $t = 0$, respectively. The evolution of the solid and fluid domain $\hat{\Omega}_0^f$ and $\hat{\Omega}_0^s$ are defined by

$$\mathcal{X}^s : \hat{\Omega}_0^s \times \mathbb{R}^+ \to \mathbb{R}^3 ,$$
$$\mathcal{A}^f : \hat{\Omega}_0^f \times \mathbb{R}^+ \to \mathbb{R}^3 ,$$

such that the range of $\mathcal{X}^s(\cdot, t)$ and $\mathcal{A}^f(\cdot, t)$ define Ω_t^s and Ω_t^f, respectively. $\mathcal{X}^s$ maps any material point $\hat{\mathbf{x}}_0^s$ from the given fixed reference configuration $\hat{\Omega}_0^s$ to the current solid material configuration Ω_t^s. The solid displacement is then defined as

$$\hat{\mathbf{u}}^s(\hat{\mathbf{x}}_0^s, t) = \mathcal{X}(\hat{\mathbf{x}}_0^s, t) - \hat{\mathbf{x}}_0^s . \tag{1}$$

The mapping $\mathcal{A}^f$ is such that $\mathcal{A}^f(\hat{\mathbf{x}}_0^f, t) = \hat{\mathbf{x}}_0^f + \hat{\mathbf{u}}^f(\hat{\mathbf{x}}_0^f, t)$, where $\hat{\mathbf{u}}^f(\hat{\mathbf{x}}_0^f, t)$ is defined as an arbitrary extension operator over the fluid domain $\hat{\Omega}_0^f$ and given by

$$\hat{\mathbf{u}}^f(\hat{\mathbf{x}}_0^f, t) = \mathrm{Ext}(\hat{\mathbf{u}}^s|_{\hat{\Gamma}_0^i}) \quad \text{in} \quad \hat{\Omega}_0^f . \tag{2}$$

The extension operator used to evaluate the fluid region displacement is the harmonic or Laplace operator. Other similar operators can be employed as described in [15, 12, 5]. The velocity $\hat{\mathbf{w}}^f$ is defined by

$$\hat{\mathbf{w}}^f = \frac{\partial \hat{\mathbf{u}}^f}{\partial t} \quad \text{in} \quad \hat{\Omega}_0^f . \tag{3}$$

This quantity represents the velocity in terms of the reference coordinate $\hat{\mathbf{x}}_0^f$. The behavior of the fluid is described by the Navier-Stokes equations for incompressible flows. For details the interested reader can also see [20, 10].

$$\begin{aligned}
&\rho^f \frac{\partial \mathbf{v}^f}{\partial t}\bigg|_{\tilde{\mathcal{A}}} + \rho^f \left(\mathbf{v}^f - \mathbf{w}^f\right) \cdot \boldsymbol{\nabla} \mathbf{v}^f - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}^f = \rho^f \mathbf{g} && \text{in} \quad (0, T) \times \Omega_t^f , \\
&\boldsymbol{\nabla} \cdot \mathbf{v}^f = \mathbf{0} && \text{in} \quad (0, T) \times \Omega_t^f , \\
&\mathbf{v}^f|_{t=0} = \mathbf{v}_0 && \text{in} \quad \hat{\Omega}_0^f , \\
&\mathbf{v}^f|_{\Gamma_{t,D}^{1,f} \cup \Gamma_{t,D}^{2,f}} = \mathbf{g}^f && \text{in} \quad (0, T) , \\
&\boldsymbol{\sigma}_f \cdot \mathbf{n}^f|_{\Gamma_{t,N}^{1,f} \cup \Gamma_{t,}^{2,f}} = \mathbf{h}^f && \text{in} \quad (0, T) ,
\end{aligned} \tag{4}$$

where ρ^f is the constant density, $\mathbf{v}^f$ is the fluid velocity, $\mathbf{g}$ is the gravity acceleration vector, $\tilde{\mathcal{A}}$ denotes the ALE application that maps the reference fluid configuration $\hat{\Omega}_0^f$ onto the current fluid configuration Ω_t^f and $\mathbf{w}^f$ denotes the fluid domain velocity. $\mathbf{n}$ is the unit normal vector that points outward from the boundary $\partial\Omega_t^f$ and $\mathbf{g}^f$, $\mathbf{h}^f$, $\mathbf{v}_0$ are given data. The flow state variables in the incompressible case are the pressure p^f and the velocity $\mathbf{v}^f$. The contribution of external forces such as gravity is assumed to be negligible. The constitutive relation for the stress tensor in the Newtonian incompressible case reads

$$\boldsymbol{\sigma}^f = -p^f \mathbf{I} + \boldsymbol{\tau}^f = -p^f \mathbf{I} + 2\mu^f \boldsymbol{\epsilon}\left(\mathbf{v}^f\right) , \tag{5}$$

where μ^f is the dynamic viscosity of the fluid, p^f the Lagrange multiplier associated to the incompressibility constraint and $\boldsymbol{\epsilon}\left(\mathbf{v}^f\right)$ the strain rate tensor defined as

$$\boldsymbol{\epsilon}\left(\mathbf{v}^f\right) = \frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{v}^f + \left(\boldsymbol{\nabla}\mathbf{v}^f\right)^{\mathrm{t}}\right) . \tag{6}$$

The total time derivative is related to the adopted reference systems. The governing equations for structural mechanics are the following momentum equations

$$\rho^s \frac{\partial \mathbf{v}^s}{\partial t} - \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}^s(\mathbf{u}^s) = \mathbf{0} \qquad \text{in} \quad \Omega_t^s \,, \tag{7}$$

where ρ^s is the density of the solid material, $\mathbf{v}^s$ is the velocity field and $\boldsymbol{\sigma}^s$ its Cauchy stress tensor, which is a function of the solid region displacement $\mathbf{u}^s$. Since the constitutive law for the solid stress tensor is expressed in terms of displacements one must solve both the balance equations (7) and the kinematic relation

$$\mathbf{v}^s = \frac{\partial \mathbf{u}^s}{\partial t} \,. \tag{8}$$

For the reference configuration we can introduce the right Cauchy-Green deformation tensor $\mathbf{C}$ as

$$C_{ij} = F_{ki}F_{kj} \qquad\qquad \forall\, i,j = 1,\ldots,3 \,, \tag{9}$$

where $\mathbf{F}$ is the deformation gradient tensor defined by $\mathbf{F} = \mathbf{I} + \boldsymbol{\nabla}\mathbf{u}^s$. In a similar way in the current configuration we can introduce the left Cauchy-Green deformation tensor, $\mathbf{b}$, as

$$b_{ij} = F_{ik}F_{jk} \qquad\qquad \forall\, i,j = 1,\ldots,3 \,. \tag{10}$$

According with this notation we can now express the Cauchy stress tensor, $\boldsymbol{\sigma}^s$, as [20]

$$\sigma^s_{ij} = \frac{2}{J}\left[b_{ij}\left(I\, b_{ij} - b_{im}b_{mj}\right)\frac{J\delta_{ij}}{2}\right]\begin{pmatrix}\frac{\partial W}{\partial I}\\ \\ \frac{\partial W}{\partial II}\\ \\ \frac{\partial W}{\partial J}\end{pmatrix} , \tag{11}$$

where $I = C_{ii}$, $II = 1/2\,(I - C_{ij}C_{ji})$ are the first and second invariant of the right Cauchy-Green strain tensor $\mathbf{C}$ and J its determinant. The quantity $W = W(I, II, J)$ is the strain energy of the system which depends on the constitutive law of the considered material. For example for a Neo-Hookian material, with respect to the current configuration, the energy function is defined by

$$\begin{aligned} W(I,J) =& \frac{1}{2}\mu_s\left(J^{-2/3}\mathrm{tr}\mathbf{C} - 3\right) + \\ & \frac{1}{2}\left(\lambda + \frac{2}{3}\mu_s\right)\left(\frac{1}{2}(J^2-1) - \ln J\right) . \end{aligned} \tag{12}$$

In the case of incompressible solid, the third invariant is equal to one so the energy density function becomes

$$W(I,J) = \frac{1}{2}\mu_s\left(\mathrm{tr}\mathbf{C} - 3\right) \tag{13}$$

and the Cauchy stress tensor is defined by

$$\boldsymbol{\sigma}^s = -p^s \mathbf{I} + \boldsymbol{\sigma}^{s*}\,, \tag{14}$$

where $\boldsymbol{\sigma}^{s*}$ is the tensor obtained by using the equations (11) and (13). The problem defined by (4)-(7) is not well posed since we have not yet prescribed any boundary conditions at the interface Γ_t^i. The coupling between the fluid and the solid model determines the missing boundary conditions, which consist of imposing the continuity of velocity and stress at the interface Γ_t^i as

$$\mathbf{v}^f|_{\Gamma_t^i} = \mathbf{v}^s|_{\Gamma_t^i}\,, \tag{15}$$

$$\boldsymbol{\sigma}^f \cdot \mathbf{n}^f|_{\Gamma_t^i} + \boldsymbol{\sigma}^s \cdot \mathbf{n}^s|_{\Gamma_t^i} = \mathbf{0}\,. \tag{16}$$

In order to write the weak formulation of the coupled problem, let us consider the following functional spaces

$$\begin{aligned}
\mathbf{V}^t &= \{\boldsymbol{\phi} \in \mathbf{H}^1(\Omega_t^f) : \boldsymbol{\phi}|_{\Gamma_{t,D}^{1,f} \cup \Gamma_{t,D}^{2,f}} = \mathbf{0}\}\,,\\
\mathbf{V}_g^t &= \{\boldsymbol{\phi} \in \mathbf{H}^1(\Omega_t^f) : \boldsymbol{\phi}|_{\Gamma_{t,D}^{1,f} \cup \Gamma_{t,D}^{2,f}} = \mathbf{g}^f\}\,,\\
Q^t &= L^2(\Omega_t^f)\,,\\
\mathbf{M}^0 &= \{\boldsymbol{\psi} \in \mathbf{H}^1(\hat{\Omega}_0^s) : \boldsymbol{\psi}|_{\hat{\Gamma}_{0,D}^{1,s} \cup \hat{\Gamma}_{0,D}^{2,s} \cup \hat{\Gamma}_{0,D}^{3}} = \mathbf{0}\}\,,\\
\mathbf{M}_g^0 &= \{\boldsymbol{\psi} \in \mathbf{H}^1(\hat{\Omega}_0^s) : \boldsymbol{\psi}|_{\hat{\Gamma}_{0,D}^{1,s} \cup \hat{\Gamma}_{0,D}^{2,s} \cup \hat{\Gamma}_{0,D}^{3}} = \mathbf{g}^s\}\,,\\
D^0 &= L^2(\hat{\Omega}_0^s)\,,
\end{aligned}$$

where $\mathbf{H}^1(\Omega)$ is the standard Sobolev space of order s with respect to the set Ω and $L^2(\Omega)$ is the Sobolev space of order 0 containing square integrable functions over Ω. For details concerning these function spaces one may consult [8]. In addition, let us introduce the following bilinear form

$$a^f(\mathbf{v}^\mathbf{f}, \boldsymbol{\phi}) = \int_{\Omega^f} \boldsymbol{\tau}^f(\mathbf{v}^f) : \boldsymbol{\nabla}\phi \, d\mathbf{x} = \mu(\boldsymbol{\nabla}\mathbf{v}^f + (\boldsymbol{\nabla}\mathbf{v}^f)^T, \boldsymbol{\nabla}\boldsymbol{\phi})\,, \tag{17}$$

where we denote with $\boldsymbol{\tau}^f$ the fluid viscosity tensor. The variational formulation of the fluid equations can be obtained through the usual method by multiplying the equations (4) with appropriate test functions, performing integrations on the whole domain and keeping into account the boundary and interface conditions. This procedure leads, for the velocity field $\mathbf{v} \in \mathbf{V}_g^t$ and pressure $p \in Q^t$, to the following fluid momentum equation

$$\begin{aligned}
&\rho^f \left(\left.\frac{\partial \mathbf{v}^f}{\partial t}\right|_{\tilde{\mathcal{A}}}, \boldsymbol{\phi} \right) + a(\mathbf{v}^f, \boldsymbol{\phi}) - \rho^f((\boldsymbol{\nabla}\cdot\mathbf{w}^f)\mathbf{v}^f, \boldsymbol{\phi}) + \rho^f(\left(\mathbf{v}^f - \mathbf{w}^f\right)\cdot\boldsymbol{\nabla}\mathbf{v}^f, \boldsymbol{\phi}) = \\
&\qquad (p^f, \boldsymbol{\nabla}\cdot\boldsymbol{\phi}) + \int_{\Gamma_t^i} (\boldsymbol{\sigma}^f\cdot\mathbf{n}^f)\cdot\boldsymbol{\phi}\, d\gamma + \int_{\Gamma_N^f} \mathbf{h}^f\cdot\boldsymbol{\phi}\, d\gamma + \int_{\Omega^f} \rho^f \mathbf{g}\cdot\boldsymbol{\phi}\, d\mathbf{x}\,, \qquad (18)\\
&(q, \boldsymbol{\nabla}\cdot\mathbf{v}^f) = 0\,,\\
&\mathbf{v}^f|_{t=0} = \mathbf{v}_0^f\,,
\end{aligned}$$

for all $\boldsymbol{\phi} \in \mathbf{V}^t$ and $q \in Q^t$. In a similar way, we define the following bilinear form

$$a^s(\mathbf{u}^s, \boldsymbol{\psi}) = (\boldsymbol{\sigma}^s(\mathbf{u}^\mathbf{s}), \boldsymbol{\nabla}\boldsymbol{\psi})\,. \tag{19}$$

By following the procedure briefly described above, we obtain at each time t, for the velocity $\mathbf{u}^s \circ \mathcal{X}^s \in \mathbf{M}_g^0$ and pressure $p^s \circ \mathcal{X}^s \in \mathbf{D}^0$, the following weak formulation for the solid problem

$$\begin{aligned}
&\rho^s \left(\frac{\partial^2}{\partial t^2} \mathbf{u}^s, \boldsymbol{\psi} \right) + a^s(\mathbf{u}^s, \boldsymbol{\psi}) - (p^s, \boldsymbol{\nabla} \cdot \boldsymbol{\psi}) = \int_{\Gamma_t^i} (\boldsymbol{\sigma}^s \cdot \mathbf{n}^s) \cdot \boldsymbol{\psi} \, d\gamma + \int_{\Gamma_N^s} \mathbf{h}^s \cdot \boldsymbol{\psi} \, d\gamma \,, \\
&(d, \boldsymbol{\nabla} \cdot \mathbf{u}^s) = 0 \,, \\
&\mathbf{u}^s|_{t=0} = \mathbf{u}_0^s \,, \qquad \mathbf{v}^s|_{t=0} = \mathbf{v}_0^s \,,
\end{aligned}$$

for all $\boldsymbol{\psi} \circ \mathcal{X}^s \in \mathbf{M}^0$ and $d \circ \mathcal{X}^s \in \mathbf{D}^0$. Let us introduce a global weak formulation for the fluid-structure problem. If we define the functional space

$$\mathbf{S}^t = \{ (\boldsymbol{\phi}, \boldsymbol{\psi} \circ \mathcal{X}^s) \in \mathbf{V}^t \times \mathbf{M}^0 : \boldsymbol{\psi}|_{\Gamma_t^i} = \boldsymbol{\phi}|_{\Gamma_t^i} \} \,, \tag{20}$$

from (15), (16), (18) and (20), we can write the FSI problem in the coupled formulation as

$$\begin{aligned}
&\rho^f \left(\left. \frac{\partial \mathbf{v}^f}{\partial t} \right|_{\tilde{\mathcal{A}}}, \boldsymbol{\varphi} \right) + a(\mathbf{v}^f, \boldsymbol{\varphi}) - \rho^f((\boldsymbol{\nabla} \cdot \mathbf{w}^f)\mathbf{v}^f, \boldsymbol{\varphi}) + \rho^f(\left(\mathbf{v}^f - \mathbf{w}^f\right) \cdot \boldsymbol{\nabla} \mathbf{v}^f, \boldsymbol{\varphi}) - \\
&\qquad (p^f, \boldsymbol{\nabla} \cdot \boldsymbol{\varphi}) + \rho^s(\frac{\partial^2}{\partial t^2} \mathbf{u}^s, \boldsymbol{\varphi}) + a^s(\mathbf{u}^s, \boldsymbol{\varphi}) - (p^s, \boldsymbol{\nabla} \cdot \boldsymbol{\varphi}) \qquad (21) \\
&\qquad\qquad - \int_{\Gamma_N^s} \mathbf{h}^s \cdot \boldsymbol{\varphi} \, d\gamma - \int_{\Gamma_N^f} \mathbf{h}^f \cdot \boldsymbol{\varphi} \, d\gamma = \int_{\Omega^f} \rho^f \mathbf{g} \cdot \boldsymbol{\varphi} \, d\mathbf{x} \,, \quad \forall \boldsymbol{\varphi} \in \mathbf{S}^t \\
&(q, \boldsymbol{\nabla} \cdot \mathbf{v}^f) = 0 \qquad (d, \boldsymbol{\nabla} \cdot \mathbf{u}^s) = 0 \,, \qquad (22) \\
&\mathbf{v}^f|_{t=0} = \mathbf{v}_0^f \qquad \mathbf{u}^s|_{t=0} = \mathbf{u}_0^s \qquad \mathbf{v}^s|_{t=0} = \mathbf{v}_0^s \,.
\end{aligned}$$

It is worth noting that by using the coupling conditions (15), (16) and this particular choice of the fluid-structure test functions, the boundary terms that appear in the fluid-solid interface Γ_t^i cancel out. This assures that forces at the interface are always computed in an exact way.

2.2 Multiphase problem

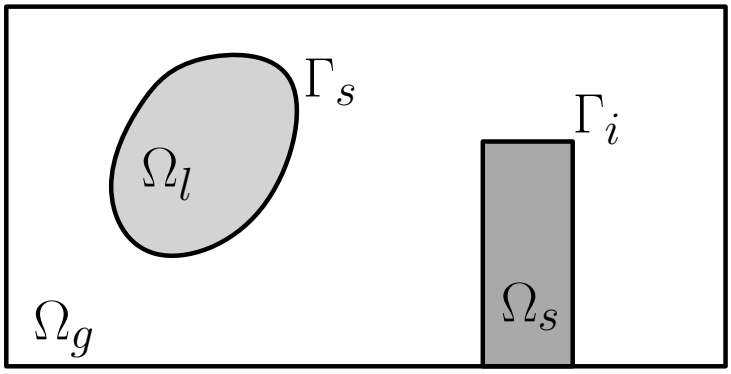


Figure 2: Computational domain. The liquid in Ω_l and Ω_l are the reference fluid phases and the boundary between the two fluid phases is denoted by Γ_s. Ω_s mark the solid region of the domain and the interface between the solid and the fluid region is labeled with Γ_i.

In this Section the mathematical model of the Multiphase Fluid Structure Interaction problem is presented. Let us consider a domain as shown in Figure 2, where Ω marks the whole domain, Ω_l is the portion of the domain occupied by the reference fluid phase and Ω_g is the sub-domain with the secondary phase. The boundary between the two immiscible fluids is denoted by Γ_s, and its topology can vary during the evolution of the system since each sub-domain evolves in time. The solid domain is labeled with Ω_s and the interface between the fluid phases and the solid region is marked Γ_i. As explained in Section 2.1 the balance equations that hold in a fluid structure interaction problem are

$$\rho^f \frac{\partial \mathbf{v}^f}{\partial t}\bigg|_{\tilde{\mathcal{A}}} + \rho^f \left(\mathbf{v}^f - \mathbf{w}^f\right) \cdot \boldsymbol{\nabla} \mathbf{v}^f - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}^f = \rho^f \mathbf{g} \qquad \text{in} \quad (0,T) \times \Omega_t^f \cup \Omega_t^g \,, \tag{23}$$

$$\rho^s \frac{\partial \mathbf{v}^s}{\partial t} - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}^s(\mathbf{u}^s) = \mathbf{0} \qquad \text{in} \quad (0,T) \times \Omega_t^s \,, \tag{24}$$

$$\boldsymbol{\nabla} \cdot \mathbf{v} = \mathbf{0} \qquad \text{in} \quad (0,T) \times \Omega_t \,, \tag{25}$$

$$\mathbf{v}^f|_{t=0} = \mathbf{v}_0 \qquad \text{in} \quad \hat{\Omega}_0 \,, \tag{26}$$

$$\mathbf{v}|_{\Gamma_{t,i}} = \mathbf{g}^f \qquad \text{in} \quad (0,T) \,, \tag{27}$$

$$\boldsymbol{\sigma} \cdot \mathbf{n}|_{\Gamma_{t,i}} = \mathbf{h}^f \qquad \text{in} \quad (0,T) \,. \tag{28}$$

In order to take into account the presence of a multiphase fluid that interacts with a solid domain, we modify the first equation of (23) while the rest of the problem is treated as described in the previous Section. The momentum balance equations in the fluid domain are formally written as in single phase formulation

$$\begin{aligned} &\rho^f \frac{\partial \mathbf{v}^f}{\partial t}\bigg|_{\tilde{\mathcal{A}}} + \rho^f \left(\mathbf{v}^f - \mathbf{w}^f\right) \cdot \boldsymbol{\nabla} \mathbf{v}^f - \nabla \cdot (\mu(\nabla \mathbf{v}^{\mathbf{f}} + (\nabla \mathbf{v}^{\mathbf{f}^\mathsf{T}})) + \nabla p = \rho^f \mathbf{g} \,, \\ &\nabla \cdot \mathbf{v}^{\mathbf{f}} = 0 \,, \qquad \mathbf{x} \in \Omega, t \in [0,T] \,, \end{aligned} \tag{29}$$

and the difference lies in the definition of the physical properties ρ^f and μ^f. If we denote with l the properties of the reference phase and with g the values associated to the secondary phase, we can define the physical properties ρ^f and μ^f as

$$\rho^f = \rho_l \chi + \rho_g (1 - \chi) \,, \tag{30}$$

$$\mu^f = \mu_l \chi + \mu_g (1 - \chi) \,, \tag{31}$$

where χ is the color or indicator function. This function describes the distribution of the two phases in the domain. It is equal to 1 in the reference phase and 0 in the secondary phase. We note that the function χ is discontinuous on the interface Γ_s. We can define χ as

$$\chi(\mathbf{x}, t) = \int_{\Omega_l(t)} \delta(\mathbf{x}' - \mathbf{x}) \, d\mathbf{x}' \quad \forall \mathbf{x} \in \Omega \,. \tag{32}$$

The indicator function is therefore a multidimensional Heaviside function that changes value on Γ_s. We can also write that

$$\nabla \chi = -\int_{\Gamma_s} \delta(\mathbf{x}' - \mathbf{x}) \mathbf{n}' \, dS' = -\mathbf{n} \int_{\Gamma_s} \delta(\mathbf{x}' - \mathbf{x}) \, dS' = -\mathbf{n} \delta_s(\mathbf{x}) \,, \tag{33}$$

where $\delta_s(\mathbf{x})$ is the Dirac delta function that is discontinuous on Γ_s. Under the hypotheses of immiscible fluids with no phase change, the color function behaves like a passive scalar

and is purely transported by the velocity field, following the simple advection equation

$$\frac{\partial \chi}{\partial t} + (\mathbf{v^f} \cdot \nabla)\chi = 0\,, \qquad \text{in } \Omega \times [0,T]\,. \tag{34}$$

The Volume of Fluid algorithm is employed to solve (34). Given an initial state of the indicator function (32) the VOF algorithm is implemented in two steps. First the interface is advected with an explicit Lagrangian advection method using the fluid velocity as computed by the FSI solver and the normal vector to the interface. Then the interface is reconstructed based on the color function gradient and on the conservation of volume constraint. In this work we use a Piecewise Linear Interface Calculation (PLIC) reconstruction scheme together with an ELVIRA method for the computation of the interface normals. For detailed information on this method the interested reader can consult [19, 13, 7, 4, 3].

3 NUMERICAL RESULTS

In this Section we report some numerical results obtained by implementing the mathematical model just described in a computational platform where the dedicated solvers for the FSI and the VOF problems are coupled. Inside this platform they can exchange data in a fast and efficient way. In all the test cases the fluid is composed of two phases with different densities and a portion of the domain is a solid region that undergoes to deformations due to the fluid motion. We use standard units of measure for all the results reported.

3.1 Test 1: Dam break

In this first test case we consider the dam break problem. This is a typical benchmark for two-phase flow simulations but in this case we consider a deformable solid on which the dam fluid impacts. The initial configuration of the problem is shown in Figure 3 in particular in the right part of this Figure one can see the solid, the primary and the secondary fluid regions marked by Ω_s, Ω_l, Ω_g, respectively. The global domain $\Omega = \Omega_s \cup \Omega_l \cup \Omega_g$ is the square $[0,1] \times [0,1]$. The geometries of the different parts of the domain are specified by giving the coordinate of some reference points. According to

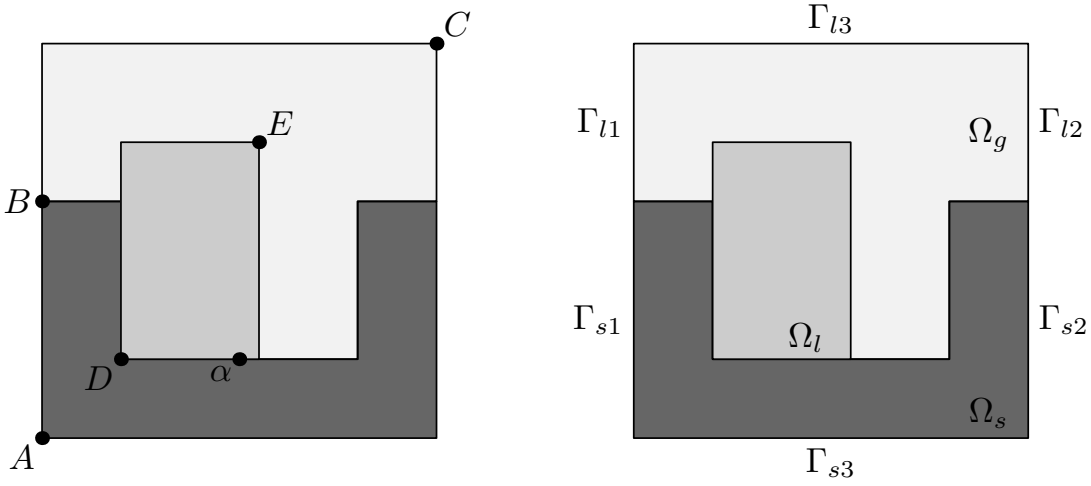
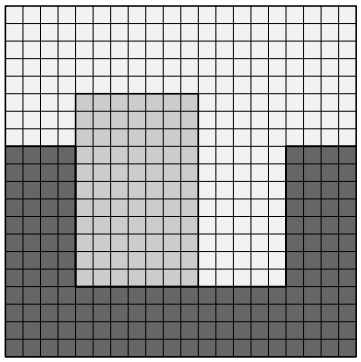


Figure 3: Test 1: Domain overview. On the left some reference points: $A = (0,0)$, $B = (0,0.6)$, $C = (1,1)$, $D = (0.2,0.2)$, $E = (0.5,0.7)$ and $\alpha = (0.2,0.5)$. On the right labeling of the surfaces.



Parameter	Value
ρ_l, ρ_s	500 Kg/m^3
μ_l	0.005 Pa s
ρ_g/ρ_l	0.001
μ_g/μ_l	0.01
Young modulus	$4 \cdot 10^4$ Pa
Poisson coefficient	0.4

Figure 4: Test 1. Coarse computational grid on the left, physical parameters on the right.

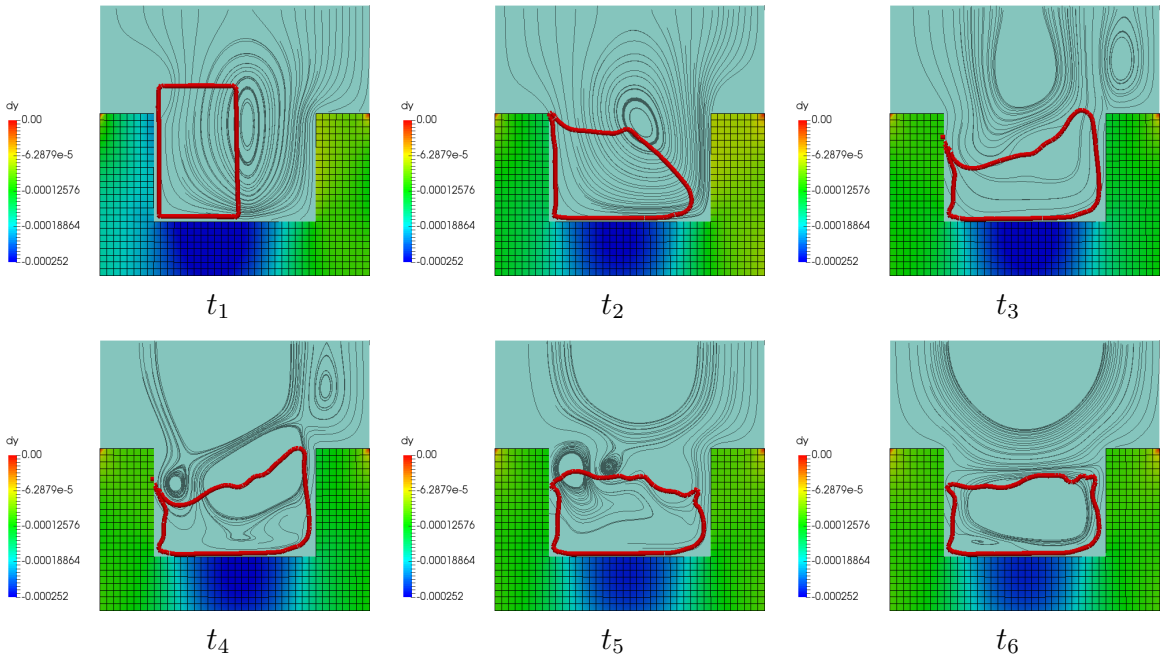


Figure 5: Test 1. Solution overview at different time steps: $t_1 = 0.005$ s, $t_2 = 0.04$ s, $t_3 = 0.08$ s, $t_4 = 0.095$ s, $t_5 = 0.16$ s and $t_6 = 0.2$ s.

the nomenclature shown on the left of Figure 3, the coordinate of the reference points in standard units of measure are: $A = (0, 0)$, $B = (0, 0.6)$, $C = (1, 1)$, $D = (0.2, 0.2)$ and $E = (0.5, 0.7)$. Due to the different densities of the two fluid phases (Ω_l and Ω_g) and the gravitational field, Ω_l falls down into the solid container Ω_s which deforms due to the interaction with both the fluid phases.

The coarse computational grid, both for the FSI and VOF modules, is generated subdividing Ω into 400 cells (20 subdivision per edge) and it is shown on the left of Figure 4. The coarse grid is then refined multiple times to allow for a proper grid resolution for the VOF solver. According to notation shown in Figure 3, we impose a vanishing velocity field on $\Gamma_{l1} \cup \Gamma_{l2}$ and at the boundary fluid structure interface $\{\Gamma_{l1} \cap \Gamma_{s1}\} \cup \{\Gamma_{l2} \cap \Gamma_{s2}\}$. A homogeneous Neumann condition is imposed of the fluid region Γ_{l3} and on the solid boundary Γ_{s3}. The physical properties of the different phases of the problem are summarized in the Table on the right of Figure 4, in particular both of the fluid phases are modeled as Newtonian incompressible fluids while the solid is represented by a compressible linear elastic material.

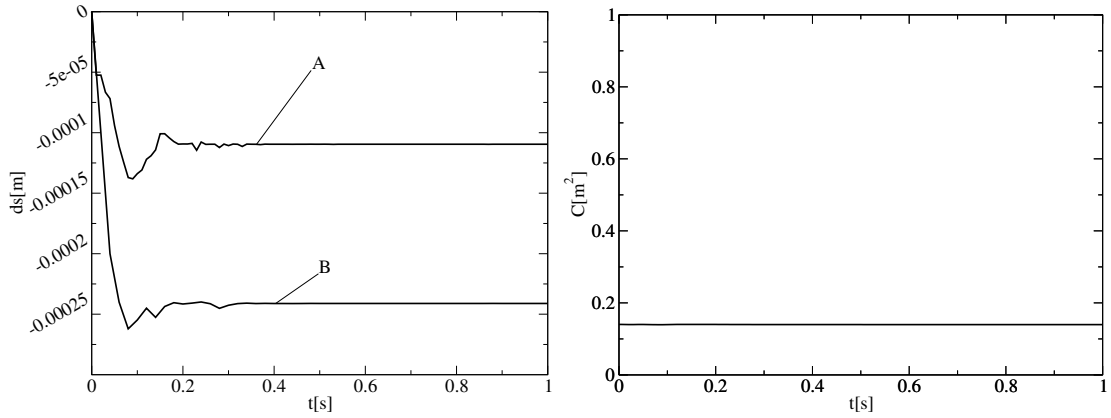


Figure 6: Test 1. On the left, transverse (A) and axial (B) displacement over time. On the right, color function integral over the computational domain over time.

In Figure 5 the solution overview is shown at different time step $t = 0.005$, 0.04, 0.08, 0.095, 0.16 and 0.2 s. In this Figure the axial displacement field is shown in the solid region (Ω_s) while in the fluid part we mark in red the interface between the primary and the primary phase and in light gray the streamlines of the fluid velocity field. We can notice that, as expected, the primary phase with higher density falls down into the solid container that is deformed due to its weight.

The transverse and axial displacements of point α (see left part of Figure 3) over time are shown in the left part of Figure 6 with curves A and B respectively. We can notice that the displacement field after some initial oscillations reaches a steady state configuration and it is always smaller than the characteristic length of the computational grid. In such condition the displacement field can be neglected in the projection of the velocity and the color function during the data exchange between the FSI and VOF meshes. In order to check the performance of the coupled FSI-VOF solver we can evaluate the primary phase volume by computing the color function integral

$$\int_{\Omega_l(t)} d\mathbf{x} = k\,. \tag{35}$$

We remark that the fluid is represented with an incompressible model so because of the free divergence constraint of the velocity field the total primary phase volume k must remain constant over time. The quantity k as computed from (35) is plotted over time on the right of Figure 6.

3.2 Test 2: Tank filling

In the second test case we consider a filling problem in which, as in the previous case, two fluid phases and a solid region are considered. The domain overview is shown in Figure 7, in particular on the left of that Figure the specific geometries of the different parts of the domain are specified by giving the coordinates of some reference points: $A = (0, 0)$, $B = (0.5, 0.15)$, $C = (0.5, 0.5)$ and $D = (1, 1)$. On the right of the same Figure one can see the solid, the primary and the secondary fluid regions marked as Ω_s, Ω_l, Ω_g, respectively. The global domain $\Omega = \Omega_s \cup \Omega_l \cup \Omega_g$ is a square with a surface of $1m^2$.

According to the surface labeling shown in Figure 7, we impose a vanishing velocity field on $\Gamma_{l1} \cup \Gamma_{l2} \cup \Gamma_{l3} \cup \Gamma_{s1} \cup \Gamma_{s2}$. A homogeneous Neumann condition is imposed on the solid region Γ_{s3}. Concerning the multiphase advection problem, the color function

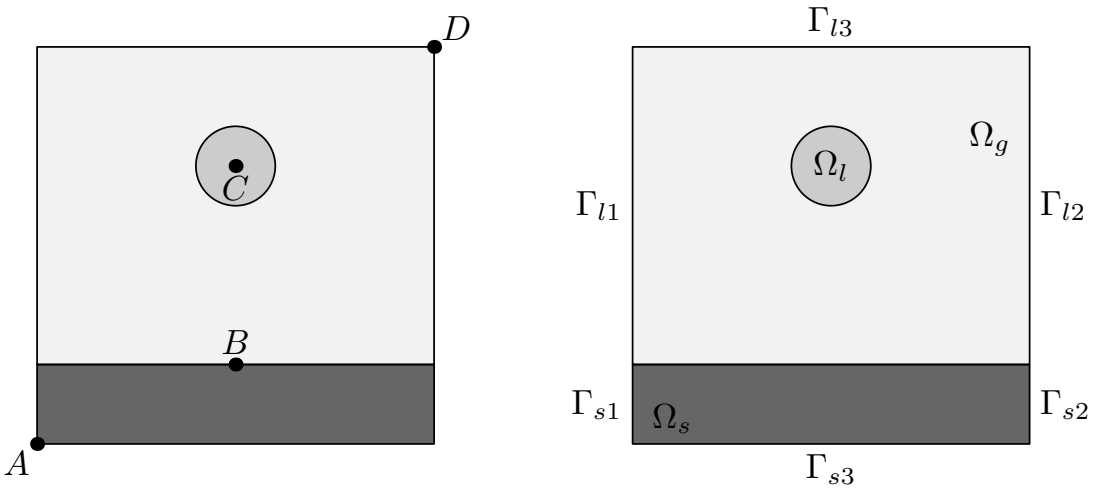


Figure 7: Test 2: Domain overview. On the left reference point: $A = (0m, 0m)$, $B = (0.5m, 0.15,)$, $C = (0.5m, 0.5m)$ and $D = (1m, 1m)$. On the right part labeling of the surfaces.

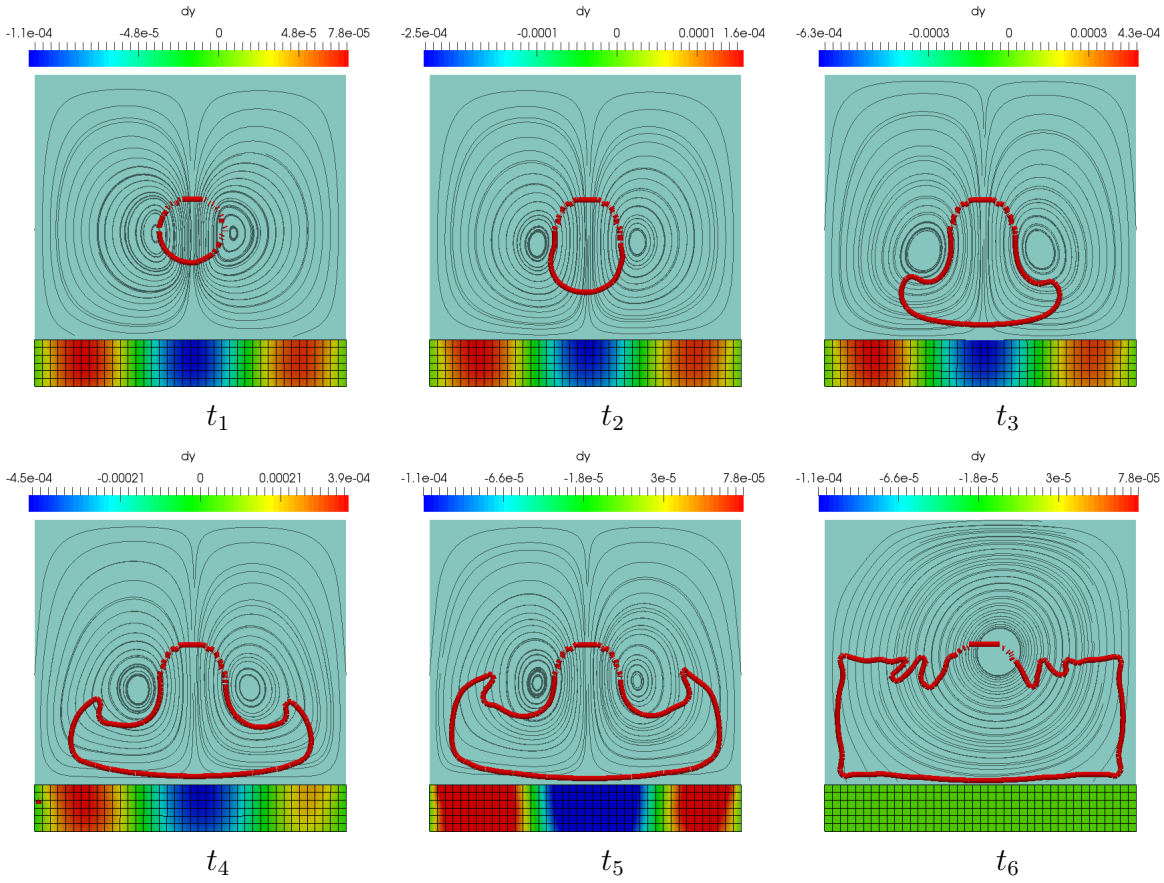


Figure 8: Test 12. Solution overview at different time steps: $t_1 = 0.005s$, $t_2 = 0.09s$, $t_3 = 0.2s$, $t_4 = 0.3s$, $t_5 = 0.4s$ and $t_6 = 2s$.

is set to 1 in the Ω_l region. The physical properties of the different materials present in the problem are the same used in the previous case and can be seen in the Table on the right of Figure 4. The color function is set to 1 in the Ω_l domain for the full time of the simulation, so this volume represents an inflow for the primary phase with velocity computed by the FSI solver. This phase flows inside the domain until it cover the circular

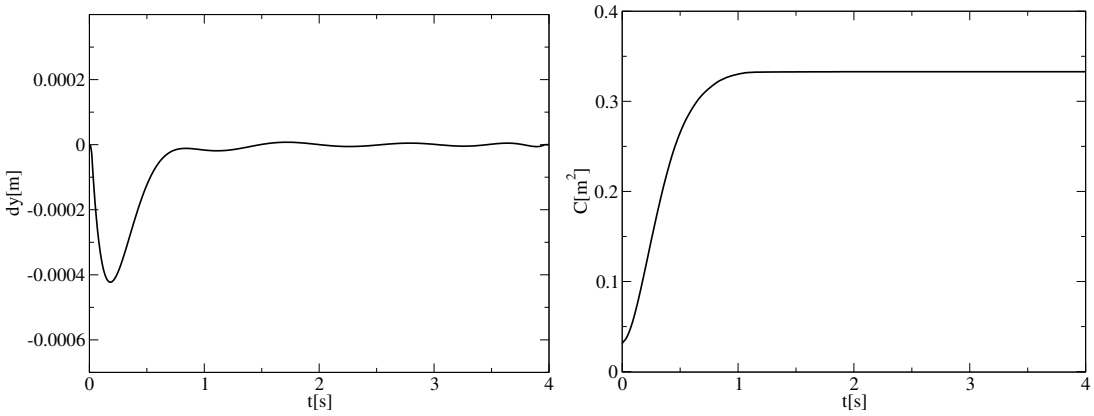


Figure 9: Test 2. On the left, axial displacement over time. On the right, color function integral over the computational domain over time.

inlet region and a steady state configuration is reached.

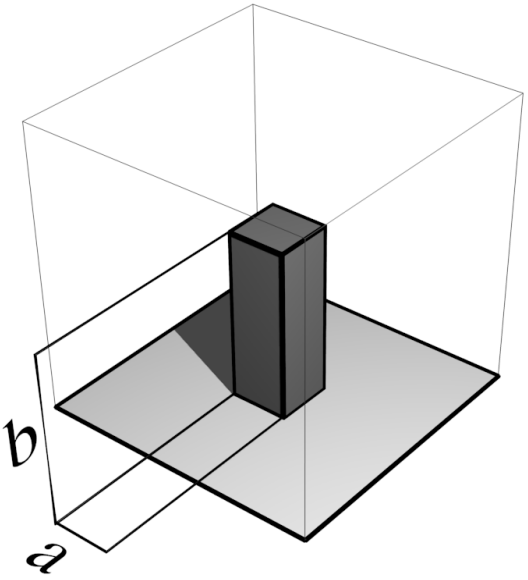
In Figure 8 the solution overview is shown at different time steps $t = 0.005$, 0.09, 0.2, 0.3, 0.4 and 2 s. In the solid region Ω_s we report with colors the axial displacement field while in the fluid region $\Omega_l \cup \Omega_g$ we show the interface between the two phases $\Omega_l \cap \Omega_g$ as a red solid line. The streamlines of the fluid velocity field are reported in light gray in the same Figure. We can notice that as the heavy phase falls because of the gravitational field, more secondary phase is injected in the domain until it cover the inlet region. The axial displacement of the central point of the solid region over time is shown on the left of Figure 9. We can notice that the displacement field shows a great oscillation due to the first contact between the secondary phase and the solid region, after this impact the average axial deformation decreases and reaches a steady negative value when the injection of the secondary phase stops. Also in this case it is worth to notice that the deformations that occurs are always smaller than the characteristic length of the computational grid.

As in the previous case, in order to check the performance of the FSI-VOF coupled modules, we can evaluate the integrated color function (35). In this case because of the inlet of the primary phase the total primary phase volume k must increase over time. The result of (35) along time is shown on the right of Figure 9. we can notice that the value increases as the secondary phase is injected into the domain and reaches a steady value as the injection ends.

3.3 Test 3: Bending rod in two-phase flow

In this third test we consider the flow of a multiphase fluid around a solid deformable obstacle. The domain overview is shown on the left of Figure 10. The global domain is a cube characterized by an edge of 1 m, the obstacle is placed in the center of the cube base with a square transverse section a^2 of 0.01 m^2 and a height b of 0.5 m. The physical parameters employed for this test case are reported in Figure 10 on the right. In this Table we see that the fluid and the solid have the same density, while the density ratio between the primary and secondary phase is close to the water over air ratio. The solid has been modeled as an incompressible solid and the fluids with an incompressible Newtonian model.

The boundary and initial conditions of the problem can be seen in the set of Figures 11. In the colored surface of the sub figure a we set an homogeneous Neumann boundary condition to simulate the outlet of the domain. In the colored surfaces of sub figure



Parameter	Value
ρ_l, ρ_s	500 Kg/m^3
μ_l	0.005 Pa s
ρ_g/ρ_l	0.001
μ_g/μ_l	0.01
Young modulus	$4 \cdot 10^4$ Pa
Poisson coefficient	1/2

Figure 10: Test 3. Domains overview on the left and physical parameters on the right.

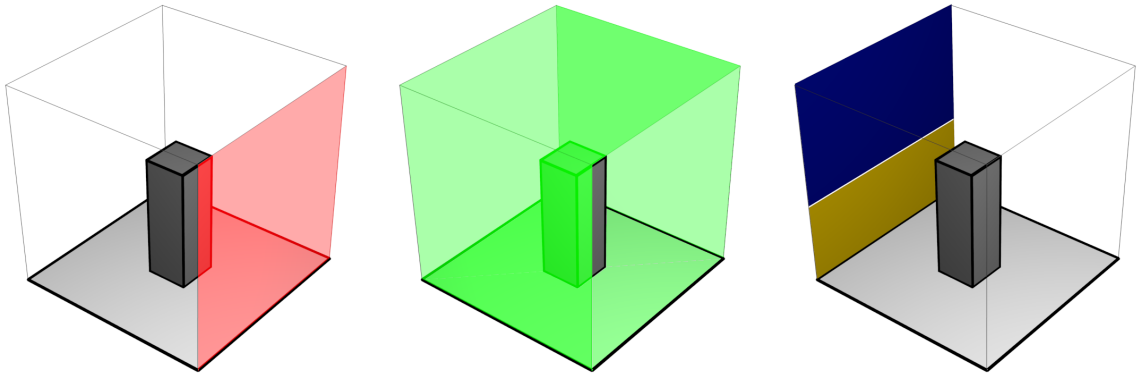


Figure 11: Test 3. Boundary conditions overview: a homogeneous Neuman, b homogeneous Dirichlet and c non homogeneous Dirichlet.

b we impose a no slip condition while in the colored surfaces of c a non homogeneous Dirichlet is set as an inlet. In that surface we impose a vanishing transverse velocity and a non-vanishing axial flows. In particular the fluid flows into the domain with a constant velocity of 1 m/s. Concerning the multiphase problem we consider a stratified flow with the primary heavy phase on the bottom and the secondary phase on top. As initial condition we consider the domain filled with the secondary phase while the boundary condition applied as inlet is visible in sub figure c of Figures 11. The yellow region is occupied by the primary phase with a height of 0.3 while in the blue region the secondary phase has a height of 0.7. During the simulation the heavy phase enters in the domain and flows on the bottom of the domain because of the gravity, following a typical stratified flow pattern.

In this test we consider the obstacle as a deformable solid rod, so the impact with the heavy phase displaces the rod with a force based on the weight and inertia of the fluid phase. One of the main interesting features of the coupled FSI-VOF solver is that the

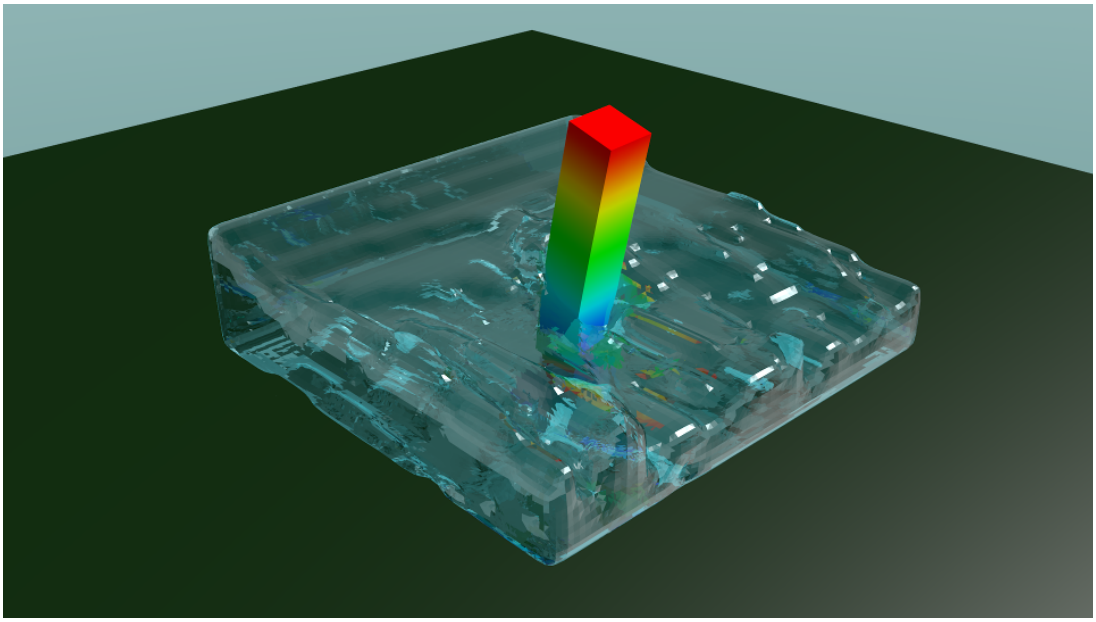


Figure 12: Test 3. Solution overview at $t = 10$ s.

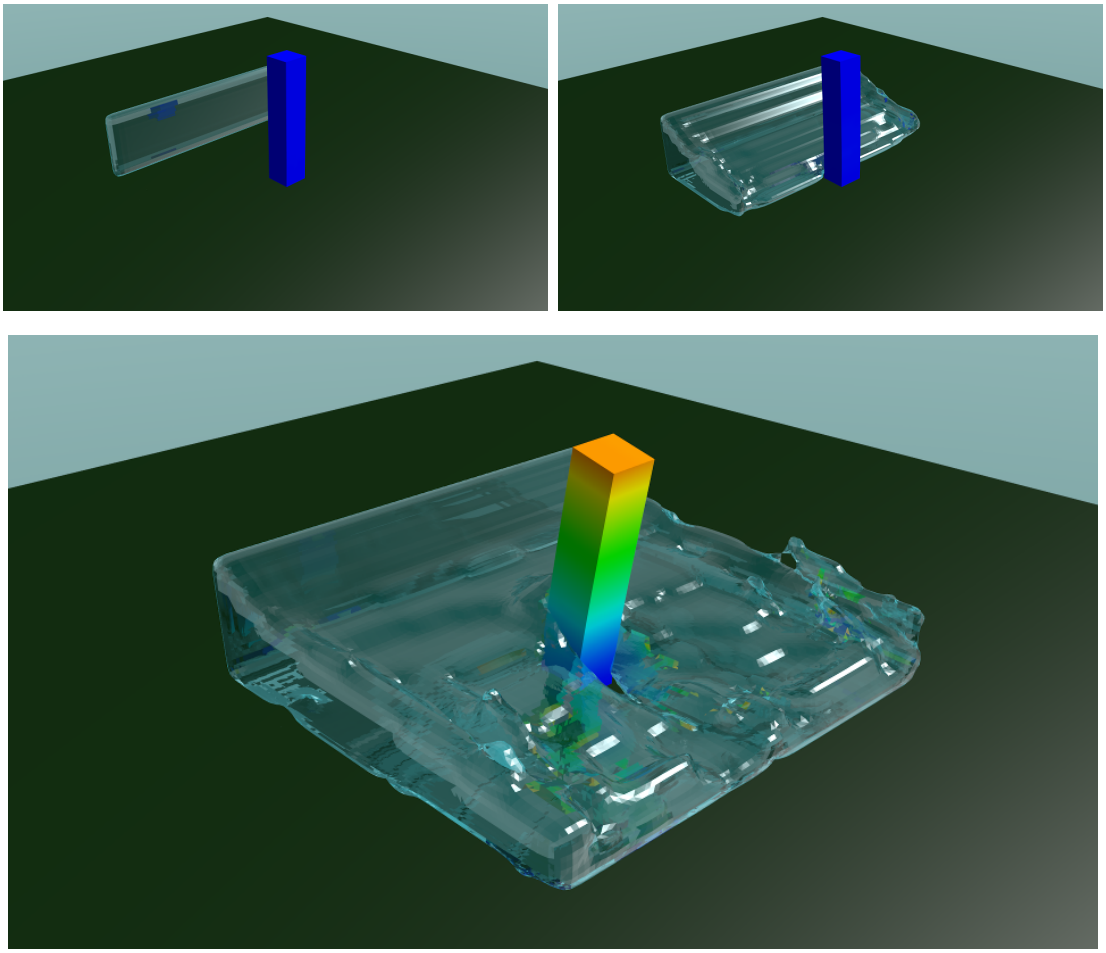


Figure 13: Simulation of the two-phase flow over a bending rod. From top left to bottom right three different time steps: $t_1 = 0.1$ s, $t_2 = 2.5$ s and $t_3 = 5$ s.

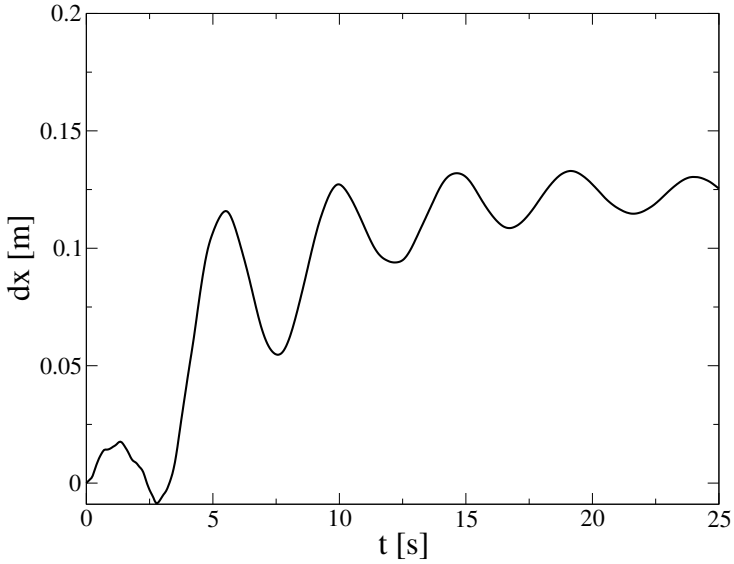


Figure 14: Displacement dx in the main flow direction on the most stressed point of the rod, reported as a function of time.

displacements of the solid can be precisely evaluated together with the stresses inside the structure. Moreover the bending of the solid has an influence on the flow field which could not be taken in consideration without the FSI solver.

In Figure 12 an overview of the solution is shown when the steady state condition is being approached. From this Figure we can see how the multiphase fluid flows into the computational domain and deforms the bending rod which is colored with the displacement in the main direction of the flow. The primary-secondary phase interface is represented with a transparent surface. In Figure 13 the solution overview is shown at different time steps: $t = 0.1$ s, $t_2 = 2.5$ s and $t_3 = 5$ s. From this sequence we can appreciate the evolution of the secondary phase flow together with the deformation of the bending rod. The color of the rod represents the displacement in the main flow direction. It can be seen that, due to the impact with the secondary phase, the rod bends and begins to oscillate. In Figure 14 the displacement of the central point of the top surface of the solid part in the main flow direction is reported as a function of time. The oscillating damped behavior of the rod in the main direction (x axis) can be clearly seen in this Figure. We remark that although the deformation in the top part of the rod are large, where the solid interact with the secondary phase the deformation field is still smaller then the characteristic length of the VOF grid. In such condition the displacement can be neglected in the projection of the computational fields among the different grids.

4 CONCLUSION

In this paper we have presented a model for the multiphase simulation of Fluid Structure Interaction problems based on a monolithic FSI approach and a Volume of Fluid method. The domain consists of a fluid and a solid region. Inside the fluid two phases are present with different physical properties. The solution of this problem is accomplished through the coupling of the FSI and VOF solvers using dedicated data exchange libraries. The results reported show that this method is capable to predict the displacements and stresses that a solid undergoes while interacting with a two phase fluid. Some improvements for future works could be to take into account the surface tension for the secondary

phase and consider cases where large displacements occur where the secondary phase is present.

REFERENCES

[1] F. Aubert, E. Aulisa, S. Manservisi, and R. Scardovelli. Interface tracking with dynamically-redistributed surface markers in unstructured quadrangular grids. *Computers & fluids*, 35 (10):1332–1343, 2006.

[2] J.P. Aubry. *Beginning with Code Aster. A Practical Introduction to Finite Element Method using CodeAster, Gmsh and Salome*. Framasoft, 2013.

[3] E. Aulisa, A. Cervone, S. Manservisi, and P. Seshaiyer. A multilevel domain decomposition approach for studying coupled flow applications. *Communications in Computational Physics*, 6(2):319–341, 2009.

[4] E. Aulisa, S. Manservisi, and P. Seshaiyer. A computational multilevel approach for solving 2d navier-stokes equations over non-matching grids. *Computer Methods in Applied Mechanics and Engineering*, 195(33-36):4604–4616, 2006.

[5] Daniele Cerroni, Sandro Manservisi, and Filippo Menghini. A new moving mesh technique for monolithic multigrid fluid-structure interaction solver. *Recent Advances in Civil Engineering and Mechanics*, 1:146–152, 2014.

[6] Daniele Cerroni, Sandro Manservisi, and Filippo Menghini. An improved monolithic multigrid Fluid-Structure Interaction solver with a new moving mesh technique. *International Journal of Mathematical Models and Methods in Applied Sciences*, 9:227–234, 2015.

[7] A Cervone, Sandro Manservisi, and Ruben Scardovelli. A fem solver coupled to a multilevel vof method for simulation of axisymmetric jets and to a front-tracking method for simulation of spreading droplets. *Atomization and Sprays*, 20(2), 2010.

[8] P.G. Ciarlet. *The finite element method for elliptic problems*, volume 40. Siam, 2002.

[9] J. Glimm, M. J. Graham, J. Grove, X. L. Li, T.M. Smith, D. Tan, F. Tangerman, and Q. Zhang. Front tracking in two and three dimensions. *Computers & Mathematics with Applications*, 35(7):1–11, 1998.

[10] Arthur Guittet, Maxime Theillard, and Frédéric Gibou. A stable projection method for the incompressible navier–stokes equations on arbitrary geometries and adaptive quad/octrees. *Journal of Computational Physics*, 292:215–238, 2015.

[11] Matthias Heil. An efficient solver for the fully coupled solution of large-displacement fluid–structure interaction problems. *Computer Methods in Applied Mechanics and Engineering*, 193(1):1–23, 2004.

[12] P. Le Tallec and J. Mouro. Fluid structure interaction with large structural displacements. *Computer Methods in Applied Mechanics and Engineering*, 190(24):3039–3067, 2001.

[13] J.E. Pilliod and E. G. Puckett. Second-order accurate volume-of-fluid algorithms for tracking material interfaces. *Journal of Computational Physics*, 199(2):465–502, 2004.

[14] G Ryskin and LG Leal. Numerical solution of free-boundary problems in fluid mechanics. part 1. the finite-difference technique. *Journal of Fluid Mechanics*, 148:1–17, 1984.

[15] PA Sackinger, PR Schunk, and RR Rao. A newton–raphson pseudo-solid domain mapping technique for free and moving boundary problems: a finite element implementation. *Journal of Computational Physics*, 125(1):83–103, 1996.

[16] Ruben Scardovelli and Stéphane Zaleski. Interface reconstruction with least-square fit and split eulerian–lagrangian advection. *International Journal for Numerical Methods in Fluids*, 41(3):251–274, 2003.

[17] T. Tezduyar, Y. Bazilevs, and K. Takizawa. *Computational Fluid-Structure Interaction*. Wiley, 2013.

[18] D. Thakore. *Finite Element Analysis using Open Source Software*. Moonish Enterprises, 2014.

[19] Grétar Tryggvason, Ruben Scardovelli, and Stéphane Zaleski. *Direct numerical simulations of gas–liquid multiphase flows*. Cambridge University Press, 2011.

[20] Olgierd Cecil Zienkiewicz, Robert Leroy Taylor, Olgierd Cecil Zienkiewicz, and Robert Lee Taylor. *The finite element method*, volume 3. McGraw-hill London, 1977.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A METHOD OF CHAINED ANALYTICAL WAVE STRUCTURES FOR LARGE-SCALE STRATIFIED TWO-PHASE PIPE FLOWS (ECCOMAS CONGRESS 2016)

A.H. Akselsen[1]

[1]Department of Energy and Process Engineering, Norwegian University of Science and Technology
Kolbjørn Hejes v. 1B, 7491 Trondheim, Norway
e-mail: andreas.h.akselsen@ntnu.no

Keywords: Chained Analytical Waves; Roll-Waves; Two-Phase Flow; Pipe Flow; Two-Fluid Model

Abstract. *Pipeline transport systems can span vast distances, making finely resolved simulation techniques computationally unaffordable. This article examines the use of analytical steady wave solutions in a dynamic simulation framework as a technique for obtaining physical wave dynamics in a flow regime simulator without discretising individual wave structures. A scheme based on this principle is presented wherein a family of steady roll-waves profile solutions is generalised to allow for a dynamic profile evolution. These solutions are then implemented into a finite volume scheme where a wave solution constitutes a single dynamic grid cell. Waves interact dynamically to construct a wave regime evolving in time. Predictions on flow development are compared with finely resolved direct numerical simulations on fixed grids. Steady wave solutions, once subjected to coordinate stretching, behaves appropriately in the dynamic frame, predicting the regime development in time, the wave coalescences and the final steady state or slug transition.*

1 INTRODUCTION

1.1 Review

Roll-wave trains are periodic occurrences of moving hydraulic jumps and constitutes a shallow-water flow regime in both channels and pipes. The wave dynamic of this flow regime is important in two-phase pipe flows as it affect the momentum exchange between phases, and because it may initiate the transition to a slugging flow.

Profile solutions of steady-state hydraulic jumps in open channels was published by Brass already in 1868.

Thomas [17] presented, in 1937 and 1939, a family of explicit, mechanistic, free surface roll-wave solutions using a reference system relative to the waves through a moving belt analogy.

Dressler [7] later formalised these profile solutions into a closed form composed of trains of piecewise monotonic solutions connected by shocks. He formulated conservation conditions for these shocks. Dressler also went on to prove the uniqueness of these solutions through entropy considerations, and to also construct viscous, continuous profile solutions. Thomas and Dressler both showed, by different means, that some amount of friction was a necessary requirement for the formation of roll-waves.

Miya et. al. [14] went on to derive similar profile solutions for gas-liquid channel flows, also including shape factors for the velocity profiles. They further investigated the pressure distribution and compared the profile solutions to experimental data.

Watson [18] reformulated the gas-liquid solution for flow in pipes. This model had a form similar to that for channel flow, but with a geometrical complexity making it unsuited for analytical integration. Algebraically explicit profile solutions are therefore unavailable for pipe flow. The increased complexity of these equations also makes solution uniqueness difficult to prove; this was instead assumed.

Johnson et. al. [11, 10] continued the work on discontinuous and viscous continuous roll-wave models for stratified pipe flow. They constructed a database system for retrieving profile solutions at given flow conditions. In [12], Johnson et. al. compare high-pressure, upward inclined pipe flow experiments to the continuous roll-wave model to find good agreement.

Comparisons between roll-wave experiments and predictions from finely resolved numerical representations have been made by multiple authors. For instance, Holmås [9] compared a pseudospectral representation (using fast Fourier transformation) of the incompressible two-fluid pipe flow model with the above cited experiments of Johnson. The Biberg model [2] for pre-integrated turbulent shear and velocity profiles was here incorporated.

Similar comparisons were made by Cao et. al. [5], including $k-\varepsilon$ turbulent closures to their Riemann solver + MUSCL method. They report good agreement with the experimental data of Brock's [3] after the $k-\varepsilon$ extension.

Richard and Gavrilyuk [15] extended Dressler's roll-wave solutions to account for turbulent shear and dissipation. Reynolds' stresses are here related to enstrophy in the wave, providing wave-breaking as a model extension. Very good agreement with Brock's experimental data was found [3], appropriately breaking off the sharp wave tip of the Dressler solutions.

In the work of Brook et. al. [4], the solutions of Dressler were compared to roll-wave simulation results of the shallow water equations using a second order Godunov method.

A final work worth mentioning in relation to these steady wave solutions is Lahey's piece on dispersed bubbly pipe flow [13]. Here, void waves were analysed using non-linear theory, finding similar sets of structures and profile solutions.

In regard to the method principle, some ideas have been drawn from the work of De Leebeeck

and Nydal [6]. They presented a phenomenological model treating large waves as a fixed-length choke object and used object oriented programming to simulate intermittent occurrences of such wave representations.

1.2 Aim

The aim in this work is to whether solutions for steady wave trains can be used in modelling dynamically evolving wavy flows. Will a set of spurious analytical waves merge and adjust to a 'fully developed' flow pattern? Does the method have potential for use in engineering simulator tools?

The concept is tested through a method based on the principle of singleton wave solutions in the frame of a finite volume method. Predicting the development of a *flow regime* in long pipeline simulations is usually more important than predicting individual waves. We need therefore not demand exact conformity with the base model, but that the appropriate statistical development is ensured and that predictions on regime transition are reliable. Similar methods can be devised for any type of regime flow where reasonably steady intermittent structures appear.

1.3 Introduction to a method of dynamically chained analytical wave structures

In this article, the steady state solutions presented by the aforementioned authors will be generalised and introduced into a dynamic, computational framework for simulating transient wave development. (The term 'steady state' will here refere to solutions which are steady in a reference system moving with constant translation velocity.) For ease of reference, the resulting method will be dubbed *the method of chained analytical waves,* CAW for short.

The concept of the CAW method is presented in Figure 1.1 with explanations in the caption. It consists of four basic steps:

i) A profile and shock celerity reconstruction from the average state properties in a relative reference system.

ii) Integrating the average wave properties in time using fluxes and sources obtained from the chained profiles.

iii) Letting the wave objects move according to the shock celerities.

iv) Reconstructing the profiles and shock celerities anew from the new average properties.

The two-fluid model is presented in Section 2. A conservative finite volume method for dynamic cells is then presented in Section 3.1, with the average volume equations in Section 3.1.1 and the shock conditions in Section 3.1.2. The analytical wave model is found in Section 3.2, with the general profile equations presented in Section 3.2.2. Section 3.3 deals with selecting a wave profile from the family of profile solutions on the basis of the average properties. Putting these pieces together, Section 3.4 presents the simulation routine.

Figure 1.2 illustrates the main difference between the CAW method and conventional, fine-gridded CFD methods, namely the much greater number of grid cells required by the latter. The numerical stability of both types of solution methods will be restricted by CFL-type criteria which limit the length of the time integration step in proportion to the grid cell lengths. Consequently, the CAW method should be capable of remaining stable while using significantly longer time steps, at a time scale better suited for wave dynamics.

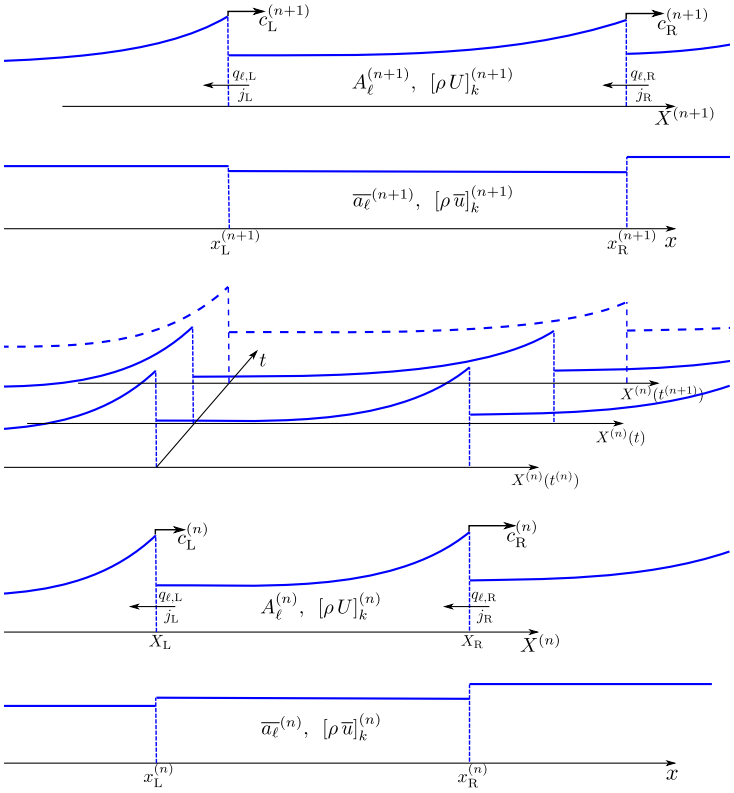
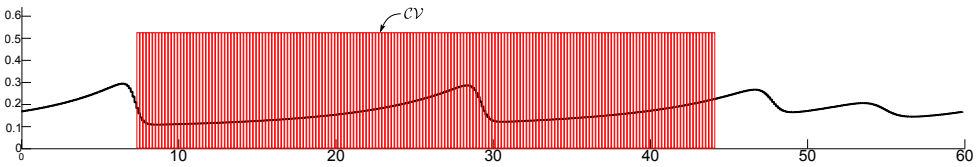
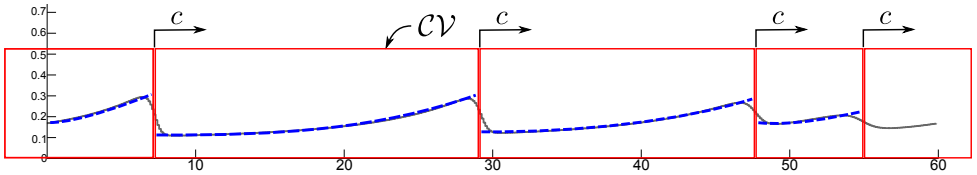


Figure 1.1: Principle illustration.
• Bottom: Available average state $\overline{\boldsymbol{\psi}}$
• 2nd from bottom: Adopt a reference system relative to the wave. Reconstruct the waves with the analytical wave model and determine the shock celerities and the flux exchanges between waves (Section 3.2-3.3.)
• Middle: Integrate in time using the fluxes and shock speeds obtained at $t^{(n)}$ from the analytical wave regime. Allow the wave to stretch and contract in the strip $t \in \left[t^{(n)}, t^{(n+1)}\right\rangle$, (Section 3.1.)
• 2nd from top and top: The time integration provides the average state $\overline{\boldsymbol{\psi}}^{(n+1)}$ at time $t^{(n+1)}$, and the cycle repeats for the new time step.



(a) Conventional capturing simulation with a fine, stationary grid – 500 grid cells.



(b) Chained analytical wave simulation – 4 grid cells.

Figure 1.2: Control volumes

2 THE TWO-FLUID MODEL

In addition to the presence of a gas, the complication of a pipe geometry is a key difference between the explicit shallow water roll-wave solutions of Dressler and the numerically computed pipe flow solutions of Watson – Figure 2.1 shows a schematic. The pipe introduces new

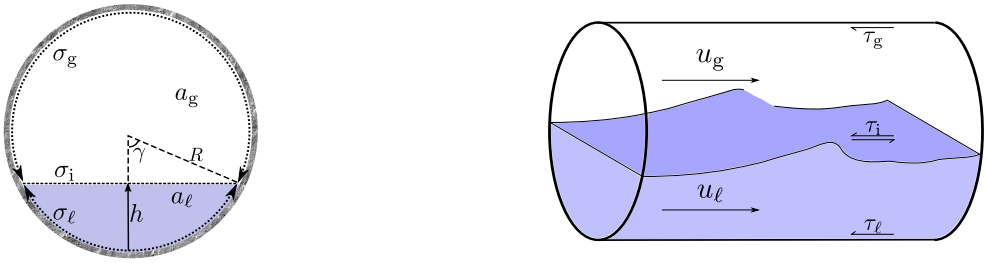


Figure 2.1: Pipe cross-section

geometric variables such as the level height h, partial cross-section areas a_k and perimeter length σ_k for the gas ($k = \mathrm{g}$) and liquid ($k = \ell$) phases. These are all algebraically interchangeable through a geometric function $a_\ell = \mathcal{A}_\ell(h)$:

$$\mathcal{A}_\ell(h) = R^2\left(\gamma - {}^1\!/_2 \sin 2\gamma\right), \qquad \sigma_\ell = 2R\gamma,$$
$$\mathcal{A}'_\ell = \frac{\mathrm{d}\mathcal{A}_\ell}{\mathrm{d}h} = \sigma_\mathrm{i}, \qquad \sigma_\mathrm{g} = 2R\left(\pi - \gamma\right),$$
$$\gamma = \arccos\left(1 - h/R\right), \qquad \sigma_\mathrm{i} = 2R\sin\gamma.$$

Though expressions of the same property, the level height h is beat suited as an independent modelling variable as a_k and σ_k are explicit functions in h, whereas the inverse is not true. R is here the pipe inner radius and γ the interface half-angle. Subscript 'i' indicates the gas-liquid interface.

The compressible, isothermal, four-equation two-fluid model for stratified pipe flow results from an averaging of the conservation equations across the cross-sectional area and is commonly written

$$\partial_t\left(a\rho\right)_k + \partial_x\left(a\rho u\right)_k = 0, \tag{2.1a}$$
$$\partial_t\left(a\rho u\right)_k + \partial_x\left(a\rho u^2\right)_k + a_k\partial_x p_\mathrm{i} + a_k\rho_k g\cos\theta\,\partial_x h = s_k, \tag{2.1b}$$

where phase subscript $k \in \{\mathrm{g}, \ell\}$. p_i is the interface pressure, coupled to the fluid densities ρ_k by some equation of state.

The phase properties represent cross-sectional averages in each field, and, together, the two last left-hand terms in the momentum equation represents a hydrostatic approximation to the average pressures over the cross-section. A flat, fully turbulent velocity profile is here assumed.

The numerical treatment of the pressure term is a key issue in simulating the two-fluid model. It introduces sonic characteristics and stability restrictions. In this work, as well as in most of those hereto cited, pressure issues are numerically sidestepped by assuming both phases to be incompressible. This eliminates the sonic characteristics while the system still retains its conditionally hyperbolic nature. It also allows for a conservative, two-equation system formulation. The interfacial pressure, essentially linking the momentum equations of each phase, is eliminated by first reducing each momentum equation by their respective mass equations, and then

divided each by their respective specific areas a_k. Subtracting one from the other, the pressure term is eliminated. The resulting conservation equations reads

$$\partial_t \boldsymbol{\psi} + \partial_x \boldsymbol{f} = \boldsymbol{s} \tag{2.2a}$$

with

$$\boldsymbol{\psi} = \begin{pmatrix} a_\ell \\ [\rho u]^\ell_{\mathrm{g}} \end{pmatrix}, \qquad \boldsymbol{f} = \begin{pmatrix} q_\ell \\ j \end{pmatrix}, \qquad \boldsymbol{s} = \begin{pmatrix} 0 \\ s \end{pmatrix}. \tag{2.2b}$$

Flux and source component symbols have here been introduced and are

$$q_k = a_k u_k, \qquad j = \left[\rho \frac{u^2}{2}\right]^\ell_{\mathrm{g}} + w_y\, h, \tag{2.2c}$$

$$s = -w_x - \left[\frac{\tau\sigma}{a}\right]^\ell_{\mathrm{g}} + \tau_{\mathrm{i}}\sigma_{\mathrm{i}}\left(\frac{1}{a_\ell} + \frac{1}{a_{\mathrm{g}}}\right), \tag{2.2d}$$

and phase differencing operation $[\cdot]^\ell_{\mathrm{g}} = (\cdot)_\ell - (\cdot)_{\mathrm{g}}$ which will be used throughout. The algebraic relations

$$a_\ell + a_{\mathrm{g}} = \mathcal{A}, \qquad q_\ell + q_{\mathrm{g}} = \mathcal{Q}, \tag{2.2e}$$

close the model. This is essentially the same formulation as used by Holmås in [9]. $\mathcal{A}$ may in general be a parametric function of x and $\mathcal{Q}$ of t, though they are here treated as constants. The second relation in (2.2e) is obtained from adding together the two mass equations (2.1a) and applying the first relation. Specific weight terms, also parametric, are

$$w_x = [\rho]^\ell_{\mathrm{g}}\, g \sin\theta, \qquad w_y = [\rho]^\ell_{\mathrm{g}}\, g \cos\theta,$$

with g being the gravitational acceleration and θ the pipe inclination angle, positive above datum. The friction model τ_k is based on the rather simple Taitel and Dukler model [16] for turbulent flow, quickly summarised below:

$$\begin{aligned} &\tau_k = {}^1\!/_2\, C_{\mathrm{f},k}\, \rho_k u_k |u_k|, && \tau_{\mathrm{i}} = {}^1\!/_2\, C_{\mathrm{f,i}}\, \rho_{\mathrm{g}} (u_{\mathrm{g}} - u_\ell)|u_{\mathrm{g}} - u_\ell|, \\ &C_{\mathrm{f},k} = 0.046\, Re_k^{-0.2}, && Re_k = (ud/\nu)_k, \\ &d_\ell = 4\, a_\ell/\sigma_\ell, && d_{\mathrm{g}} = 4\, a_{\mathrm{g}}/(\sigma_\ell + \sigma_{\mathrm{i}}). \end{aligned}$$

The interfacial friction factor is $C_{\mathrm{f,i}} = m_{\mathrm{x}} \max\{C_{\mathrm{f,g}}, 0.014\}$ where m_{x} is a crude model parameter employed by numerous authors in order to achieve $C_{\mathrm{f,i}} > C_{\mathrm{f,g}}$ as observed experimentally. ν is the kinematic viscosity.

A study on how m_{x} affect the wave regime was carried out in [11] and will not be repeated here. Instead, m_{x} is rather arbitrarily given the value $m_{\mathrm{x}} = 5.0$. The choice of m_{x} affects the range of flow conditions in which the stable roll-wave regime is observed.

3 A METHOD OF CHAINED ANALYTICAL WAVE STRUCTURES

Details of the CAW method are described in this section. The process of averaging the base equation system (2.2) into movable, stretching and contracting control volumes is described in Subsection 3.1, while details of the wave model are described in Subsection 3.2. Finally, a routine description is given in Subsection 3.4.

Readers who are foremost interested in the method principle (as presented in the introductory section) may skip past the method details presented here and still gather an impression from the results section that follows.

3.1 A finite volume method framework

3.1.1 Average volume equations

A control volume is placed over a wave, as illustrated in Figure 1.2b. Integration of (2.2a) is performed across the control volume, first in space from the left shock at x_L to the right one at x_R, and then in time from the present time t_n to the next time level t_{n+1}:

$$\int_{t_n}^{t_{n+1}} \int_{x_\mathrm{L}(t)}^{x_\mathrm{R}(t)} \left(\partial_t \boldsymbol{\psi} + \partial_x \boldsymbol{f} - \boldsymbol{s}\right) \mathrm{d}x \, \mathrm{d}t = 0. \tag{3.1}$$

Using Leibniz' rule, the first transient term evaluates to

$$\int_{x_\mathrm{L}(t)}^{x_\mathrm{R}(t)} \partial_t \boldsymbol{\psi} \, \mathrm{d}x = \partial_t \big(\Delta x \overline{\boldsymbol{\psi}}\big) - \left[c \, \boldsymbol{\psi}\right]_\mathrm{L}^\mathrm{R} \tag{3.2}$$

where $c = \frac{\mathrm{d}x}{\mathrm{d}t}$ is the control volume border velocity, chosen equal to the shock speed of the wave front discontinuity, such that the control volume follows the wave front. $\Delta x(t) = x_\mathrm{R}(t) - x_\mathrm{L}(t)$ is the wavelength, and the notation for evaluating the left-right difference $[\cdot]_\mathrm{L}^\mathrm{R} = (\cdot)_\mathrm{R} - (\cdot)_\mathrm{L}$ is adopted. The bar will indicate the cell average and is defined

$$\overline{\phi}(t) = \frac{1}{\Delta x(t)} \int_{x_\mathrm{L}(t)}^{x_\mathrm{R}(t)} \phi(x,t) \, \mathrm{d}x. \tag{3.3a}$$

Also introducing the temporal average $\langle \cdot \rangle^{(n)}$,

$$\big\langle \phi(x) \big\rangle^{(n)} = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \phi(x,t) \, \mathrm{d}t \tag{3.3b}$$

with $\Delta t = t_{n+1} - t_n$, the integral equation (3.1) is cast as

$$\big(\Delta x \overline{\boldsymbol{\psi}}\big)^{(n+1)} = \big(\Delta x \overline{\boldsymbol{\psi}}\big)^{(n)} + \Delta t \left\langle \Delta x \overline{\boldsymbol{s}} - [\boldsymbol{f}_\mathrm{r}]_\mathrm{L}^\mathrm{R} \right\rangle^{(n)}, \tag{3.4}$$

where the flux terms have been made relative to the shock propagation, i.e., $\boldsymbol{f}_\mathrm{r} = (q_{\ell,\mathrm{r}}, j_\mathrm{r})^T = \boldsymbol{f} - c\boldsymbol{\psi}$. Relative notation will be adopted throughout to indicate when fluxes are relative a moving frame. The mass and momentum fluxes of the relative frame are key variables and are therefore emphasised:

$$q_{\ell,\mathrm{r}} = a_\ell \left(u_\ell - c\right), \qquad\qquad j_\mathrm{r} = \left[\rho u \left(\frac{u}{2} - c\right)\right]_\mathrm{g}^\ell + h w_y. \tag{3.5}$$

It is important to note that Equation (3.4) is still exact.

In case of a first order time integration, $\langle \cdot \rangle^{(n)}$ simply evaluates to

$$\langle \cdot \rangle^{(n)} \xrightarrow{\text{explicit}} (\cdot)^{(n)}, \qquad\qquad \langle \cdot \rangle^{(n)} \xrightarrow{\text{implicit}} (\cdot)^{(n+1)}.$$

Explicit time integration $\langle \cdot \rangle^{(n)} \to (\cdot)^{(n)}$ is used in the CAW method. Equation (3.4) then reads

$$\begin{aligned}
\left(\Delta x \overline{a_\ell}\right)^{(n+1)} &= \left(\Delta x \overline{a_\ell}\right)^{(n)} - \Delta t \left[q_{\ell,\mathrm{r}}^{(n)}\right]_\mathrm{L}^\mathrm{R}, \\
\left(\Delta x \left[\rho \overline{u}\right]_\mathrm{g}^\ell\right)^{(n+1)} &= \left(\Delta x \left[\rho \overline{u}\right]_\mathrm{g}^\ell\right)^{(n)} + \Delta t \left(\Delta x \overline{s} - [j_\mathrm{r}]_\mathrm{L}^\mathrm{R}\right)^{(n)}.
\end{aligned}$$

3.1.2 Shock conditions

The coordinate translation velocity $C(H)$ may dynamically vary over a wave as it stretches or contracts, meaning that celerity c adhere to the shocks themselves rather to a wave as a whole. A shock must obey the condition that the relive fluxes $\boldsymbol{f}_{\mathrm{r}} = (q_{\ell,\mathrm{r}}, j_{\mathrm{r}})^T$ are shock invariant.[1]

Following a shock front $i + \frac{1}{2}$, the common notation for a left and right discontinuity limit is adopted, namely $\phi^- = \phi_{\mathrm{R},i}$ and $\phi^+ = \phi_{\mathrm{L},i+1}$. The conservation conditions $j_{\mathrm{r}}^- = j_{\mathrm{r}}^+$ and $q_{\ell,\mathrm{r}}^- = q_{\ell,\mathrm{r}}^+$ imply $J_{\mathrm{r}}(h^-, q_{\ell,\mathrm{r}}, c) = J_{\mathrm{r}}(h^+, q_{\ell,\mathrm{r}}, c)$. Subtracting $\frac{1}{2}\rho_k c$ on both sides illustrates that this condition corresponds to maintaining a Bernoulli invariant

$$\left[\left[\tfrac{1}{2}\rho\,(u-c)^2\right]_{\mathrm{g}}^{\ell} + w_y\, h\right]_{-}^{+} = 0, \tag{3.6}$$

where $[\cdot]_-^+ = (\cdot)^+ - (\cdot)^-$. This is the jump condition similar to that presented by Watson [18]. Solving for c yields

$$c = \frac{\left[\rho\,[u^2/2]_-^+\right]_{\mathrm{g}}^{\ell} + w_y\,[h]_-^+}{\left[\rho\,[u]_-^+\right]_{\mathrm{g}}^{\ell}}.$$

Conservation of the relative discharge $q_{\ell,\mathrm{r}}$ can be archived through the model choice of $Q_{\ell,\mathrm{r}}(H)$ – see Section 3.2.3.

Once the shock celerities and the integral values (3.4) have been computed, the control volume borders are translated according to $x_{i+\frac{1}{2}}^{(n+1)} = x_{i+\frac{1}{2}}^{(n)} + \Delta t\, c_{i+\frac{1}{2}}^{(n)}$. New wavelengths $\Delta x_i^{(n+1)} = x_{i+\frac{1}{2}}^{(n+1)} - x_{i-\frac{1}{2}}^{(n+1)}$ yield the new average states $\overline{\psi}_i^{(n+1)} = \left(\Delta x \overline{\psi}\right)_i^{(n+1)} / \Delta x_i^{(n+1)}$. A wave sketch with the discussed variable is presented in Figure 3.1.

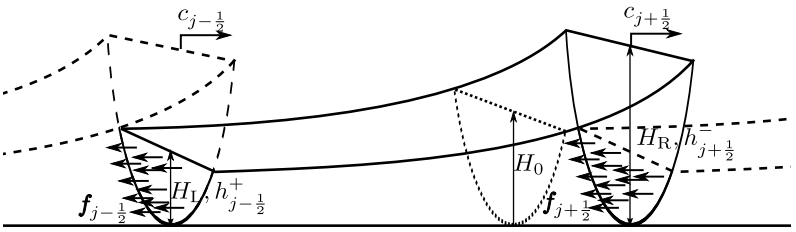


Figure 3.1: A sketch of a wave solution interacting in a wave train.

3.2 An analytical wave model

A local coordinate system $X(x,t)$, moving with celerity $C(x)$, is defined

$$X^{(n)} = x - C^{(n)}t, \qquad \begin{aligned} x_{\mathrm{L}} \le x \le x_{\mathrm{R}}, \\ t_n \le t < t_{n+1}, \end{aligned}$$

[1] To see that both components of $\boldsymbol{f}_{\mathrm{r}}$ are invariants across a shock, let C be the shock speed c and evaluate (3.7a) over an infinitely narrow integral. All but $\boldsymbol{f}_{\mathrm{r}}^+ - \boldsymbol{f}_{\mathrm{r}}^-$ disappears.

where $C^{(n)}$ is generally a function of x, but constant in a steady wave train. Next, translated variables are introduced as $\Phi(X,t) = \phi(x,t)$, reserving upper-case characters for these. Equations (2.2) are now recast in the relative coordinates, yielding

$$\partial_t \boldsymbol{\Psi} + \partial_X \boldsymbol{F}_\mathrm{r} + \boldsymbol{\Psi}\partial_X C = \boldsymbol{S}, \tag{3.7a}$$

$$A_\ell + A_\mathrm{g} = \mathcal{A}, \qquad Q_\ell + Q_\mathrm{g} = \mathcal{Q}, \tag{3.7b}$$

with $\boldsymbol{F}_\mathrm{r} = \boldsymbol{F} - C\boldsymbol{\Psi} = (Q_{\ell,\mathrm{r}}, J_\mathrm{r})^T$ being the fluxes with relative velocities. $Q_{\ell,\mathrm{r}}$ and J_r are of course similar to (3.5) with moving-frame variables.

Note that a non-conservative term $\boldsymbol{\Psi}\partial_X C$ has arisen to represent coordinate stretching. It manifests as a source term because changes in the space within which $\boldsymbol{\psi}$ is conserved are invisible from the stretching reference system.

System (3.7) may also be written in the form of a relative Jacobian $\mathbb{B}_\mathrm{r} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\Psi}} - C\mathbb{I}$, i.e.,

$$\left(\partial_t + \mathbb{B}_\mathrm{r}\partial_x\right)\boldsymbol{\Psi} = \boldsymbol{S} \tag{3.8}$$

where

$$\mathbb{B}_\mathrm{r} = \frac{1}{\rho^*}\begin{pmatrix} (\rho U_\mathrm{r})^* & 1 \\ \kappa^2 & (\rho U_\mathrm{r})^* \end{pmatrix}$$

and

$$\kappa = \sqrt{\frac{\rho^* w_y}{\mathcal{A}'_\ell} - \frac{\rho_\ell \rho_\mathrm{g}}{A_\ell A_\mathrm{g}}\left(U_\mathrm{g} - U_\ell\right)^2}. \tag{3.9}$$

Here, the intrinsic weighting operation

$$\phi^* = \frac{\phi_\ell}{A_\ell} + \frac{\phi_\mathrm{g}}{A_\mathrm{g}}$$

makes its first appearance as an useful shorthand. Eigenvalues of (3.2) are

$$\lambda^\pm_\mathrm{r} = \frac{(\rho U_\mathrm{r})^* \pm \kappa}{\rho^*}. \tag{3.10}$$

3.2.1 The steady profile equation

The profile equation for steady roll-waves are time-invariant, non-stretching solutions of (3.7), i.e., $\partial_t \boldsymbol{\Psi} = \boldsymbol{0}$, $\partial_X C = 0$. Equation (3.7a) then directly yields

$$\frac{\mathrm{d}Q_{\ell,\mathrm{r}}}{\mathrm{d}X} = 0, \qquad\qquad \frac{\mathrm{d}J_\mathrm{r}}{\mathrm{d}X} = S,$$

where the first result implies that $Q_{\ell,\mathrm{r}}$, and therefore also $Q_{\mathrm{g,r}}$, are constant. Applying the chain rule $\frac{\mathrm{d}J_\mathrm{r}}{\mathrm{d}X} = \frac{\mathrm{d}J_\mathrm{r}}{\mathrm{d}H}\frac{\mathrm{d}H}{\mathrm{d}X}$ to the second result yields

$$X' = \frac{J'_\mathrm{r}}{S}, \qquad \text{(steady wave)}. \tag{3.11a}$$

This is the profile equation, inverted. The level height H has been chosen as integration variable and so

$$\phi' = \frac{\mathrm{d}\phi}{\mathrm{d}H}$$

has been defined. Imposing $Q_{k,\mathrm{r}} = \mathrm{const}$, the profile numerator is found to be

$$J'_{\mathrm{r}} = \frac{\mathrm{d}J_{\mathrm{r}}}{\mathrm{d}H} = w_y - \mathcal{A}'_\ell \left(\rho\, U_{\mathrm{r}}^2\right)^* . \tag{3.11b}$$

Relative velocities are of course $U_{k,\mathrm{r}} = U_k - C$. The dynamic frame of the source does not cause any changes as compared to (2.2d), i.e.,

$$S = -w_x - \left[\frac{\mathcal{T}\Sigma}{A}\right]_{\mathrm{g}}^{\ell} + \mathcal{T}_{\mathrm{i}}\Sigma_{\mathrm{i}} \left(\frac{1}{A_\ell} + \frac{1}{A_{\mathrm{g}}}\right). \tag{3.11c}$$

All terms appearing in (3.11a) are functions of H alone.

For the formation of a steady solution to be possible, i.e., for $\partial_t \boldsymbol{\Psi} \to 0$, the roots of J'_{r} and S must coincide in what is known as the 'critical point' H_0: $J'_{\mathrm{r}}(H_0) = S(H_0) = 0$. The wave height increases monotonically with the region $H < H_0$ being hydraulically supercritical ($\mathrm{sgn}\lambda^- = \mathrm{sgn}\lambda^+$) and $H > H_0$ being subcritical ($\mathrm{sgn}\lambda^- \neq \mathrm{sgn}\lambda^+$). This is the profile equation similar to that presented by Watson [18] (alternatively Johnson in [10].)

The well-known 'viscous Kelvin-Helmholtz' criterion for marginal flow stability [1] can easily be inferred directly from the profile equation (3.11) as a zero-amplitude roll-wave. Even a zero-amplitude wave must contain a critical point H_0. By uniformity, all points in a marginally stable plane flow are critical points. In addition, for the wave to remain at zero amplitude, S in (3.11a) must be at an equilibrium with respect to H. Consequentially, the viscous Kelvin-Helmholtz criterion, as presented in [1], may simply be written

$$S = 0, \qquad S' = 0, \qquad J'_{\mathrm{r}} = 0.$$

The first condition is the hold-up equation, providing the steady, waveless stratified state. The second condition determines the critical wave speed while the third forms the criterion of marginal stability. In this way, viscous, stratified stability theory is a natural component of the wave profile solution; breaking wave solutions does not exist unless the uniformly plane stratified flow solution is unstable.

3.2.2 The stretching profile equation

In contrast to the steady wave solution, $\boldsymbol{\Psi}$ is here still a function of time, and C and $Q_{\ell,\mathrm{r}}$ are generally not constant within a wave. The state $\boldsymbol{\Psi}$ is however still approximated as a single variable function in H. For this to be reasonable, the time variation is assumed slow and relatable to the average properties and fluxes. Thus, time dependency is included only indirectly through locally parametric state averages $\boldsymbol{\Phi}(X,t) \to \boldsymbol{\Phi}\big(X; \overline{\psi}(t), \boldsymbol{f}_{\mathrm{L}}(t), \boldsymbol{f}_{\mathrm{R}}(t)\big)$ and differentiated as a single variable function.

$C(H; c_{\mathrm{L}}, c_{\mathrm{R}})$ and $Q_{\ell,\mathrm{r}}(H; q_{\ell,\mathrm{L}}, q_{\ell,\mathrm{R}})$ are modelled locally within the wave to remain globally continuous – Section 3.2.3. These should evolve towards constant values as the steady state solution (3.11a) is approached. Phase velocities are given directly from $U_k = Q_{k,\mathrm{r}}/A_k + C$, with $Q_{\mathrm{g,r}} = Q_{\mathrm{m}} - AC - Q_{\ell,\mathrm{r}}$.

The derivative of the state $\boldsymbol{\Psi} = \left(A_\ell, [\rho\, U]_{\mathrm{g}}^{\ell}\right)^T$, with respect to the single variable H, is

$$\boldsymbol{\Psi}' = \begin{pmatrix} 1 \\ -\left(\rho\, U_{\mathrm{r,adj}}\right)^* \end{pmatrix} \mathcal{A}'_\ell, \tag{3.12}$$

where the adjusted relative velocities

$$U_{k,\mathrm{r,adj}} = U_{k,\mathrm{r}} - \frac{Q'_{\ell,\mathrm{r}} - A_\ell C'}{\mathcal{A}'_\ell} = U_k - \frac{Q'_\ell}{\mathcal{A}'_\ell} \tag{3.13}$$

have been introduced. Equation (3.8) may now be written

$$\left(\boldsymbol{\Psi}'\partial_t + \mathbb{B}_{\mathrm{r}}\boldsymbol{\Psi}'\partial_x\right) H = \begin{pmatrix} 0 \\ S - S_{\mathrm{trans}} \end{pmatrix}, \tag{3.14}$$

where S_{trans} is an artificial source representation of rest terms $\rho^* \left(\partial_t Q_{\ell,\mathrm{r}} + A_\ell \partial_t C\right)$ from the slow transient – it will be discussed shortly. Eliminating $\partial_t H$ between the two equation components of (3.14), one obtains the stretchable inversed profile equation

$$X' = \frac{J'_{\mathrm{r,adj}}}{S - S_{\mathrm{trans}}}, \tag{3.15}$$

with

$$J'_{\mathrm{r,adj}} = w_y - \mathcal{A}'_\ell \left(\rho\, U^2_{\mathrm{r,adj}}\right)^*. \tag{3.16}$$

Note that if the flow has reached a steady state in which $[q_{\ell,\mathrm{r}}]^{\mathrm{R}}_{\mathrm{L}} = [c]^{\mathrm{R}}_{\mathrm{L}} = 0$, then both $Q_{\ell,\mathrm{r}}$ and C are constants (see Section 3.2.3.) One has from (3.13) that $J'_{\mathrm{r,adj}} = J'_{\mathrm{r}}$ and (3.15) becomes identical to the steady discontinuous solutions (3.11), provided $S_{\mathrm{trans}} \to 0$.

A truly steady wave in which $S_{\mathrm{trans}} = 0$ and $\partial_t \boldsymbol{\Psi} = 0$ also implies that the roots of J'_{r} and S coincide at the critical point H_0, as explained in Section 3.2.1. In evolving waves, on the other hand, the roots of $J'_{\mathrm{r,adj}}$ and S will not coincide perfectly. The quite simple means of making the profile integrable is here to let the S_{trans} term, representing the transient residual, provide the degree of freedom needed for accomplishing the double root. S_{trans} is made a spatial constant within each wave, expressing the discrepancy between the two roots. Precisely, $S_{\mathrm{trans}} = S(H_0)$ where H_0 is the root of $J'_{\mathrm{r,adj}}$.

Johnson et. al. [10] chose the friction multiplier m_{x} such that the double root was achieved, rather than using an artificial source term. The present approach was chosen in order to make the method independent of the friction closure.

Note that if second order differential terms are to be included into the profile model, for instance turbulent shear terms as in [12], the profile may no longer be monotonic and the variable inversion $H(X) \to X(H)$ no longer possible.

3.2.3 Modelling 'in the small' – $Q_{\ell,\mathrm{r}}(H)$ and $C(H)$

In this section the choices for the modelled discharge rate $Q_{\ell,\mathrm{r}}$ and translation velocity C are discussed. These choices then directly govern the transient $\partial_t A_\ell$ in the strip $[t_n, t_{n+1}\rangle$ through the mass equation in (3.7a).

$Q_{\ell,\mathrm{r}}$ and C are made globally continuous and locally linear in their individual variable ξ, that is

$$\phi = \phi_{\mathrm{L}} + [\phi]^{\mathrm{R}}_{\mathrm{L}} \frac{\xi_\phi - \xi_{\phi,\mathrm{L}}}{[\xi_\phi]^{\mathrm{R}}_{\mathrm{L}}}, \qquad \phi \in \{Q_{\ell,\mathrm{r}}, C\}. \tag{3.17}$$

The ξ's are chosen as

$$\xi_{Q_{\ell,\mathrm{r}}} = A_\ell^{3/2}\left(\mathcal{A}'_\ell\right)^{-1/2}, \qquad \xi_C = 1/A_\ell.$$

The choice for $\xi_{Q_{\ell,\mathrm{r}}}$ is motivated by the consideration that changes in $Q_{\ell,\mathrm{r}}$ should be consistent with the characteristic speeds (3.10), simplified in assuming $\rho_\ell \gg \rho_\mathrm{g}$ (the characteristics of the shallow water equations.)

ξ_C is chosen considering the adjusted relative liquid convection velocity $U_{\ell,\mathrm{r,adj}} = \big(Q_{\ell,\mathrm{r}} - A_\ell \frac{\mathrm{d}Q_{\ell,\mathrm{r}}}{\mathrm{d}A_\ell} - A_\ell^2 \frac{\mathrm{d}C}{\mathrm{d}A_\ell}\big)/A_\ell$ active in $J'_{\mathrm{r,adj}}$, seeking to make the influence of coordinate stretching uniform.

These sub-models do affect the wave dynamics notably, and better alternatives are likely to exist.

3.3 Finding the Correct Profile Solution

Expression (3.15) gives the inverse wave profile as function of H. The wave profile should be consistent with the average wave cell properties, i.e., wavelength, mass and momentum from the integral equations, that is

$$\left(1, \overline{a_\ell}, [\rho\overline{u}]_\mathrm{g}^\ell\right) = \frac{1}{\Delta x}\int_{h_\mathrm{L}}^{h_\mathrm{R}} \left(1, A_\ell, [\rho\, U]_\mathrm{g}^\ell\right) \frac{J'_{\mathrm{r,adj}}}{S - S_{\mathrm{trans}}}\,\mathrm{d}H. \tag{3.19}$$

These three conditions, together with the shock condition (3.6), are sufficient to uniquely determine the shock states $\{h^+, h^-, q_\ell, c\}_{i+\frac{1}{2}}$. A multi-dimensional root search algorithm is employed for the job. [2]

The momentum flux $j_{\mathrm{r,L}}$ through the left face of the control volume is computed $j_{\mathrm{r,L}} = J_\mathrm{r}(H_\mathrm{L}, q_{\ell,\mathrm{L}}, c_\mathrm{L})$, and similar for the right face. J_r will then remain globally continuous since c was defined in (3.6) to maintain the momentum invariant across shocks. Likewise, $Q_{\ell,\mathrm{r}}$ is globally continuous from (3.17). Finally, the average source term is

$$\overline{s} = \overline{S} = \frac{1}{\Delta x}\int_{h_\mathrm{L}}^{h_\mathrm{R}} S\,\frac{J'_{\mathrm{r,adj}}}{S - S_{\mathrm{trans}}}\,\mathrm{d}H.$$

The profile itself is then given by

$$X(h) = X_\mathrm{L} + \int_{h_\mathrm{L}}^{h} \frac{J'_{\mathrm{r,adj}}}{S - S_{\mathrm{trans}}}\,\mathrm{d}H; \quad h_\mathrm{L} \le h \le h_\mathrm{R}.$$

Main conditions for a physical wave profile solution are

$$h_\mathrm{L} < H_0 < h_\mathrm{R} \tag{3.20a}$$
$$J'_{\mathrm{r,adj}} > 0, \qquad H > H_0 \tag{3.20b}$$
$$S - S_{\mathrm{trans}} > 0, \qquad H > H_0 \tag{3.20c}$$
$$J'_{\mathrm{r,adj}} < 0, \qquad H < H_0 \tag{3.20d}$$
$$S - S_{\mathrm{trans}} < 0, \qquad H < H_0 \tag{3.20e}$$
$$C < U_\ell, \qquad h_\mathrm{L} < H < h_\mathrm{R} \tag{3.20f}$$

[2] The CAW method was programmed in the MATLAB language; built-in functions `integrate` and `fsolve` of version R2013b were used to evaluate (3.19) and to search for the roots, respectively.

Further details on the conditional nature of S, J'_{r} and the shock relation (3.6) is provided in [11, 10]. Conditions (3.20a) and (3.20b) imply that h_{R} must be less than the upper root of $J'_{\mathrm{r,adj}}$, at which $\frac{\mathrm{d}H}{\mathrm{d}X} \to \infty$. Likewise, (3.20a) and (3.20e) implies that h_{L} must be greater than the lower root of $S - S_{\mathrm{trans}}$, at which $\frac{\mathrm{d}H}{\mathrm{d}X} \to 0$.

3.4 Simulation routine

A simultaneous and a sequential solver approach will presently be described. The main distinction is that the simultaneous procedure requires the solution of one $4 \times N$-dimensional profile problem every time step, N being the number of waves, while the sequential procedure requires the solution of N three-dimensional problems.

3.4.1 Simultaneous solver

Each wave in a wave train will provide a shock front (jump) and append four variables (h^+, h^-, $q_{\ell,\mathrm{r}}$ and c) to the system. The shock condition (3.6) couples the individual profiles such that the profile of one wave is mutually dependent on the profile of its neighbouring waves. The number of variables in the coupled system is four times the number of distinct shocks i.e., $\{h^+, h^-, q_{\ell,\mathrm{r}}, c\}_{i+\frac{1}{2}}$, $i = 1 \ldots N$.
Figure 3.2 provides an algorithmic illustration of the simulation procedure.

3.4.2 Sequential solver

If, on the other hand, the wave solutions are decoupled by instead using the wave velocities from the previous time step, the chained wave system will consist of separate three-dimensional single-wave problems with the variables $\{h_{\mathrm{L}}, h_{\mathrm{R}}, q_{\mathrm{r},\ell,\mathrm{L}}\}_i$. $Q^{(n)}_{\ell,\mathrm{r}}$ from (3.17) may be constructed using the old differences $[q_\ell]_{\mathrm{L}}^{\mathrm{R}^{(n-1)}}$ together with $q^{(n)}_{\mathrm{r},\ell,\mathrm{L}}$. An algorithmic illustration is given in Figure 3.3. The downside of a sequential procedure is that it is prone to numerical oscillations due to the lag in celerity information. Under-relaxation of c, performed in the manner $c^{(n)} := rc^{(n-1)} + (1-r)\, c^{(n)}$, r being the relaxation factor, is an effective way of stabilising the sequential routine. Measures like reducing the time step and iterating on the next time level also improve stability, though these measures also reduce the simulation efficiency.

3.5 Implementation and management

The CAW simulator programme is implemented using object oriented programming with wave and shock objects arranged sequentially in a linked list. All flow objects are also linked to the main pipe object holding properties like pipe inclination and diameter, gathered relative to the flow objects' positions during runtime.

The CAW method requires an initial set of wavelengths. This must either be supplied as user input or retrieved from direct simulations (Section 4.)
Wave mergers, or coalescences, are identified from the collision of shocks or the diminishment of the shock height. In such events, two wave objects must merge. Mergers involve compounding the conserved integral properties into a new single wave object which covers the same spatial interval.

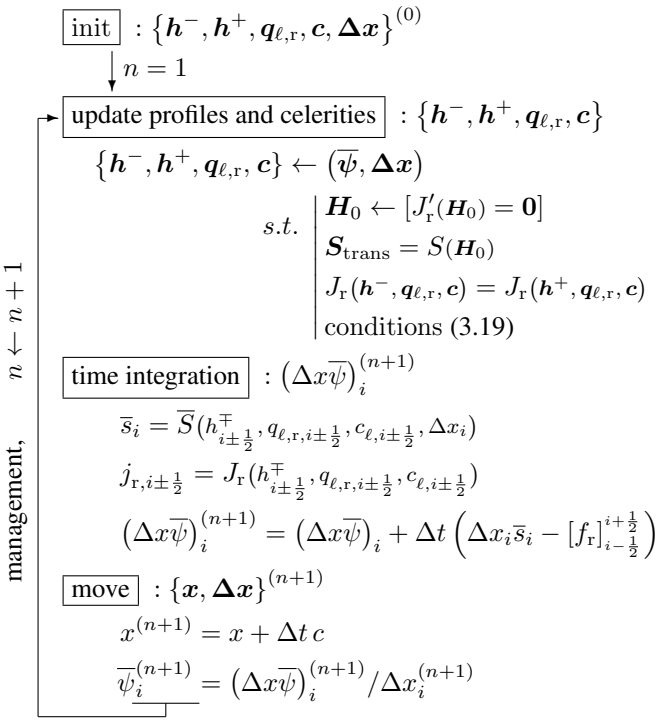


Figure 3.2: Simulation procedure. Profiles solved simultaneously. Time index (n) dropped. Vectors indicate collected shock data.

4 VALIDATION: WAVE CAPTURING THROUGH FINE DISCRETISATION

For validation, the target equations (2.2) are also solved through direct discretisation on a fixed uniform grid.

A number of the simpler common schemes were tested. Out of these, the simplest alternative, an explicit, non-staggered, first order upwind (or donor cell) scheme was preferred on the basis of simplicity. This is used for in all numerical comparisons with the number of grid cells equalling 3000 and the liquid-based CFL number equalling 0.5. For completeness, the discretised equations are presented in (4.1):

$$a_{\ell,i}^{(n+1)} = a_{\ell,i}^{(n)} - \frac{\Delta t}{\Delta x}\left(q_i^{(n)} - q_{i-1}^{(n)}\right), \tag{4.1a}$$

$$u_{\ell,i}^{(n+1)} = \left(\rho_\ell + \rho_{\mathrm{g}} \frac{a_{\ell,i}^{(n+1)}}{a_{\mathrm{g},i}^{(n+1)}}\right)^{-1} \times \left[\frac{\rho_{\mathrm{g}}\, \mathcal{Q}}{a_{\mathrm{g},i}^{(n+1)}} + \left[\rho\, u_i^{(n)}\right]_{\mathrm{g}}^{\ell} - \frac{\Delta t}{\Delta x}\left(j_i^{(n)} - j_{i-1}^{(n)}\right) + s_i^{(n)}\right], \tag{4.1b}$$

with q_ℓ and j given in (2.2c) and s in (2.2d). Simulations performed with (4.1) will simply be referred to as 'direct simulations' in the following discussion.

Although the direct simulations are finely resolved, one should keep in mind that they too are associated with inaccuracies from numerical diffusion and dispersion.

5 NUMERICAL EXPERIMENTS

This work contains a large number of parameters. In the interest of briefness the parameters listed in Table 1 will remain fixed for all numerical experiments. Fluid parameters correspond

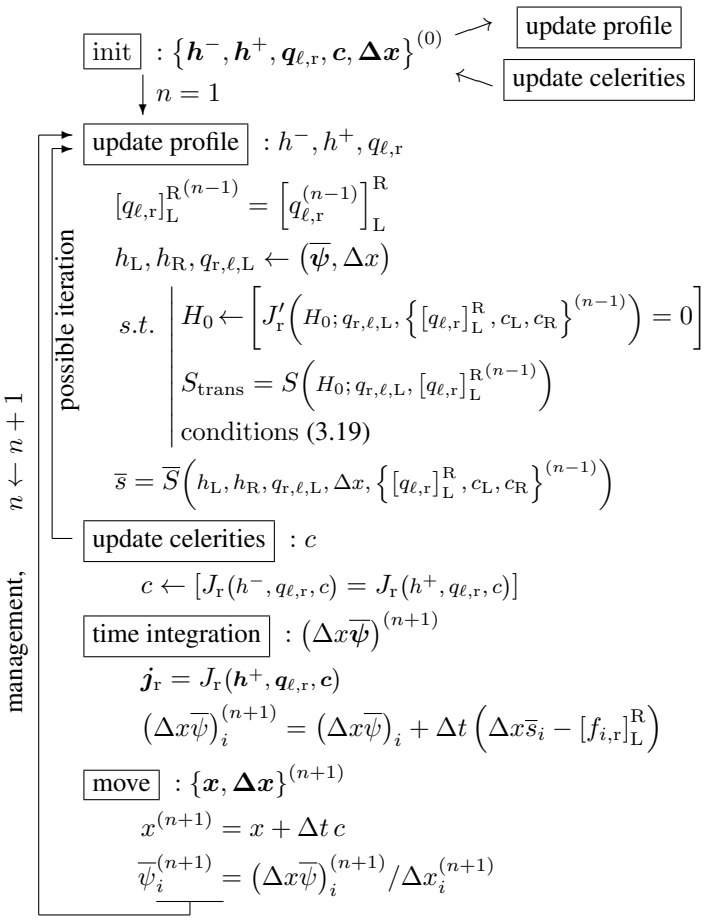


Figure 3.3: Simulation procedure. Profiles solved sequentially. Time index (n) dropped. Vectors indicate collected shock data.

to air-water flow in atmospheric conditions. Boundary conditions are made periodic (cyclic) for all test cases.

ρ_ℓ	998	$\mathrm{kg/m^3}$
ρ_g	1.205	$\mathrm{kg/m^3}$
ν_ℓ	1.005E-6	$\mathrm{m^2/s}$
ν_g	1.50E-5	$\mathrm{m^2/s}$
m_x	5.0	–
d	0.078	m
$q_\ell^{(0)}/\mathcal{A}$	0.25	$\mathrm{m/s}$

Table 1: Fixed parameters.

5.1 Simulations where the initial CAW profiles are used as the initial conditions for the direct simulation.

The first set of simulations are presented in Figure 5.1. Here, CAW simulations consist of two waves with initial length $\Delta x^{(0)} = 10\,d$ and $\Delta x^{(0)} = 20\,d$. A direct simulation runs alongside the CAW simulation in each case. Initial conditions in the direct simulations are taken as point values of the initial profile solution of the CAW. At $t > 0$ the two simulation methods run independent of each other. Initial superficial velocities of the gas phase are $q_\mathrm{g}^{(0)}/\mathcal{A} = 10.0\,\mathrm{m/s}$ in Subfigure (a) and (b), and $q_\mathrm{g}^{(0)}/\mathcal{A} = 11.0\,\mathrm{m/s}$ in Subfigure (c). The mixture rate remains unchanged from its initial value through all cases, that is $\mathcal{Q} \equiv q_\ell + q_\mathrm{g} = q_\ell^{(0)} + q_\mathrm{g}^{(0)}$.

Subfigure (a) shows the profile development in a one degree upwards inclined pipe. Wave profiles remain similar throughout the simulation, and wave coalescence occurs in close proximity in time and space. The predicted rate at which the shorter wave is eaten by the longer one is seen to differ more as the distance between wave fronts decreases. This feature may be caused by the diffusive properties of the simple direct simulation scheme, or it can be caused by an increase in the artificial source S_trans required to maintain the quasi-steady approximation in the CAW method. A third possible explanation for the slight discrepancy is that the sub-models of Section 3.2.3 are not optimal.

In Subfigure (b), the pipe inclination is one degree downwards. This manifests in higher waves and faster coalescence. (Waves pass the periodic boundaries many times in between snapshots.) The rate of coalescence is still greater in Subfigure (c), where the gas rates have been increased. The wave resulting from this wave merger is unstable, which from a modelling perspective can be interpreted at the formation of a momentary slug. In the direct simulation, this is seen as numerical instability; a sudden exponential and non-physical growth at the wave tip (also occurring in Figures 5.3,) leading to a simulation breakdown. In the CAW simulation, the instability is manifested in that no single monotone profile solution fitting the average properties exists.

Figure 5.2 presents the time evolution of the artificial source term S_trans for the two stable cases, Figure 5.1a and b. S_trans is largest initially and just before and just after a wave coalescence. The larger values of S_trans prior to a coalescence may contribute to the discrepancy in predicted celerities, as mentioned above. After coalescence, S_trans quickly diminishes to-

wards zero, meaning that the solution becomes identical to the steady wave solution presented in Section 3.2.1.

5.2 Simulations where the initial conditions of the CAW simulation are collected from a developing direct simulation.

A similar case set-up is presented in Figures 5.3 and 5.4 with a horizontal pipe. These CAW simulations are initiated with the average wave properties gathered from an intrinsically developing direct simulation. The direct simulation was initially uniform apart from a tiny, pointwise random disturbance. Instantaneous data was then stored at $t = 30\,\mathrm{s}$, before the growing waves began to break. Resetting the time to $t = 0$, CAW simulations are initiated by partitioning the direct simulation domain and averaging the properties within these partitions. The criterion for a wave partition is that it forms the midpoint between a local left peak and local right trough in liquid fraction. These wave are still developing, as can be seen from the initial profiles at $t = 0$. Figure 5.3 shows wave profile snapshots and Figure 5.4 displays the wavelengths development in time. Subfigures 5.3a and 5.4a present a simulation with sequentially computed CAW profiles (Section 3.4.2.) This simulation has been computed with a 90% under-relaxation when updating the wave celerities, keeping it numerically stable. The flow pattern evolves slowly and does not appear to be greatly affected by the considerable relaxation.

Subfigures 5.3b and 5.4b present this same test case when adopting the simultaneous solution procedure (Section 3.4.1.) Average states $\overline{\psi}$ are compared in Figure 5.5. The time and location of wave mergers are here in good agreement, keeping in mind the these simulations ran independently of each other before direct simulation waves started to break.

Finally, Figure 5.6 compares the two-wave development in a range of pipe inclinations, (a), for gravity driven flows, and, (b), for pressure driven flows.

5.3 Multiple waves in alternating pipe geometry

Finally, a larger simulation with multiple waves is performed in a snake-like pipe configuration. The pipeline geometry is shown in Table 2. $q_{\mathrm{g}}^{(0)}/\mathcal{A} = 9\,\mathrm{m/s}$. Periodic boundary conditions make this configuration continuous.



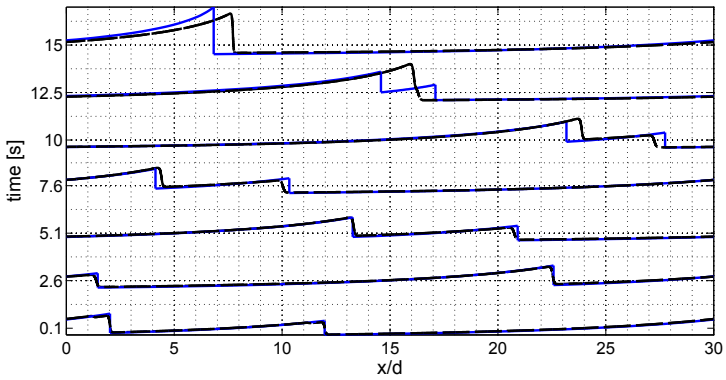
Pipe	1	2	3	4	5	6	7	8	9	10
θ	-.1°	-.3°	-.5°	-.3°	-.1°	.1°	.3°	.5°	.3°	.1°
L	10	10	20	10	10	10	10	20	10	10

Table 2: Pipeline geometry

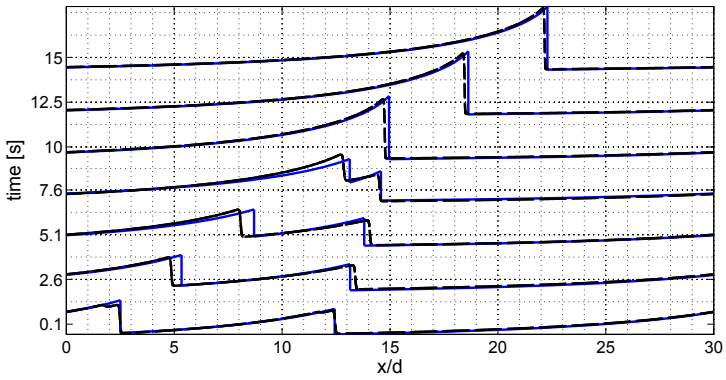
In order to avoid disturbing the profile search routine, changes in pipe inclination are accounted for by approximating a constant inclination in the profile equation. This inclination is computed from the fraction of wave found within each pipe segment. CAW profiles will thus react instantaneously to the changing inclination, inertia being supplied only though the average state. The sequential solver (Figure 3.3) is used with 95% under-relaxation.

Figure 5.7 presents the time evolution from a 12-wave simulation in a 120 diameter long domain. The two predictions remain fairly consistent throughout the simulation. Alternating pipe inclinations are manifested as a wavy evolution in the wavelength plot, Figure 5.8.

(a) $\theta = +1°$, $q_g^{(0)}/\mathcal{A} = 10\,\mathrm{m/s}$.



(b) $\theta = -1°$, $q_g^{(0)}/\mathcal{A} = 10\,\mathrm{m/s}$.



(c) $\theta = -1°$, $q_g^{(0)}/\mathcal{A} = 11\,\mathrm{m/s}$.

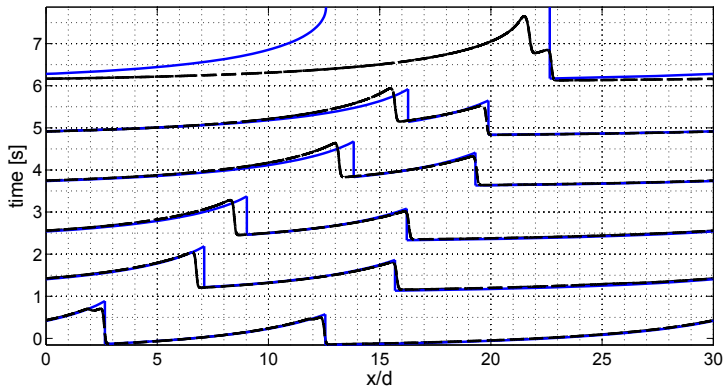


Figure 5.1: Double wave simulation using the simultaneous solution procedure. The initial state of the direct simulations is taken from the initial solutions of the analytical waves. Solid line: CAW, stippled line: direct simulation.
$\Delta x^{(0)} \in \{10\,d, 20\,d\}$. Wave heights H ranges from around $0.1\,d$ to $0.3\,d$ before coalescence. (Waves pass through boundaries many times between each profile visualisation.)

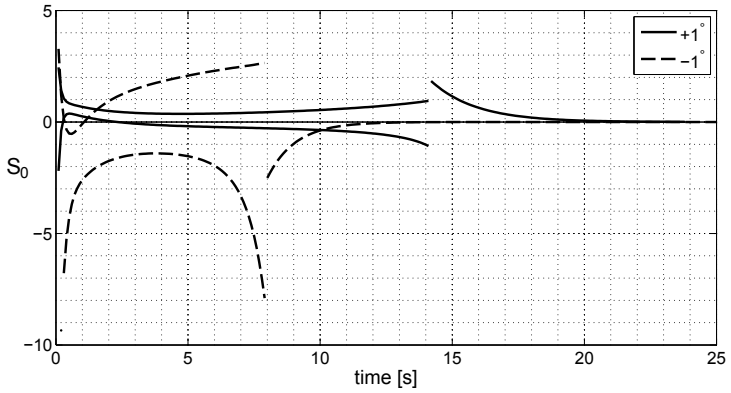
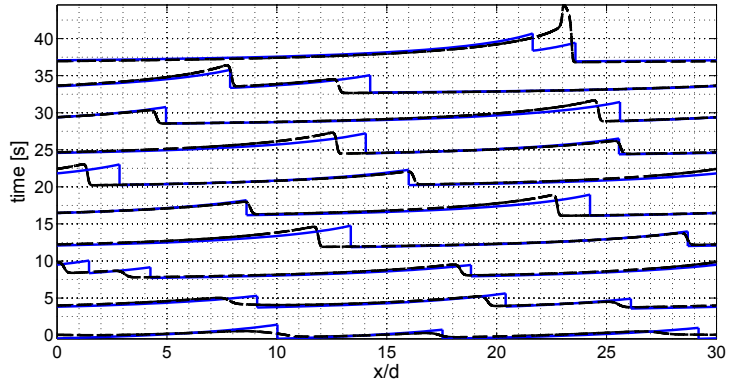
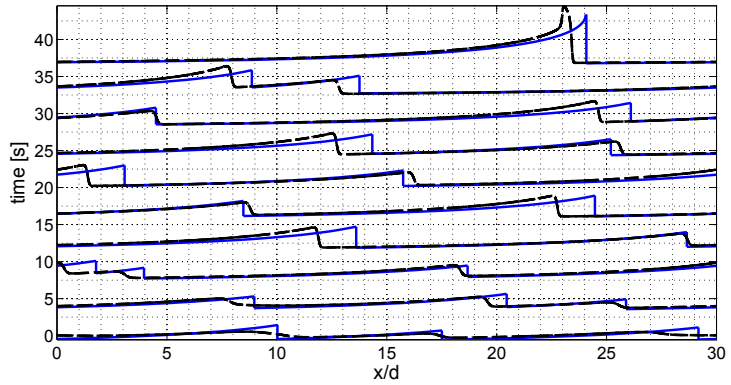


Figure 5.2: Artificial source S_{trans}. $q_{\text{g}}^{(0)}/\mathcal{A} = 10.0\,\text{m/s}$, $\theta = \pm 1°$, cf. Figure 5.1a and b.

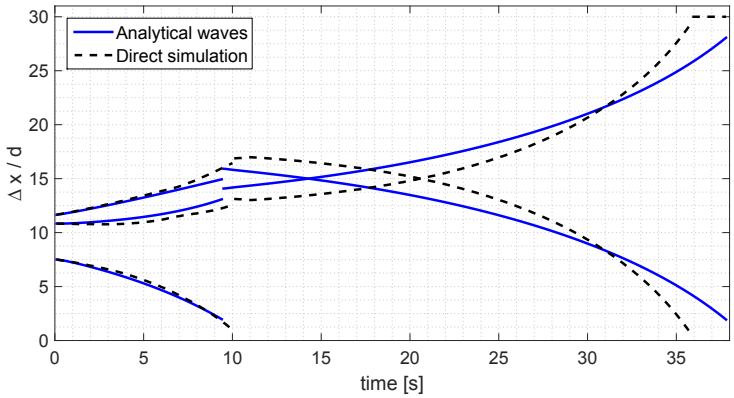


(a) Sequential procedure. 90% relaxation.

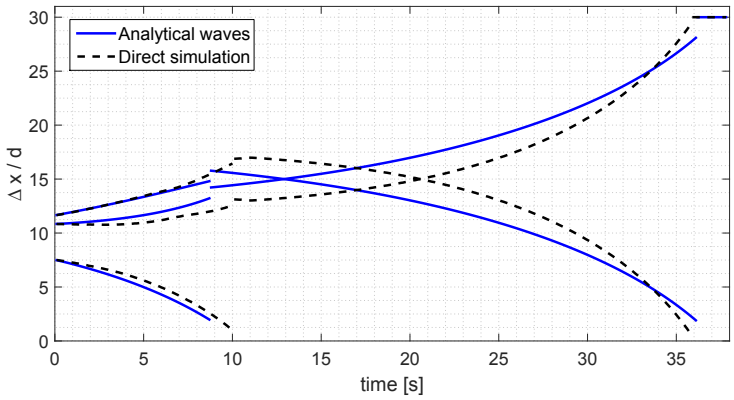


(b) Simultaneous procedure. (No relaxation.)

Figure 5.3: Profile development in time. The initial average states of the analytical wave simulation is computed by averaging partitions of a direct simulation at $t = 30\,\text{s}$. Solid line: CAW, stippled line: direct simulation.
$q_{\text{g}}^{(0)}/\mathcal{A} = 11\,\text{m/s}$, $\theta = 0°$.



(a) Sequential procedure. $Q_{\ell,\mathrm{r}}(H)$ and $C(H)$, as per Section 3.2.3. 90% relaxation.



(b) Simultaneous procedure. $Q_{\ell,\mathrm{r}}(H)$ and $C(H)$, as per Section 3.2.3. (No relaxation.)

Figure 5.4: Wavelength development in time, presented as $\Delta x/d$. Cf. Figure 5.3.

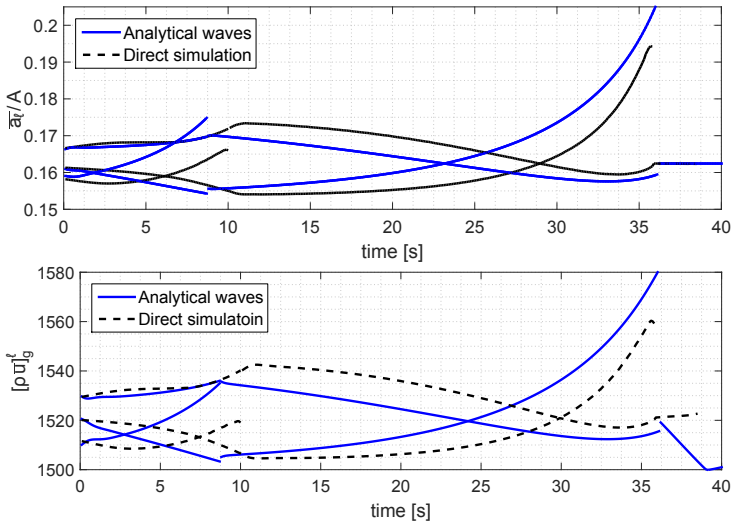
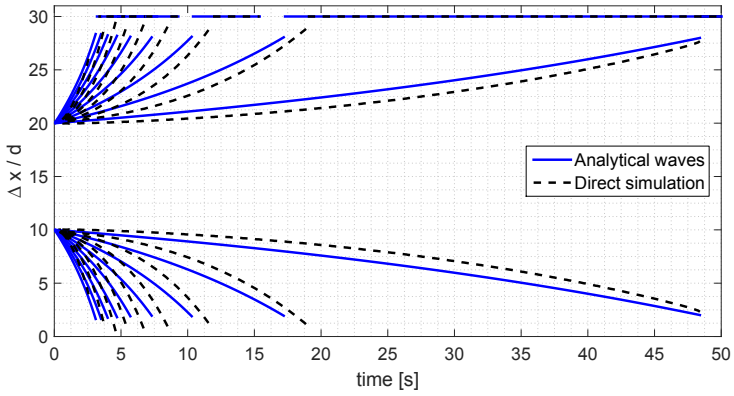
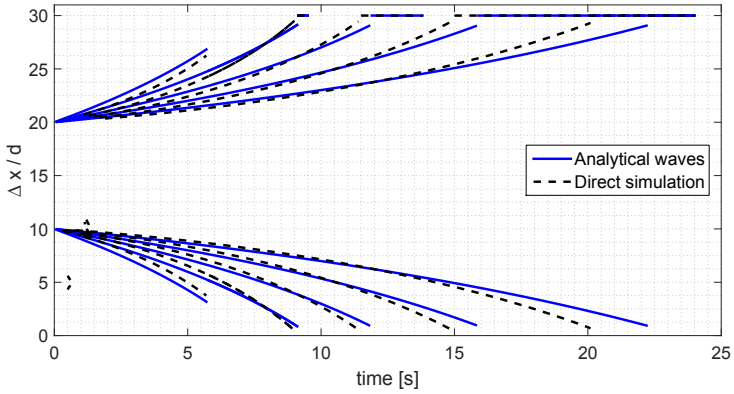


Figure 5.5: $\overline{a_\ell}/A$ and $[\rho\,\overline{u}]_{\mathrm{g}}^{\ell}$. Simultaneous procedure, cf. Figure 5.3b



(a) No gas present. $\theta \in \{-1.5°, -2.0°, \ldots, -5.5°\}$ where the lengths of the curbed lines on the 'fan' decreases with increasing inclination.



(b) $\theta = 0°$. $q_g^{(0)}/\mathcal{A} \in \{8\,\mathrm{m/s}, 9\,\mathrm{m/s}, \ldots, 12\,\mathrm{m/s}\}$ where the lengths of the curbed lines on the 'fan' decreases with increasing superficial gas velocity.

Figure 5.6: Wavelength plot. Simultaneous procedure.

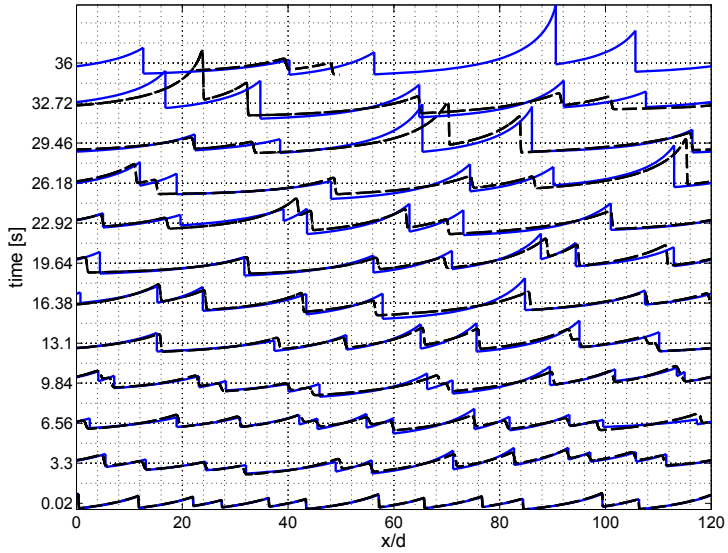


Figure 5.7: Multiple waves in the pipeline configuration of Table 2; profile evolution. Solid line: CAW, stippled line: direct simulation.

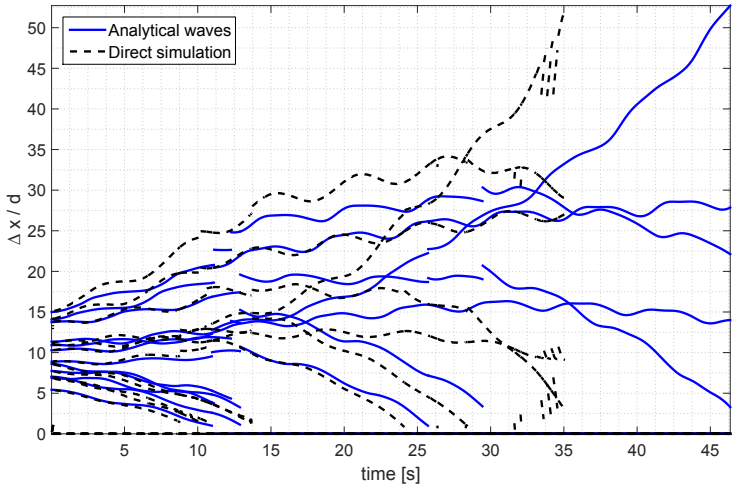


Figure 5.8: Multiple waves in the pipeline configuration of Table 2, cf. Figure 5.7; wavelength evolution.

The higher waves coalesce quicker in the capturing simulation than in the CAW during the later stages of the simulation, similar to what was observed in the two-wave simulations. Coalescence of the two highest waves takes place around $t = 35\,\mathrm{s}$ in the direct simulation, resulting in an unstable wave and the subsequent simulation break-down. This happens about ten seconds later in the CAW simulation.

Average variables $\overline{\psi}$ are again plotted against time in Figure 5.9. Oscillations of these average variables are quite consistent, despite the significant under-relaxation of the sequential CAW solver.

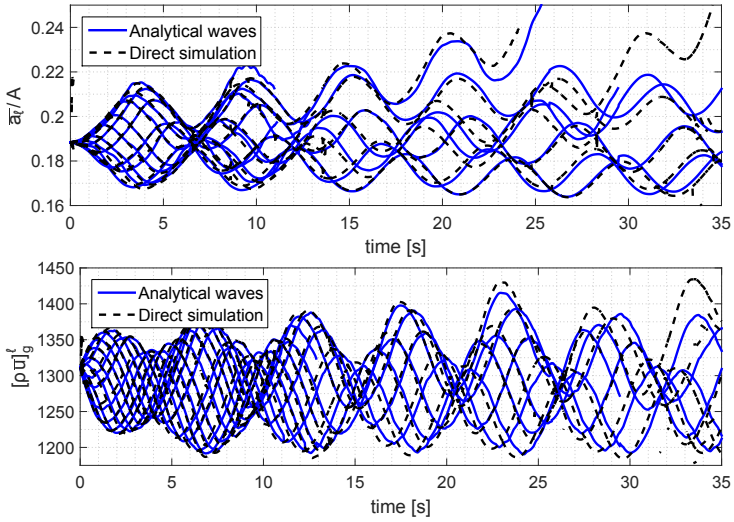


Figure 5.9: Average wave properties $\overline{\psi} = \langle \overline{a_\ell}, [\rho\,\overline{u}]_g^\ell \rangle$ cf. Table 2 and Figure 5.7

6 POTENTIAL AND EFFICIENCY

Tests presented in the previous section were of a basic nature, evaluating the main principle that analytical wave solutions would evolve appropriately if put into a dynamic frame. Obviously, the applicability such principles is tightly hinged on the degree to which defined structures (waves) are physically present.

Regarding efficiency, the bottleneck of the CAW method, as presented here, is the profile reconstruction of Section 3.3, implemented with a multi-dimensional root search algorithm. The main computational cost within this profile search routine is in turn the numerical profile integrations for retrieving average wave properties. Such costly routines can be acceptable only if that cost can be regained through significantly fewer cells and larger time steps. Simpler models, or profile approximations, which provide algebraically pre-integrable expressions would alleviate the computational cost significantly.

7 CONCLUSIONS

A simulation concept which takes advantage of naturally occurring flow structures has been presented and tested using a method of dynamically chained analytical waves. Results compare favourably with fully resolved direct simulations of the same state equation set.

In particular, the tested method seems capable of predicting wave speeds, wave evolution and wave coalescences with reasonable accuracy in time and space. The method shows the same

stability behaviour as in the fully resolved simulations, and appears to be well adopted for the range of flow parameters within which the roll-wave phenomenon is set to occur. Prediction accuracy is likely to improve with further method development.

Predictions are made using vastly larger grid cells and time steps than what is required with conventional finite-volume methods. Still, better ways of searching for profile solutions seems to be necessary for the method to become truly efficient. The method also requires initial wavelengths as input. However, once some initial wavelengths have been given, or extracted from a direct simulation, the wave regime develops either towards an appropriate steady state or a flow regime transition.

Acknowledgements

This work is financed by The Norwegian University of Science and Technology (NTNU) as a contribution to the Multiphase Flow Assurance programme (FACE.) The author thanks Kontorbamse for the inspirational support.

REFERENCES

[1] D. Barnea and Y. Taitel. Interfacial and structural stability of separated flow. *International Journal of Multiphase Flow*, 20(supplissue):387 – 414, 1994/08/.

[2] D. Biberg. A mathematical model for two-phase stratified turbulent duct flow. *Multiphase Science and Technology*, 19(1):1 – 48, 2007.

[3] R.R. Brock. Development of roll-wave trains in open channels. *American Society of Civil Engineers, Journal of the Hydraulics Division*, 95(HY4):1401 – 1427, 1969.

[4] B.S. Brook, S.A.E.G. Falle, and T.J. Pedley. Numerical solutions for unsteady gravity-driven flows in collapsible tubes: evolution and roll-wave instability of a steady state. *Journal of Fluid Mechanics*, 396:223 – 56, 1999/10/10.

[5] Zhixian Cao, Penghui Hu, Kaiheng Hu, Gareth Pender, and Qingquan Liu. Modelling roll waves with shallow water equations and turbulent closure. *Journal of Hydraulic Research*, 0(0):1–17, 0.

[6] A. De Leebeeck and O.J. Nydal. Simulation of large amplitude waves in a slug tracking scheme compared to roll wave experiments at high pressure. *International Journal of Multiphase Flow*, 36(1):40 – 50, 2010.

[7] R.F. Dressler. Mathematical solution of the problem of rollwaves in inclined open channels. *Communications on Pure and Applied Mathematics*, 2:149 – 194, 1949/06/.

[8] D. Gidaspow. Multiphase flow and fluidization. *Journal of Fluid Mechanics*, 287:405 – 405, 1995.

[9] H. Holmås. Numerical simulation of transient roll-waves in two-phase pipe flow. *Chemical Engineering Science*, 65(5):1811 – 25, 2010/03/01.

[10] George W. Johnson, Arnold F. Bertelsen, and Jan Nossen. A mechanistic model for roll waves for two-phase pipe flow. *AIChE Journal*, 55(11):2788 – 2795, 2009.

[11] G.W. Johnson. *A Stuy of Stratified Gas-Liquid Pipe Flow*. PhD thesis, Univ. Oslo, 2005. dr. scient.

[12] G.W. Johnson, J. Nossen, and A.F. Bertelsen. A comparison between experimental and continuous theoretical roll waves in horizontal and slightly inclined pipes at high pressure. pages 643 – 655, Barcelona, Spain, 2005.

[13] Jr. Lahey, R.T. Void wave propagation phenomena in two-phase flow (kern award lecture). *AIChE Journal*, 37(1):123 – 35, 1991/01/.

[14] M. Miya, D.E. Woodmansee, and T.J. Hanratty. A model for roll waves in gas-liquid flow. *Chemical Engineering Science*, 26(11):1915 – 31, 1971/11/. liquid film;high speed gas flow;flow surges;roll waves;height;wall shear stress;gas pressure;mathematical model;.

[15] G.L. Richard and S.L. Gavrilyuk. A new model of roll waves: Comparison with brock's experiments. *Journal of Fluid Mechanics*, 698:374 – 405, 2012.

[16] Y. Taitel and A.E. Dukler. A model for predicting flow regime transitions in horizontal and near horizontal gas-liquid flow. *AIChE Journal*, 22(1):47 – 55, 1976/01/.

[17] H. A. Thomas. The propagation of waves in steep prismatic conduits. In *Hydraulics Conf.*, pages 214–229, Carnegie Institute of Technology, Pittsburgh, 1939.

[18] M. Watson. Wavy stratified flow and the transition to slug flow. In C.P. Fairhurst, editor, *Multi-phase Flow – Proceedings of the 4th International Conference*, pages 495–512, Cranfield, UK, 1989.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MODELING OF EXCITATION OF CONTROLLING DISTURBANCES IN SWEPT WING BOUNDARY LAYER BY MEANS OF PLASMA ACTUATORS

Sergey L. Chernyshev [1], **Andrey Ph. Kiselev** [1], **and Aleksandr P. Kuryachii** [1]

[1] Central Aerohydrodynamic Institute (TsAGI)
1, Zhukovsky str., Zhukovsky, 140180 Moscow region, Russia
e-mail: aleksandr.kuryachiy@tsagi.ru

Keywords: Swept Wing, Cross-Flow-Type Instability, Plasma Actuator.

Abstract. *Two variants of multiple dielectric barrier discharge (DBD) actuator are proposed for generation the variable amplitude forced modes in thin boundary layer near a swept wing leading edge with the aim to cancel the critical cross-flow-type instability modes resulting in laminar-turbulent transition. Numerical simulation of the excitation of stationary forced modes by DBD-actuators in compressible 3D boundary layer on an infinite span swept wing at typical cruise flight conditions has been executed to estimate the possibility of the proposed method of transition control. Spanwise periodic localized volumetric force and heat impact on boundary layer flow at realistic physical parameters of DBD has been considered. It is shown that DBD-actuators generate three rows of pairs of counter-rotating vortices one above another along local inviscid streamline.*

1 INTRODUCTION

Development of innovative aerodynamic technologies for a more effective, environmentally friendly air transport system remains in the centre of attention of the aeronautical community. Cruise drag reduction is one of the main goals of these developments [1]. The viscous drag contributes about a half of the total cruise drag of modern airplanes. A delaying laminar-to-turbulent transition in boundary layer on aerodynamic surfaces is one of the effective methods of viscous drag reduction. Laminar-to-turbulent transition on a swept wing of typical subsonic airplane in cruise flight is caused, as a rule, by cross-flow-type instability modes, which evolve as a pairs of stationary counter-rotating vortices, whose axes are closely aligned with local inviscid streamlines [2]. Therefore any method of suppression of these instability modes would be a key to solution the problem of a swept wing drag reduction, if this method is constructive reliable and energy acceptable.

Saric et al. [3] have shown that the linear predicted most unstable mode leading to transition (termed the critical mode) could be suppressed, thereby delaying transition, by distributing spanwise-periodic discrete surface roughness elements (DREs) near the wing leading edge at spacing approximately equal to one-half of the spanwise wavelength of the critical mode. The micron-high roughness elements produce a rapid growth of the forced controlling mode and then its decay. This forced mode suppresses the critical mode, thus delaying transition beyond the natural location that would occur in the absence of any artificial roughness.

The three principal types of DREs include: appliqué, pneumatic, and dielectric barrier discharge (DBD) actuators [4]. Note, that DREs have to be placed near the neutral stability point rather upstream. Both Reynolds number and the angle of attack can change during cruise flight. Different Reynolds number produce different wavelengths of the critical mode and different angles of attack produce different positions of neutral stability point. Therefore some wavelength and chord-position variability is required in any DREs system. It is evident that appliqué DREs don't satisfy this requirement. Pneumatically driven actuator proposed in [5] is capable of producing variable height of surface roughness thereby ensuring variable amplitude of the exited controlling mode. Different amplitude can be fitted to different neutral stability point position. Main disadvantages of the pneumatic variable roughness actuator [5] consist in 1) difficulty to produce controlling mode with very short wavelength and 2) high probability of destruction of thin flexible tape covering orifices by solid or ice particles.

Dielectric barrier discharge actuators with streamwise oriented external electrodes have been used for generation of near-surface streamwise vortices first in [6] and then have been investigated in details in [7, 8]. Sandwich DBD-actuator with an additional control electrode has been considered in [9]. But long external electrodes placed parallel to the free stream direction are necessary to generate strong enough streamwise vortices in these configurations of DBD actuators. Streamwise vortices can be generated also in DBD actuators with serrated [10] or serpentine [11] external electrodes which can have small extension in streamwise direction. But these configurations seem to be too power-consuming because the discharge takes place along the whole external electrode in spanwise direction.

2 MULTIPLE DIELECTRIC BARRIER DISCHARGE ACTUATORS

Simple design of multi-DBD actuator recently proposed at TsAGI permits to generate the variable forced modes in thin boundary layer near a swept wing leading edge. The mode intensity can be adjusted to varying position of neutral stability point by changing voltage or frequency of DBD. Unfortunately this design does not permit to vary the spanwise wavelength of the generated modes. Main advantages of the proposed DBD actuator consist

in low power consumption and the possibility of significant miniaturization for generation short-wave modes.

Two variants of multi-DBD actuator are presented on Fig. 1. Here *1* is the wing skin which must be electricity-conductive and equipotential in accordance with electrostatic safety requirements, *2* is the dielectric insert, *3* is the buried accelerating electrode, *4* is the exposed electrode in the form of a comb, *5* is the high voltage power supply, *6* are the regions of DBD. It is supposed that the external electrode *4* is flush mounted and the tips of the comb teeth are placed at some distance upstream from the position of neutral stability point. The spanwise period of the teeth λ is equal to the wavelength of the exited controlling mode. The exposed electrode *4* is electrically linked with the wing skin *1* and one output of the power supply *5* is connected with the skin. Other output of the power supply is connected with the buried electrode *3*. One edge of the buried electrode *3* is arranged directly beneath the tips of the comb teeth *4*. The other edge of the buried electrode is distant enough from the skin to avoid the discharge ignition near the skin border.

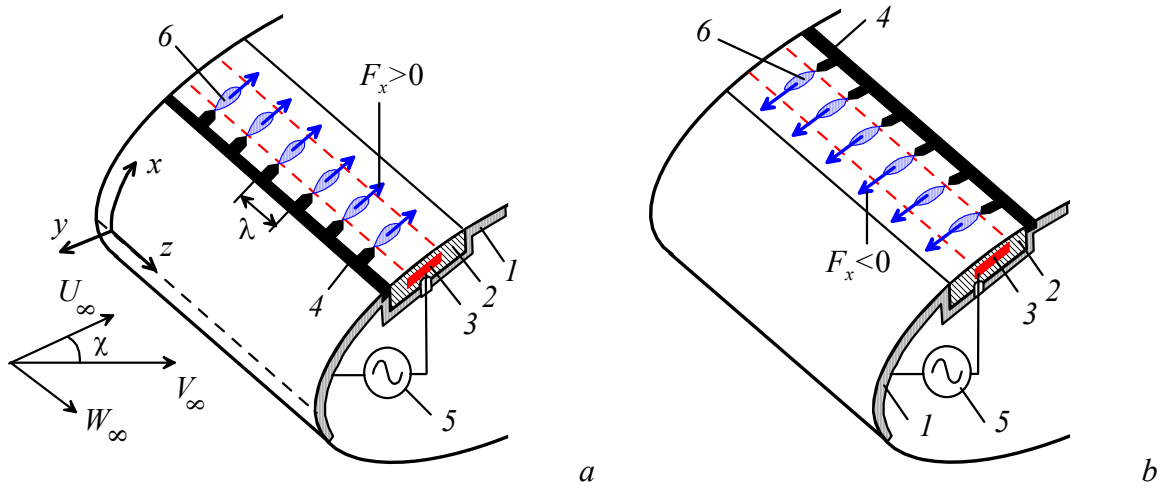


Figure 1: Design of multi-DBD actuators.

The time averaged volumetric force arising in the discharge regions *6* has two components. The main component F_x is directed along a wing surface perpendicular to a wing leading edge, i. e. along the *x*-axis, and the other component is directed perpendicular to a wing surface. The spanwise component of the generated force can be neglected. The *x*-component of the volumetric force is directed downstream ($F_x > 0$) in the variant 1 shown on Fig. 1, *a* and upstream ($F_x < 0$) in the variant 2 shown on Fig. 1, *b*. Volumetric force impact directed perpendicular to the leading edge in both variants results in a vorticity generation both in *x*- and *z*-direction and, as a result, a formation of vortex system aligned with external streamline.

3 CHARACTERISTICS OF CROSS-FLOW-TYPE INSTABILITY

The possibility of multi-DBD actuator to generate the streamwise vortices is estimated on the example of a subsonic flow over infinite span swept wing with sweep angle of $\chi = 30^o$. The static pressure $p_\infty = 2.6 \cdot 10^4$ Pa, the air temperature $T_\infty = 223$ K corresponding to the flight altitude of 10 km, and Mach number $M_\infty = 0.8$ are taken as the main free-stream parameters. They determine other flow parameters necessary for inviscid and boundary layer calculations: the flow velocity $V_\infty = 240$ m/s, the air density $\rho_\infty = 0.41$ kg/m^3, the dynamic viscosity coefficient $\mu_\infty = 1.33 \cdot 10^{-5}$ kg/(m·s).

The external boundary conditions for calculations of the compressible boundary layer have been obtained from a calculation of 2D inviscid flow over experimental LV6 DLR airfoil at zero angle of attack on the base of the Euler equations. The boundary layer flow in the vicinity of a wing leading edge is characterized by Reynolds number $\mathrm{Re}_l = \rho_\infty V_\infty l/\mu_\infty$ determined by streamwise length $l = V_\infty/[du_e(0)/dx]$ which is equal approximately to a half radius of the leading edge curvature, where u_e is the x-component of the external velocity obtained from inviscid calculation. The airfoil chord length L normal to a leading edge is related with the characteristic length l as $L = l\, du_e^{/}(0)/dx^{/}$, where the dimensionless velocity $u_e^{/}$ and the coordinate $x^{/}$ are measured in $V_\infty\cos\chi$ and L, respectively. According to executed 2D inviscid flow calculation, $du_e^{/}(0)/dx^{/} = 107.7$. The value $l = 0.03$ m was taken in the present estimations, hence, the airfoil chord length equals $L = 3.23$ m. The specified above free-stream parameters and characteristic length l determine the value of the Reynolds number $\mathrm{Re}_l = 2.03\cdot10^5$ and the estimation for the boundary layer thickness on the critical line of the wing leading edge $\delta_0 \approx 5\mathrm{Re}_l^{-1/2} l \approx 0.33$ mm.

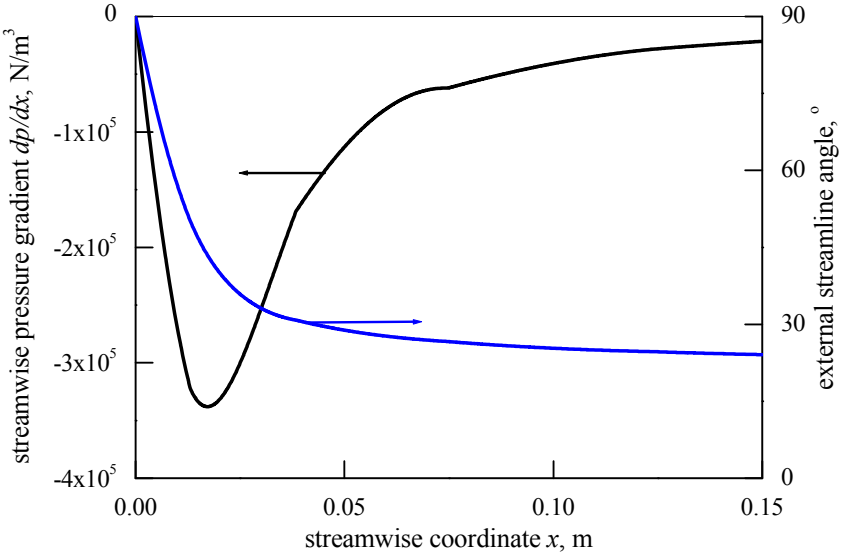


Figure 2: Streamwise distributions of the static pressure gradient and the angle of the inviscid streamline.

The inviscid flow over LV6 airfoil is characterized by very short part (less than 5% of the chord length) of strong flow acceleration as it is seen on Fig. 2. Here the static pressure gradient along a wing surface perpendicular to the leading edge and the angle ψ between the x-axis and the external inviscid streamline are shown.

The compressible boundary layer flow on a swept wing is described by the velocity components u, v, w, the static pressure p, the air density ρ, the static enthalpy $h = c_p T$ which are governed by the following system of equations and boundary conditions in Cartesian coordinate system shown on Fig. 1, *a*:

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0$$

$$\rho u\frac{\partial u}{\partial x} + \rho v\frac{\partial u}{\partial y} + \rho w\frac{\partial u}{\partial z} = \frac{\partial}{\partial y}\left(\mu\frac{\partial u}{\partial y}\right) - \frac{dp}{dx} + F_x$$

$$\rho u\frac{\partial w}{\partial x} + \rho v\frac{\partial w}{\partial y} + \rho w\frac{\partial w}{\partial z} = \frac{\partial}{\partial y}\left(\mu\frac{\partial w}{\partial y}\right)$$

$$\rho u\frac{\partial h}{\partial x}+\rho v\frac{\partial h}{\partial y}+\rho w\frac{\partial h}{\partial z}=\frac{\partial}{\partial y}\left(\frac{\mu}{\mathrm{Pr}}\frac{\partial h}{\partial y}\right)+u\frac{dp}{dx}+\mu\left[\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial w}{\partial y}\right)^2\right]+Q$$

$$\rho=\frac{\gamma}{\gamma-1}\frac{p}{h} \tag{1}$$

$$y=0:\ u=v=w=\partial h/\partial y=0;\quad y=y_e:\ u=u_e,\ w=w_e,\ h=h_e$$

$$x=0:\quad u=xU(y),\quad \rho U+\frac{d(\rho v)}{dy}=0,\quad \frac{d}{dy}\left(\mu\frac{dU}{dy}\right)-\rho v\frac{dU}{dy}-\rho U^2=0$$

$$\frac{d}{dy}\left(\mu\frac{dw}{dy}\right)-\rho v\frac{dw}{dy}=0,\quad \frac{d}{dy}\left(\frac{\mu}{\mathrm{Pr}}\frac{dh}{dy}\right)-\rho v\frac{dh}{dy}+\mu\left(\frac{dw}{dy}\right)^2=0$$

$$w_e=V_\infty\sin\chi,\quad h_e=h_\infty+0.5\left(V_\infty^2-u_e^2-w_e^2\right),\quad p_e=p_\infty\left(h_e/h_\infty\right)^{\gamma/(\gamma-1)}$$

$$\mu(T)=1.47\cdot10^{-6}\frac{T^{3/2}}{T+114},\quad \gamma=1.4,\ \mathrm{Pr}=0.72$$

Here F_x and Q are the spatial distributions of the time averaged streamwise volumetric force and Joule dissipation generated in discharge near every tooth of actuator and determined in analytical form below. The subscript e denotes the values of functions on the external boundary y_e of the boundary layer. The wing surface is supposed to be heat-insulated.

The gas velocity vector at some point inside a boundary layer can be represented by u- and w-components or by main-flow V_{MF} and cross-flow V_{CF} components related by the following expressions:

$$V_{MF}=u\cos\psi+w\sin\psi,\quad V_{CF}=u\sin\psi-w\cos\psi,\quad \mathrm{tg}\,\psi=w_e/u_e,\quad w_e=V_\infty\sin\chi \tag{2}$$

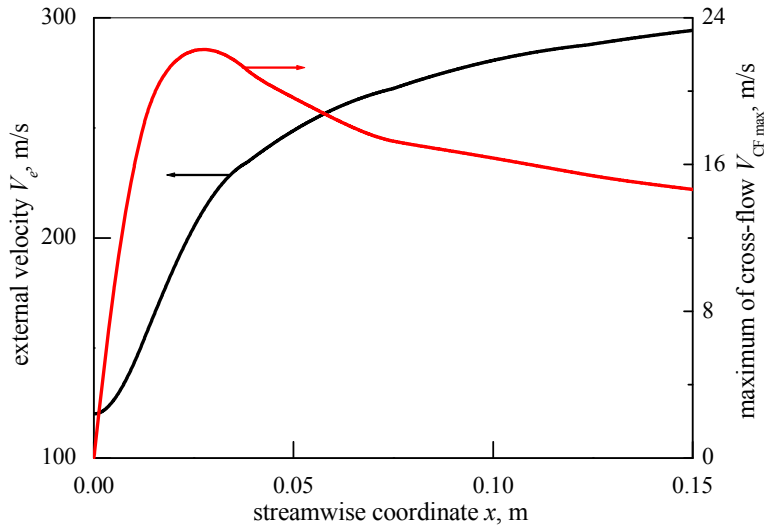


Figure 3: Streamwise distributions of the maximal main-flow and cross-flow velocities.

The streamwise distributions of the maximal main-flow velocity V_e achieving on the outer border of the boundary layer and the maximal cross-flow velocity $V_{CF\max}$ calculated without volumetric force and heat sources are presented on Fig. 3. It is seen that strong flow

acceleration results in very intensive cross-flow with maximal value of the cross-flow velocity achieving about 10% of the free stream velocity.

The stability characteristics of the considered boundary layer flow have been calculated in [12] in the framework of the linear stability equation system of Dunn-Lin [13]. Only steady cross-flow-type disturbances were considered. The well-known e^N-method was used to estimate the position of laminar-turbulent transition [2]. So the disturbances of all flow functions q and the N factor were defined as follows:

$$q = q^*(y)\exp(-\alpha_i x)\exp[i(\alpha_r x + \beta z)], \quad N(x) = -\int_{x_0}^{x} \alpha_i dx \tag{3}$$

Here q^* is the complex eigenfunction, $\alpha = \alpha_r + i\alpha_i$ is the complex eigenvalue, α_r and β represent the wave number components in x- and z-directions, α_i represents increment of spatial growth ($\alpha_i < 0$) or decrement of decay ($\alpha_i > 0$) of disturbances, x_0 is the initial coordinate where α_i obtains negative value, i.e. the disturbance becomes unstable. The so-called fixed β strategy [2] was used for N factor computation. That is the streamwise distributions of eigenvalue α and N factor are calculated according to (5) for a set of fixed spanwise wave numbers β.

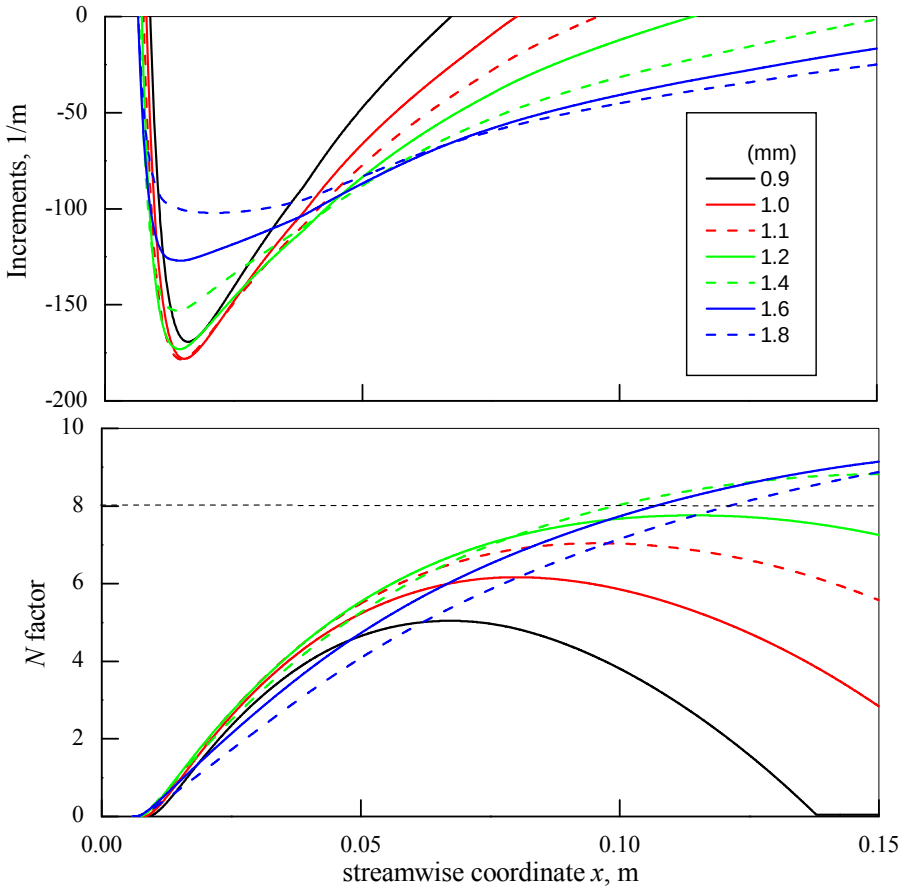


Figure 4: Streamwise distributions of increments of spatial growth and N factors for different spanwise wavelengths.

The mentioned above strong cross-flow in boundary layer results in a high flow instability. The calculated streamwise distributions of the increments of spatial growth of steady disturbances and N factors for different spanwise wavelength $\lambda = 2\pi/\beta$ in the range of 0.9-1.8

mm are shown on Fig. 4. The e^N-method implies that laminar-turbulent transition occurs when N factor for any spanwise wave number β (or corresponding wavelength λ) reaches some predefined value N_T. The position of the transition induced by steady cross-flow-type instability is estimated by $N_T = 8\text{-}10$ for the fixed β strategy [2]. According to lower Fig. 4, the disturbances with spanwise wavelengths $\lambda < 1.2$ mm are nonhazardous. The transition can be induced by disturbances with wavelengths $\lambda > 1.4$ mm at the distance $x_T \geq 0.1$ m from a critical line on the leading edge.

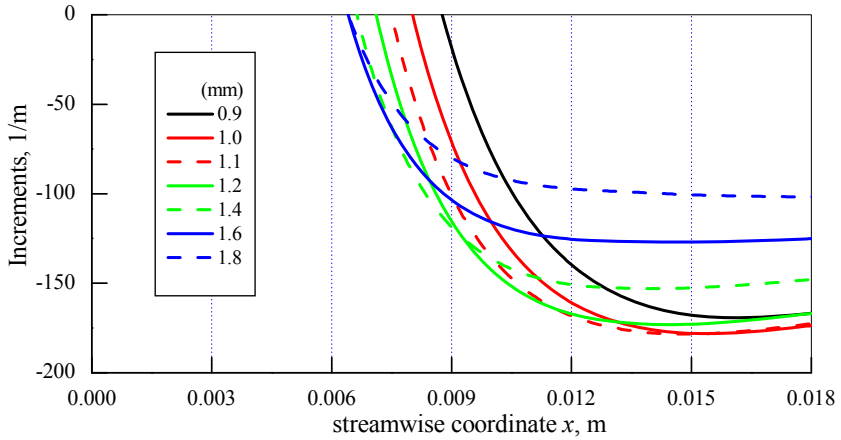


Figure 5: Distributions of increments of spatial growth in the vicinity of neutral point.

The disturbance with wavelengths $\lambda = 0.9$ mm seems to be an appropriate candidate for controlling forced mode to suppress the dangerous modes. This mode is amplified up to $x = 0.067$ m where the increments of dangerous modes attenuate appreciably, according to Fig. 4. The peculiarity of the considered particular case consists in the fact that short wave unstable modes arise downstream with respect to longer wave modes, as it is seen on Fig. 5 where the increments of the considered modes in the vicinity of neutral stability point are shown.

4 VORTICITY GENERATION BY IMPACT OF MILTI-DBD ACTUATOR

The impact of multi-DBD actuator on boundary layer flow is estimated in the framework of the system of equations and boundary conditions (1) for span-periodic flow. Numerical solution of the system (1) has been executed with the use of the expansion of all dependent variables and volumetric sources defined below in finite Fourier series and subsequent solution of the resulting 2D equation systems.

The spatial distributions of volumetric force and heat sources are taken in analytic form. The distribution of the force source in x-y plane for both positive F_+ and negative F_- force is described by the simple formula proposed in [14] and giving satisfactorily agreement with experimental data concerning the flow velocity induced by DBD actuators. The spanwise distribution is taken as a half of sine function, so the final expressions take the following form:

$$F_{\pm}(x,y,z) = F_t \frac{\pi}{2} \frac{x^{/} y^{/}}{x_0 y_0 z_0} \exp\left(\mp x^{/} - y^{/}\right) \sin\left(\pi z^{/}\right)$$

$$x^{/} = \left(x - x_{in}\right)/x_0, \quad y^{/} = y/y_0, \quad z^{/} = z/z_0, \quad 0 \leq y \leq y_e, \quad 0 \leq z \leq z_0 \qquad (4)$$

$$x_{in} \leq x \leq \infty \text{ for } F_+, \quad 0 \leq x \leq x_{in} \text{ for } F_-$$

Here F_t is the total source intensity in N, x_0 and y_0 are the coordinates of the distribution maximum (for F_+) or minimum (for F_-) in x-y plane, z_0 is the source size along a span, x_{in} is the coordinate of the tooth tip of the exposed electrode. It is supposed that the teeth tips are placed at same coordinate x_{in}= 6 mm, i.e. immediately in front of the neutral stability point, taking into account Fig. 5, in both variants of actuator shown on Fig. 1. The discharge regions propagate from x_{in} downstream in the variant 1 (F_+) or upstream in the variant 2 (F_-).

The spanwise period of multi-DBD actuator is equal to λ = 9 mm and the size of the force source along a span equals z_0= 4.5 mm. The following considerations have been taken into account at a choice of the other parameters of the distributions (4). The relation of the characteristic longitudinal to transversal sizes of the time averaged volumetric force distributions is estimated as 4:1 according to various experimental data [15-17]. The force distribution varies in the homothetic manner at variation the amplitude or the frequency of applied voltage [17]. The local maximum force density can reach about F_{max}= 10^4·N/m^3 [18].

The total force per one tooth of the exposed electrode is estimated according to (4) as $F_t = 2e^2 F_{max} x_0 y_0 z_0/\pi \approx 4.74\, F_{max} x_0 y_0 z_0$. The following characteristic sizes have been taken in calculations: x_0 = 0.4 mm, y_0 = 0.1 mm, hence F_t = 8.5·10^{-7}·N at $F_{max}\approx$ 10^4·N/m^3.

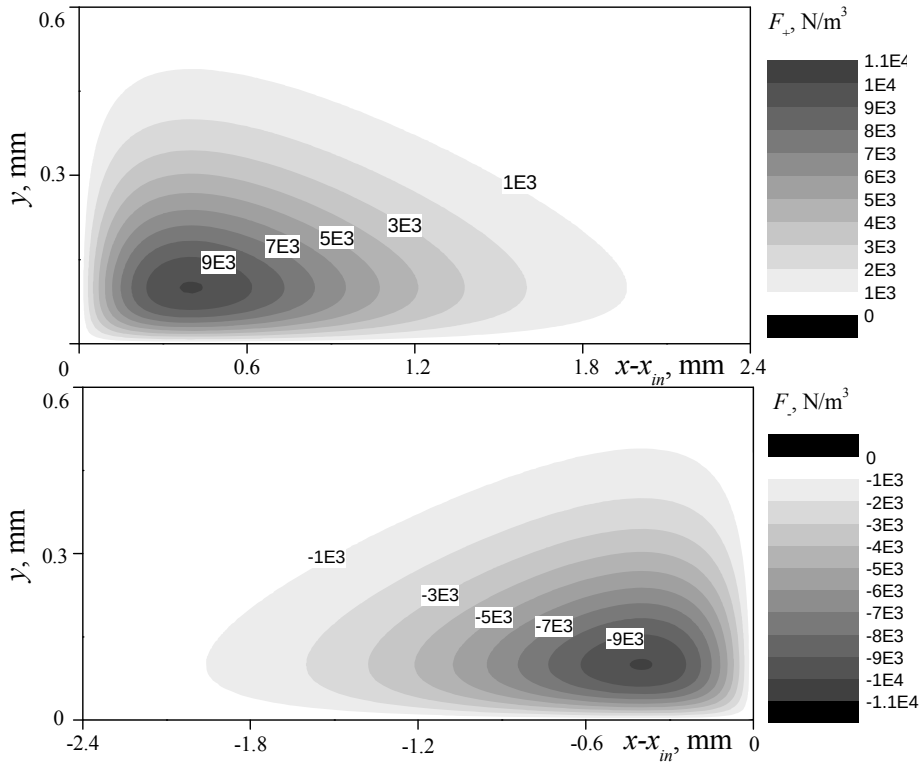


Figure 6: Spatial distributions of the volumetric force sources in x-y plane.

The spatial distributions of the positive and negative force sources in x-y plane at $z = 0.5z_0$ are shown on Fig. 6. It is seen that the vertical size of the sources slightly exceeds the boundary layer thickness on the critical line of the wing leading edge δ_0. The external boundary in the present calculations has been taken as y_e = 1.5 mm. The selected above parameters of the force distributions (4) can bee considered as ultimate for this specific case of boundary layer flow. Further increase in the source intensity would lead to its protrusion in external inviscid flow.

The distribution of the volumetric heat source has been taken in the analytic form approximately accounting for characteristic features of the Joule dissipation distribution in

DBD revealed in numerical simulation [12]. The heat sources in both variants of actuator are naturally positive and take the forms

$$Q_{\pm}(x,y,z) = \pm Q_t \frac{\pi}{2} \frac{x_q^{/}}{x_{q0} y_{q0} z_{q0}} \exp\left(\mp x_q^{/} - y_q^{/}\right) \sin\left(\pi z_q^{/}\right)$$
$$x_q^{/} = \left(x - x_{in}\right)/ x_{q0}, \quad y_q^{/} = y / y_{q0}, \quad z_q^{/} = z / z_{q0}, \quad 0 \le y \le y_e, \quad 0 \le z \le z_{q0} \qquad (5)$$
$$x_{in} \le x \le \infty \text{ for } Q_+, \quad 0 \le x \le x_{in} \text{ for } Q_-$$

The maximal value of heat release according to (5) is reached on the wing surface. The accepted energy efficiency of the actuator $E \equiv F_t/Q_t = 10^{-4}$ s/m corresponds to medium value of efficiency obtained in experiments [18]. The following parameters of the distributions (5) have been used in calculations: $x_{q0} = 0.2$ mm, $y_{q0} = 0.05$ mm, $z_{q0} = 4.5$ mm, $Q_t = 8.5 \cdot 10^{-3}$·W. The spatial distribution of the heat source Q_+ in x-y plane at $z = 0.5z_0$ is shown on Fig. 7. The distribution of the heat source Q_- is reflection symmetric relative to coordinate x_{in}.

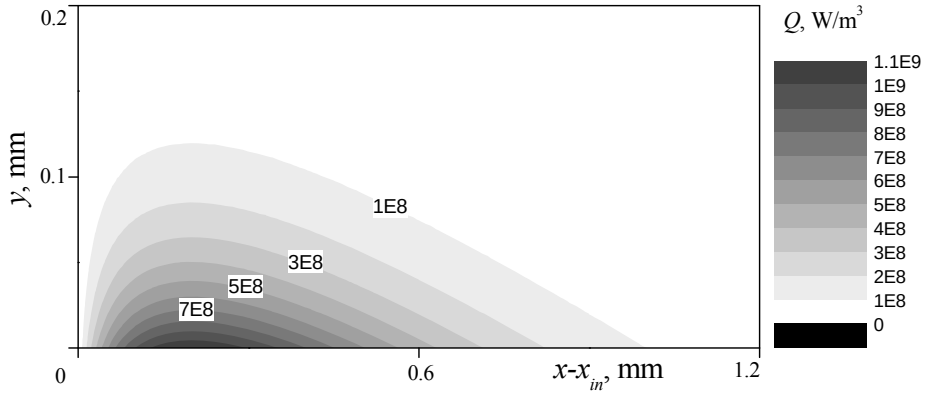


Figure 7: Spatial distribution of the volumetric heat source in x-y plane.

Volumetric force and heat impact of actuator results in span-periodic perturbation of the boundary layer velocity. Spanwise distribution of disturbances of the main-flow velocity component determined by formulas (2) at the distance $x = 6.3$ mm for variant 1 is shown on Fig. 8 as an example. The extremal values of the disturbances of the main-flow and cross-flow velocity components are presented on Fig. 9 both for combined force and heat impact (solid curves) and for only force impact (dashed curves). The effect of only force impact on the flow velocity is weak and the amplitude of the velocity disturbances is of the order of 10^{-2} m/s. The heat impact results in up to the order greater effect.

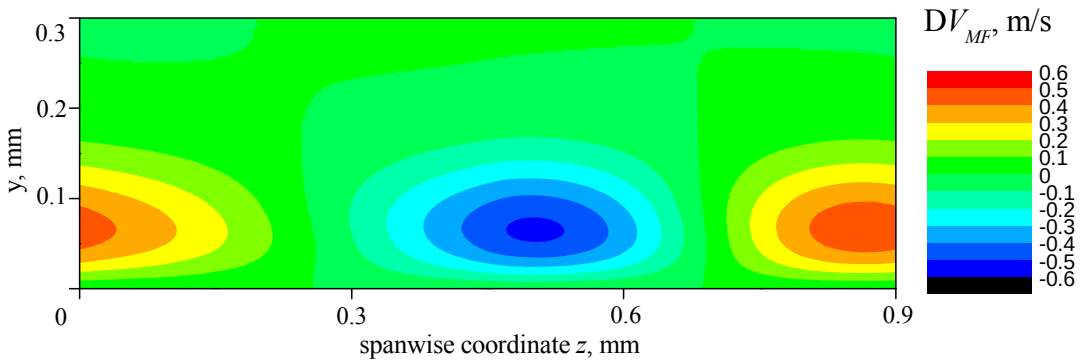


Figure 8: Distribution of main-flow velocity disturbances in z-y plane at $x = 6.3$ mm (variant 1).

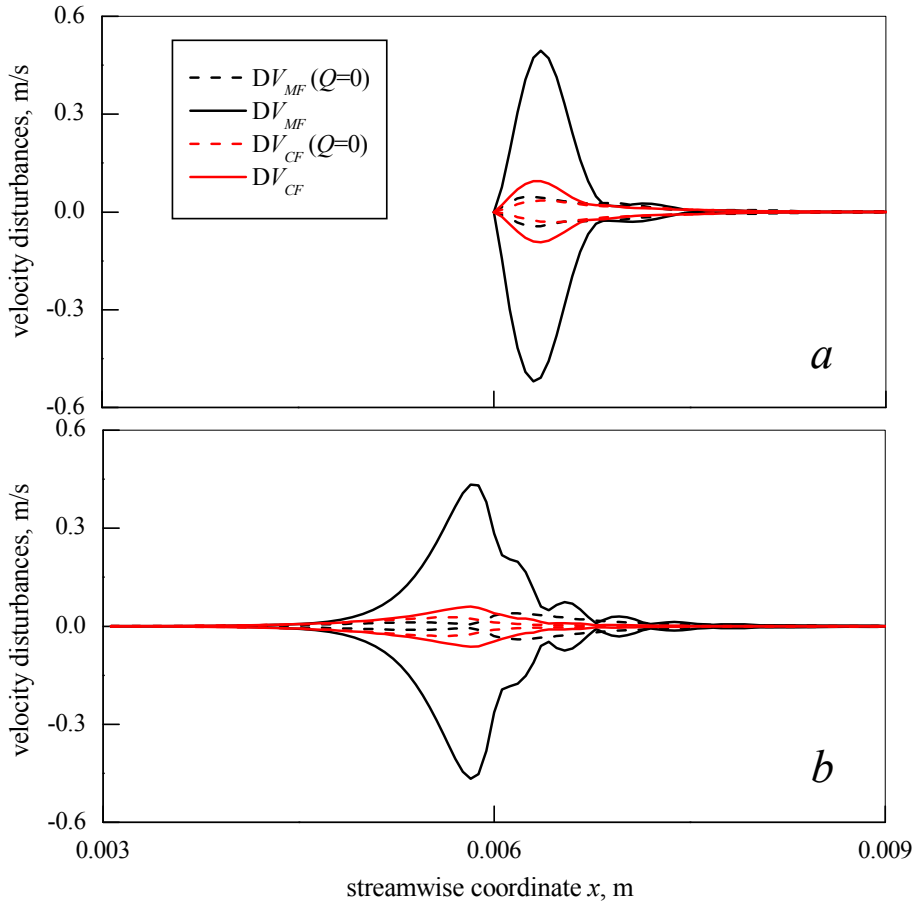


Figure 9: Extremal values of velocity disturbances in variants 1 (*a*) and 2 (*b*).

This appreciable effect of heat impact is explained by the influence of heat release in the boundary layer at a presence of streamwise pressure gradient on streamwise velocity component presenting in equations (2) [19]. Note, the external velocity at the distance $x = 6$ mm reaches $V_e \approx 130$ m/s (see Fig. 3). That is the maximal main-flow velocity disturbances do not exceed 0.4% of the external flow velocity.

The perturbations of the flow vorticity induced by force and heat impact of actuators are the most of interest. The vorticity component oriented along the main-flow streamline $\mathrm{rot}_{MF}\mathbf{V}$ in boundary layer approximation is determined as follows:

$$\mathrm{rot}_{MF}\mathbf{V} = \frac{u_e \mathrm{rot}_x \mathbf{V} + w_e \mathrm{rot}_z \mathbf{V}}{\left(u_e^2 + w_e^2\right)^{1/2}} \equiv \cos\psi\, \mathrm{rot}_x \mathbf{V} + \sin\psi\, \mathrm{rot}_z \mathbf{V}$$
$$\mathrm{rot}_x \mathbf{V} \cong \frac{\partial w}{\partial y}, \quad \mathrm{rot}_z \mathbf{V} \cong -\frac{\partial u}{\partial y} \tag{6}$$

Distributions of the perturbed vorticity oriented along the main-flow streamline at $x = 6.3$ mm for variant 1 (*a*) and $x = 6$ mm for variant 2 (*b*) are shown on Fig. 10. The presented data demonstrate that the force and heat impact of actuators result in three rows of pairs of counter-rotating vortices strongly flattened in vertical direction. The intensities of vortices situated in two rows nearest to the wall are almost equal. The upper row of vortices is significantly weaker. The presented picture of perturbed main-flow vorticity is qualitatively similar to that in the case of solid cylindrical roughness elements [20].

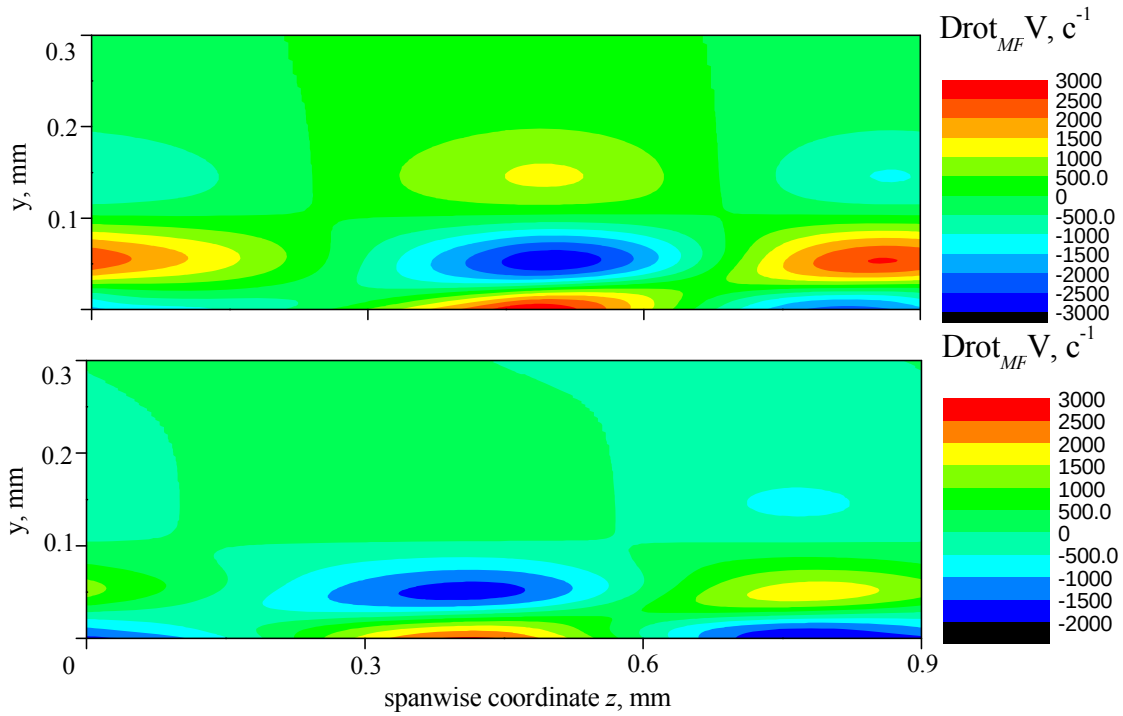


Figure 10: Distributions of perturbed main-flow vorticity in z-y plane for variants 1 (a) and 2 (b).

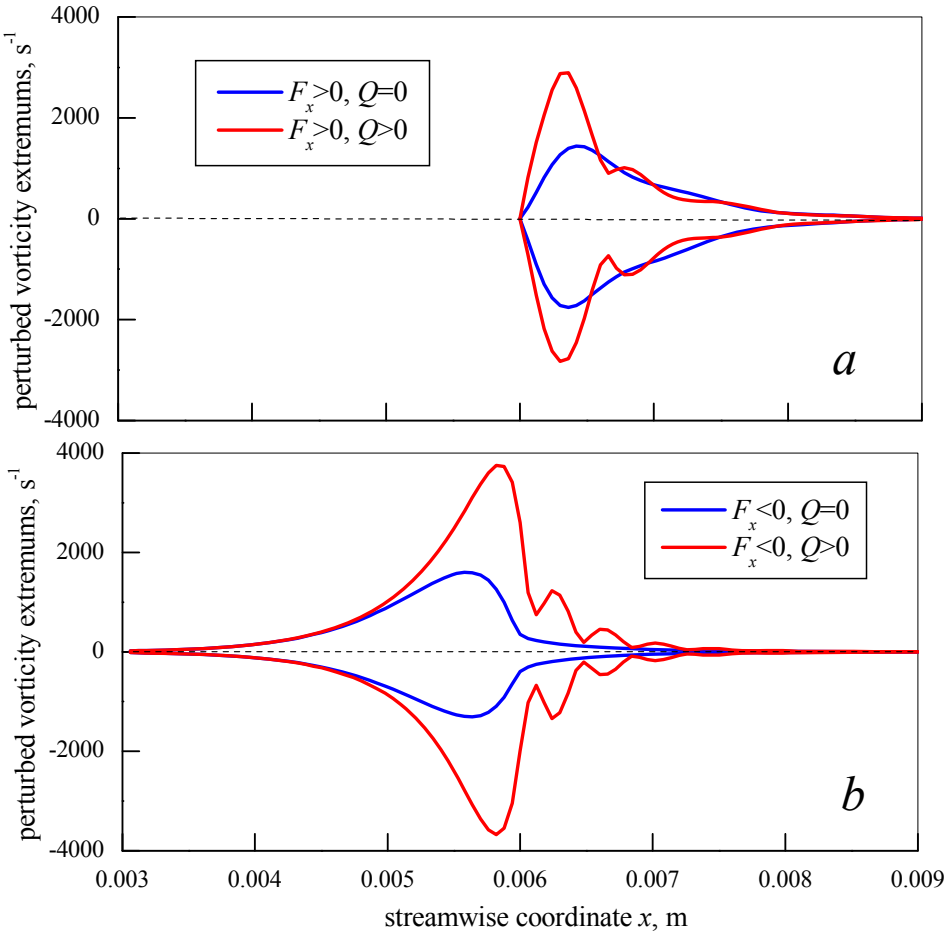


Figure 11: Extremal values of main-flow vorticity perturbation in variants 1 (a) and 2 (b).

The streamwise distributions of extremal values of the perturbed main-flow vorticity are presented on Fig. 11 for force and heat impact and only for force impact. Significant effect of heat release in discharge regions on vorticity perturbation is evident. Moreover the heat impact results in non-monotonic decay of extremal values of perturbations downstream. This phenomenon demands an explanation.

Comparison of Figs. 11 and 6 reveals that very strong vorticity is generated immediately in regions of volumetric force and this vorticity then decays rapidly downstream. For example, the perturbed main-flow vorticity at the distance $x = 8.1$ mm varies in the range -100÷100 s^{-1} in variant 1 and -15÷15 s^{-1} in variant 2. But one can notice that the total streamwise length of the domain of actuator influence is a few greater in the case of variant 2. It seems that the problem of critical mode suppression in the considered peculiar case can be solved by incorporation both variants shown on Fig. 1 in single multi-DBD actuator with common buried electrode. The upstream part of this combined actuator in the form of variant 1 acting downstream from $x \approx 6$ mm (see Fig. 5) will prevent to origin of long wave modes and the downstream part in the form of variant 2 acting upstream from $x \approx 9$ mm will excite controlling forced mode.

5 CONCLUSIONS

- Both considered variants of multi-DBD actuator generate three rows of pairs of counter-rotating vortices situated one above another along external streamline. The upper row is significantly weaker as compared to two lower rows.
- Thermal impact of DBD on vortices generation is comparable with force impact.
- High intensive vorticity (the order of 10^3 s^{-1}) is generated immediately in regions of volumetric force and heat sources and then decays rapidly downstream.

REFERENCES

[1] A. Abbas, J. de Vicente, E. Valero, Aerodynamic technologies to improve aircraft performance. *Aerospace Sci. & Technol*., **28**, 100–132, 2013.

[2] D. Arnal, G. Casalis, Laminar-turbulent transition prediction in three-dimensional flows. *Prog. Aerospace Sci.*, **36**, 173–191, 2000.

[3] W.S. Saric, R.B. Carrillo, M.S. Reibert, Leading-edge roughness as a transition control mechanism. *AIAA Paper*, **781**, 1-13, 1998.

[4] W.S. Saric, D.E. West, M.W. Tufts, H.L. Reed, Flight test experiments on discrete roughness element technology for laminar flow control. *AIAA Paper*, **539**, 1-23, 2015.

[5] E.B. White, W.S. Saric, Application of variable leading-edge roughness for transition control on swept wing. *AIAA Paper*, **283**, 1-11, 2000.

[6] J.R. Roth, D.M. Sherman, S.P. Wilkinson, Electrohydrodynamic flow control with a glow–discharge surface plasma. *AIAA Journal*, **38**, 1166-1172, 2000.

[7] D. Schatzman, M. Wicks, P. Bowles, F. Thomas, T. Corke, A parametric investigation of plasma streamwise vortex generator performance. *AIAA Paper*, **824**, 1-10, 2012.

[8] T. N. Jukes, K.-S. Choi, On the formation of streamwise vortices by plasma vortex generators. *J. Fluid Mech*., **733**, 370–393, 2013.

[9] I. Moralev, V. Sherbakova, P. Kazansky, A. Efimov, V. Bityurin, The structure of the dielectric barrier discharge and its effect on the discharge inintiated gasdynamic disturbances. *Proceedings of EUCASS-2015, Report FP EUCASS-357*, Krakow, Poland, 29 June-3 July, 2015.

[10] R. Joussot, A. Leroy, R. Weber, H. Rabat, S. Loyer, D. Hong, Plasma morphology and induced airflow characterization of a DBD actuator with serrated electrode. *J. Phys. D: Appl. Phys.*, **46**, 125204 (12 pp), 2013.

[11] R. J. Durscher, S. Roy, Three-dimensional flow measurements induced from serpentine plasma actuators in quiescent air. *J. Phys. D: Appl. Phys*., **45**, 035202 (9 pp), 2012.

[12] S.L. Chernyshev, A.P. Kuryachii, S.V. Manuilovich, D.A. Rusyanov, V.V. Skvortsov, Attenuation of cross-flow-type instability in compressible boundary layer by means of plasma actuators. *AIAA Paper*, **321**, 1-16, 2013.

[13] D. Dunn, C Lin, The stability of the laminar boundary layer in a compressible fluid for the case of three-dimensional disturbances. *J. Aeronautical Sci.,* **19**, 491–502, 1952.

[14] D.A. Russianov, M.V. Ustinov, A.A. Uspensky, A. Yu. Urusov, Action of multiple-electrodes DBD actuators on the boundary layer: theory and experiment. *Proceedings of EUCASS-2015, Report FP EUCASS-220*, Krakow, Poland, 29 June-3 July, 2015.

[15] N. Benard, A. Debien, E. Moreau, Time-dependent volume force produced by a non-thermal plasma actuator from experimental velocity field. *J. Phys. D: Appl. Phys*. **46**, 245201 (12 pp), 2013.

[16] J. Kriegseis, C. Schwarz, C. Tropea, S. Grundmann, Velocity-information-based force-term estimation of dielectric-barrier discharge plasma actuators. *J. Phys. D: Appl. Phys*. **46**, 055202 (13 pp), 2013.

[17] N. Benard, M. Caron, E. Moreau, Evaluation of the time-resolved EHD force produced by a plasma actuator by particle image velocimetry – a parametric study. *Journal of Physics: Conference Series* **646**. 012055 (4 pp), 2015.

[18] N. Benard, E. Moreau, Electrical and mechanical characteristics of surface AC dielectric barrier discharge plasma actuators applied to airflow control. *Exp. Fluids.* **55**, 1846 (43 pp), 2014.

[19] A.P. Kuryachii A, S.V. Manuilovich, Attenuation of cross-flow-type instability in a 3D boundary layer due to volumetric force impact. *TsAGI Sci. Journal*, **42**, 345-360, 2011.

[20] D.P. Rizzeta, M.R. Visbal, H.L. Reed, W.S. Saric, Direct numerical simulation of discrete roughness on a swept-wing leading edge. *AIAA Journal*, **48**, 2660-2673, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL SIMULATION OF INTERACTION PROCESS BETWEEN DIELECTRIC BARRIER DISCHARGE AND DUCT FLOW

P. A. Semenev, D. E. Pudovikov, and P. D. Toktaliev

[1] Baranov Central Institute of Aviation Motors (CIAM)
111116, Russian Federation, Moscow, Aviamotornaya str. 2
e-mail: semeyov@gmail.com

Keywords: Dielectric barrier discharge, duct flow, turbulent flow, computational fluid dynamics, RANS, DPM.

Abstract. *Numerical procedure of DBD modeling based on RANS equations has been developed and tested in present study both for the standard DBD and DBD working in constricted mode. Qualitative and approximate quantitative correspondence between known experimental data and numerical results has been achieved. Typical volume force and heat magnitudes has been evaluated, these values could be used as estimation in steady approximation calculations. Both DBD actuators models has shown up that required volume force magnitude has to be increased by 2-3 orders influence on flow separation in S-type engine ducts at relatively high inlet velocity Different method of volume force calculation using actuator electrical parameters (DPM) has been carried out. Continuous volume force components fields have been calculated as a function of DBD actuator potential difference and charge density amplitude.*

1 INTRODUCTION

Modern aviation engine improvement requires high-level scientific maintenance of developing process and research of characteristics increase approaches for different operation stages. Therefore scientific research significance in modern engines design is increasing.

Short diffusers application with a high ratio between inlet and outlet areas with the strongly curved central line helps to reduce engine installation overall dimensions and mass. However, flow separation occurs in these ducts. Longitudinal pressure gradient from centrifugal force effect along with transverse pressure gradient induces secondary flows contorting pressure field.

A wide spectrum of different "passive" (e.g., interceptors) and "aggressive" (e.g., synthetic jets, DBD and SHF-discharges) methods is applying as a way of influence on these undesirable effects by its partial or total removal.

Use of electric discharge (DBD, SHF) influence on gas-dynamic flow has been studying intensively recently.

Interest in DBD actuators application is connected with its advantages: small weight, no flow disturbance by mounted electrodes, absence of complicated mechanical or pneumatic systems, and electrodes ability of being placed in particular location of disadvantageous effects initiation, and also relatively simple active control system for critical regimes only could be created on basis of these actuators.

DBD discharge with its relatively low required power could be an example of significant flow changing (e.g., pressure loss decrease) with low energy contribution, this effect and its gas-dynamic application realization could be succeeded by reasonable placement on streamline surface.

DBD actuator consists of two electrodes placed on surface and divided by dielectric. High voltage alternating current on electrodes causes weak ionization in the nearby area. Ionized gas (plasma) under electric field gradient is exposed to volume force, it leads to additional external flow velocity distortions.

The most common approach to numerical simulation of aircraft construction external flow and engine internal flow currently is usage of different ways of Navier-Stokes equations solving. RANS/URANS is more fine-tuned computational procedure at present; LES requires greater computational powers, but allows obtaining fuller and more reliable separated flow characteristics. Perspective DNS - methods currently can be applied only to model problems due to its significant computational powers requirement. Therefore, integration of DBD flow influence models with CFD solvers seems to be perspective way of DBD numerical simulation.

The method of DBD actuator flow influence accounting by bringing in additional volume force vector in Navier-Stokes equation is largely used [1, 3, 4]. This force could be estimated as a product of charge density and electric filed potential. This approach requires Maxwell equation solving for electric field parameters obtaining or experimental data usage as an estimation of the volume force magnitude.

2 STANDARD DBD SIMULATION

Model duct studied both experimentally and numerically in [1] has been selected as an object for numerical procedure validation. In that paper two DBD actuators create vortex structure that helps to delay boundary layer transition from a laminar to a turbulent state. Computational domain inlet velocity amounts 5 m/s, volume force magnitude nearby the DBD actuator amounts 10^3 N/m3.

An annular model diffuser duct (Fig.1) has studied numerically as main simulation object; application of other ways of flow control such as interceptors and synthetic jets has been also studied in different works.

The axisymmetric duct DBD flow control has considered as a two-dimensional problem. It helps to estimate to a first approximation possibility of flow separation partial or total removal and DBD influence effectiveness in principle and volume force magnitude required for such removal.

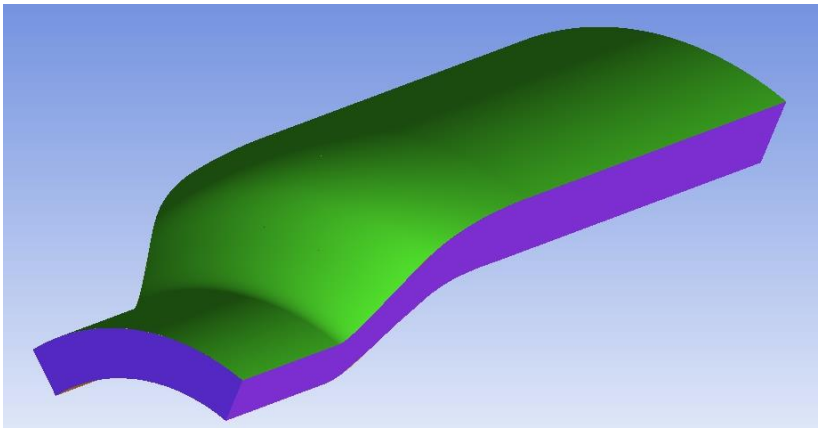


Figure 1: Simulated duct sector (60 degrees).

Volume force effect area is located in near-wall region when the flow separation is started (Fig. 2), size of this area is about 5 mm in a streamwise direction and 1 mm transversally. Exact volume force effect area x-coordinate, volume force magnitude and its vector direction has been considered as variable parameters in this research.

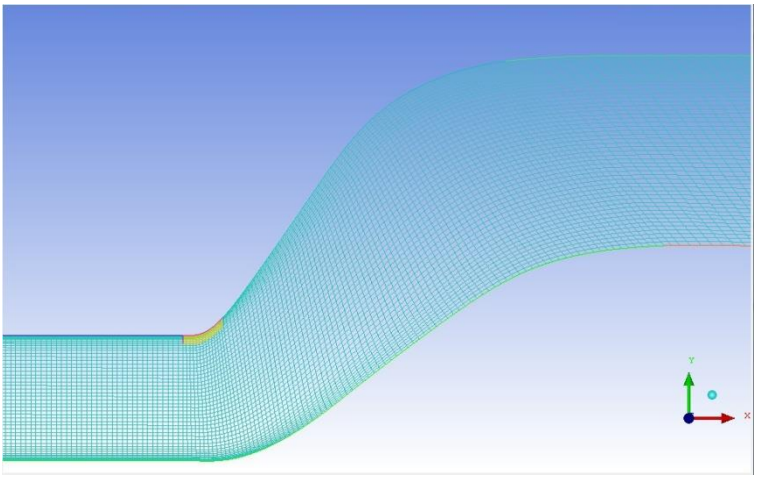


Figure 2: Calculation domain and mesh fragment example (orange area denotes volume force effect area).

The structured 2D-mesh consists of 8 blocks and about 14000 elements. Wall boundary condition has been set on top and bottom domain boundaries (Fig. 3).

Boundary conditions on inlet (left boundary): total pressure 101325 Pa, turbulence intensity - 2%, turbulence scale - 0.001 m; on outlet (right boundary): static pressure 90000 Pa.

2D steady calculation with Roe-FDS flux type solution method has been used. RANS system of equations has supplemented by Spalart - Allmaras turbulence model with standard wall functions.

Volume force magnitude in DBD effect area has varied between 1 and $3 \cdot 10^6$ N/m3. At lesser volume force magnitudes no observable influence on the flow has been noticed. Volume force vector direction (relative to duct axis i.e. relative to inlet velocity direction) has varied between 30 and 60 degrees (Fig. 4). All these variations has been made for a particular volume force effect area location.

Calculation results have shown that volume force with $2 \cdot 10^6$ N/m3 lowers flow separation and volume force with 3•106 N/m3 practically removes it. Volume force vector direction variation (due to X and Y components alteration with the constant magnitude) appears to be insignificant in comparison with force (momentum) effect. Therefore volume force vector direction has been set at a value of 45 degrees (approximately parallel with the nearby upper duct wall)

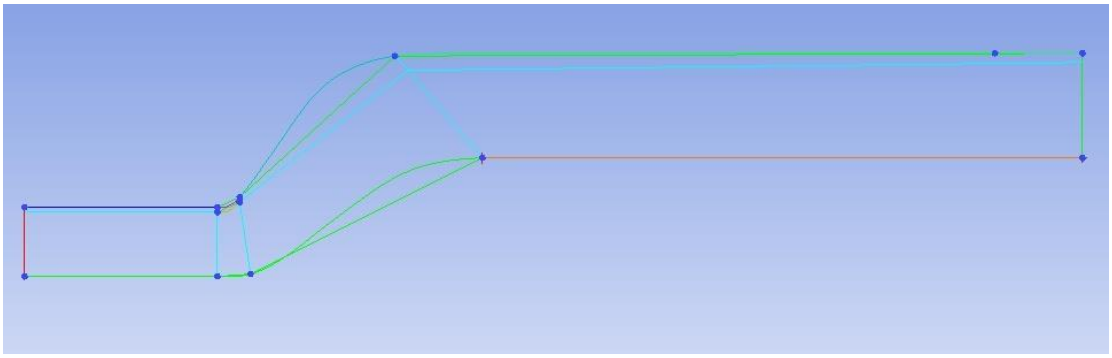


Figure 3: Computational domain block structure.

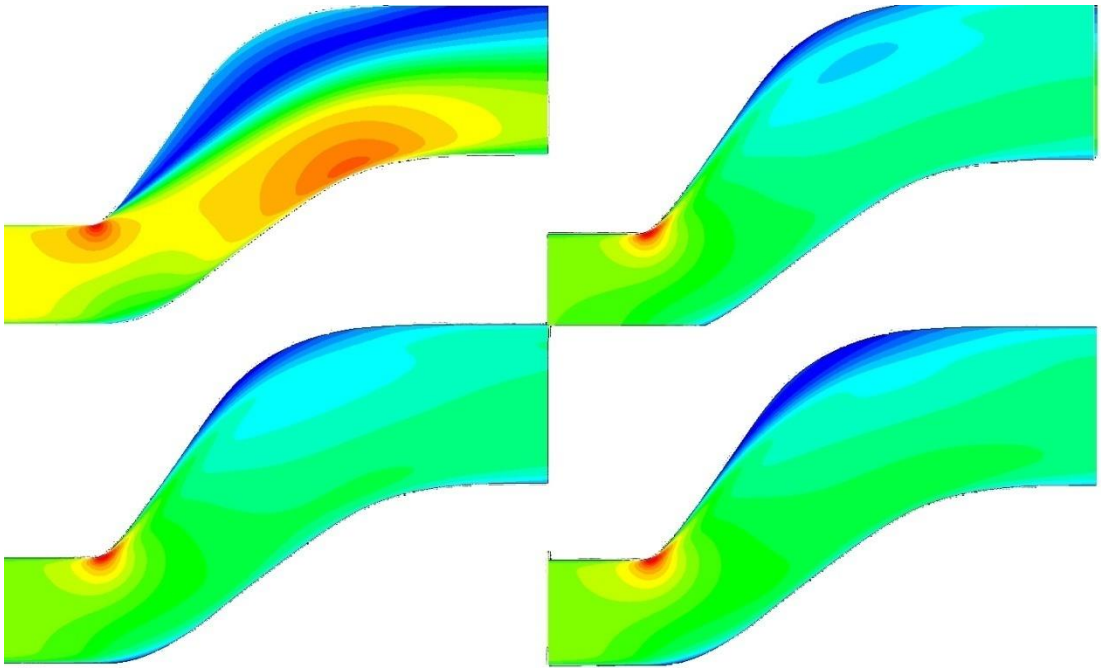


Figure 4: Velocity fields in separation flow area (left to right): zero volume force, volume force vector direction (relative to duct axis) 30 degrees, 45 degrees, 60 degrees with magnitude $3 \cdot 10^6$ N/m^3.

Flow separation appears to be unresponsive to a small volume force effect area movements in a streamwise direction (1...3 mm) also. On the contrary, force (momentum) influence intensity is the most significant parameter.

However, intensity of force (momentum) influence that requires for flow separation removal has considerably large volume force magnitude in comparison with experimental data [1,2]. The experimental upper limit of this force is connected with DBD actuators voltage limitations due to dielectric properties.

3 CONSTRICTED DBD SIMULATION

Along with standard DBD discharges simulation, which could be used for a flow control purpose, dielectric barrier discharges working in a constricted (saturated) mode or constricted DBD, has also considered in this paper.

This mode is connected with significant power input (voltage amplitude increase) which leads to the constriction of the discharge due to overheating instability. A local increase in the conductivity leads to the increase in the current density, and a bright highly conductive filament is formed. (Fig. 5, [2])

Formation of a region with a higher conductivity in the discharge should significantly disturb electric field and charge distribution, providing body force to be radial relatively to a discharge channel (filament) to a first approximation (real structure is more complicated)

Constricted DBD disturbance structure in practically immobile gas is presented in [2]. It has shown that flow structure induced by constricted DBD includes jets perpendicular to dielectric surface, which is placed between discharge channels (Fig. 5)

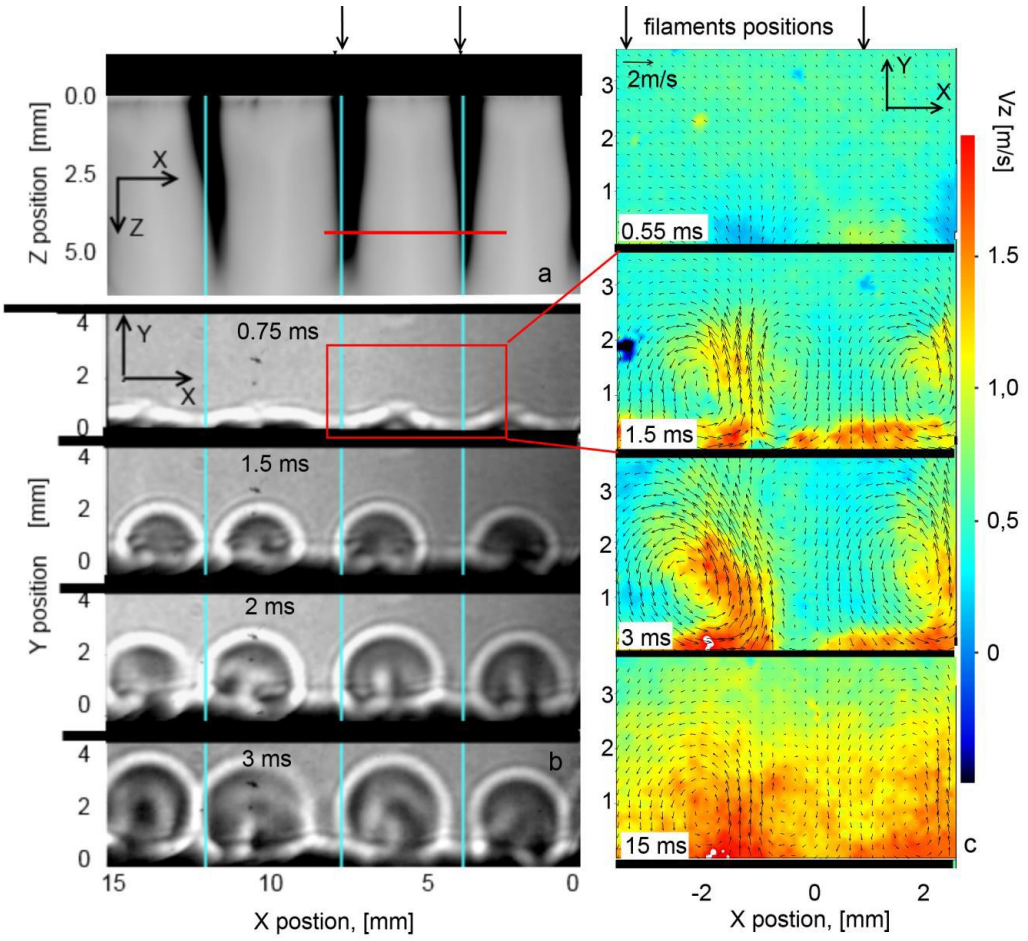


Figure 5: "a": top view (X–Z) of the discharge (negative image). Red line marks the position of the cross section for PIV measurements, "b": shadow pictures of discharge-induced disturbances and "c": enlarged PIV images of the flow in the X–Y plane at the position Z = 4 mm. [2].

Jets forming occur with gas suction from the upper layers. At next stage vortex pairs rotating in different directions are forming attached to discharge channels positions.

An interaction between formed discharge plasma channels and a medium could be described by two mechanisms: energetic mechanism (volume heat release inside the discharge channel and nearby regions), which has sufficiently larger intensity than standard DBD and force (momentum) mechanism which has lesser intensity than standard DBD. These interaction mechanisms are simulating in this study for the purpose of recreating pair vortex structures numerically and estimation of mechanisms intensity.

A validation of numerical procedure along with volume force and volume heat magnitude estimation has been carried out at 3D model of DBD actuator from [2]. Three domain versions ("a" - "c") which help to estimate necessary domain width are presented at Fig.6.

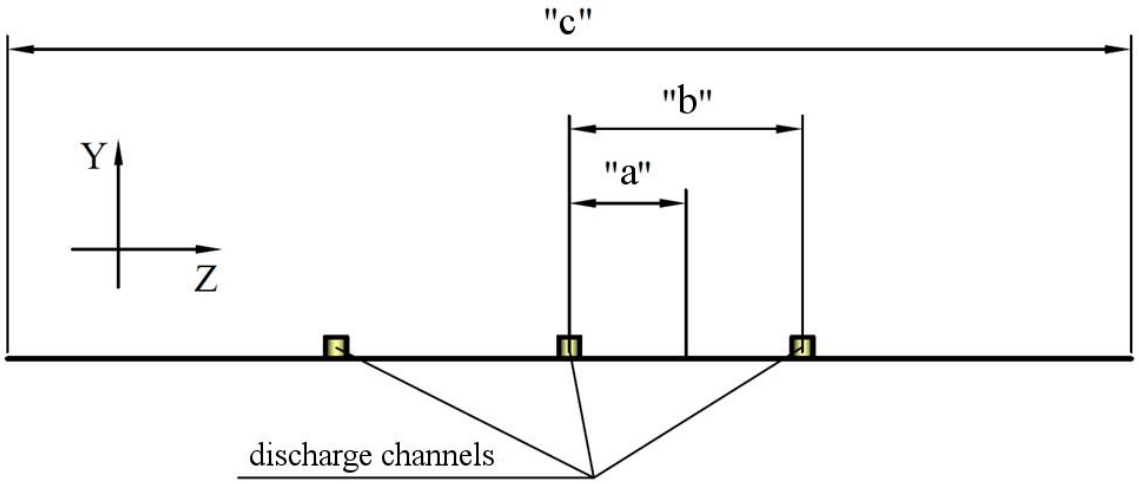


Figure 6: Three computational domain types which differ in Z-dimension: "a" - half inter-channel space, "b" - one inter-channel space, "c" two inter-channel spaces with left and right boundaries moved aside.

Inside the constricted discharge channels matter is in plasma state (temperature experimental estimation is 1500 K at the basis of channel) hence it does not pass electric field inside and volume force magnitude equates zero; experimental estimation of heat source is 10^{10} W/m^3.

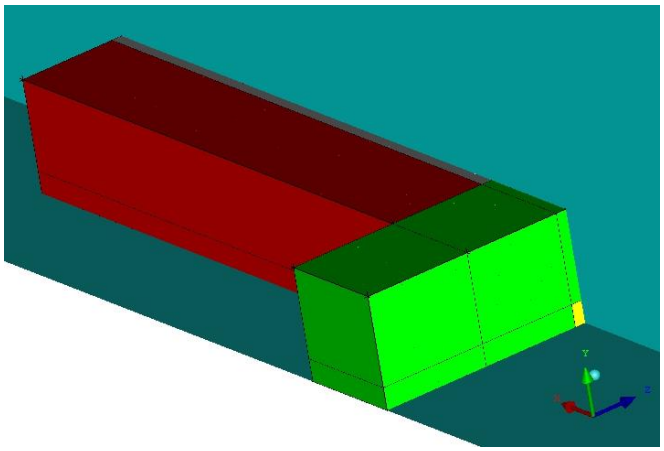


Figure 7: Configuration of computational areas with the source terms ("a" - domain type): green color denotes an area with the longitudinal volume force (X-direction), red color denotes the transversal volume force (Z-direction), and yellow color denotes discharge channel with heat sources and zero volume force.

The volume force effect area size has selected on the ground of experimental and numerical research data [1-4]. An area with longitudinal volume force (X-direction) has been set at the basis of the channel (Fig. 7); an area with transversal volume force (Z-direction) has been set along the discharge channel region.

Exact value of volume force magnitude has been clarified during the validation simulations. As a first estimation results from [1-2] has been used, qualitative correspondence to experiment [2] was the main criteria during the volume force magnitude refinement.

Three computational domain structural 3D-meshesconsist of 36 blocks and about 2 million elements, 72 blocks and about 4 million elements, 180 blocks and about 4 million elements for "a", "b" and "c" domain types respectively (Fig.6). Mesh and block structure fragments are shown at Fig. 8.

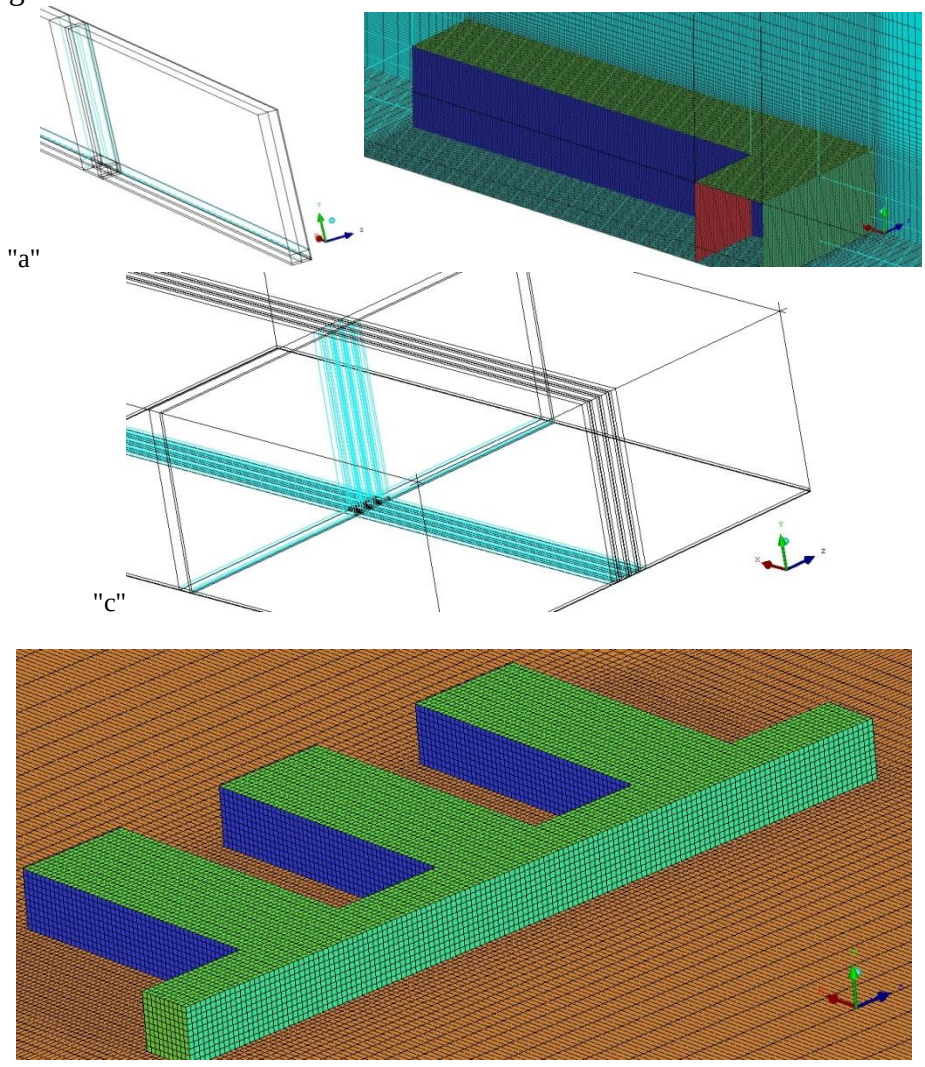


Figure 8: Computational mesh and block structure fragments for "a" and "c" domain types.

Boundary conditions on domain inlet : total pressure 101325 Pa, temperature 300K turbulence intensity - 2%, turbulence scale - 0.001 m; on outlet : static pressure 101324,7 Pa. Co-current flow velocity has value of 0,5 m/s under this conditions and equals experimental flow velocity.

3D steady calculation with Roe-FDS flux type solution method has been used. RANS system of equations has supplemented by Spalart - Allmaras turbulence model with standard wall functions.

Experimental estimation of heat source (1010 W/m3) has been based on one-channel power input estimation (5 W). The value of volume heat source magnitude has been also refined for a steady numerical problem in b domain type (Fig. 6). A correspondence of experimental and numerical temperature (600...1500 K) was the refinement criteria. The corrected volume heat source magnitude has been set at $2 \cdot 10^9$ W/m3 for the further calculations.

The selected volume force magnitude value has been set at $6 \cdot 10^3$ N/m3 for the further calculations. At this force (momentum) intensity in "b" and "c" domain types experimental pair vortex structures have been reproduced numerically. It has to be mentioned that "c" domain type shows the best experiment approximation of three options (Fig. 9).

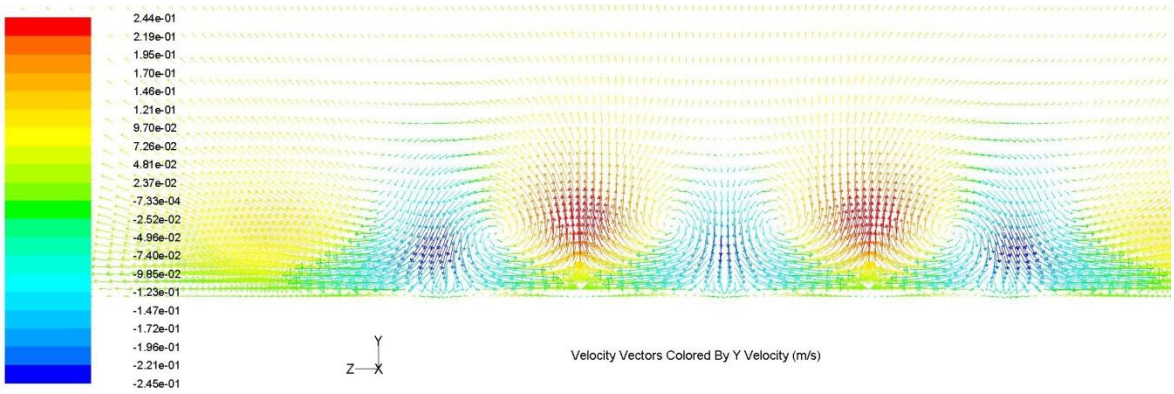


Figure 9: Y-velocity vector field at the last third of discharge channel cross-section ("c" domain type).

Actuators based on constricted DBD are perspective device for the real engine duct separation and pressure loss control. An annular model diffuser duct (Fig.1) with constricted DBD actuators has studied numerically as an application effectiveness estimation of this device.

Z-curvature of the segment of annular model duct (Fig.10) was disregarded due to small computational region width.

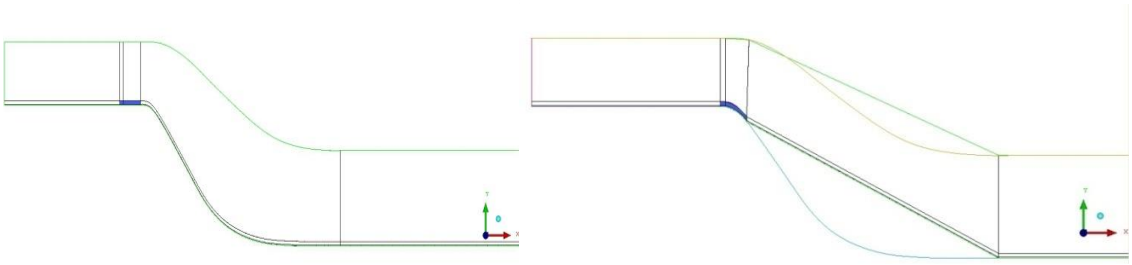


Figure 10: Geometry and block structure of the aggressive model duct, blue areas denote to different actuator and volume force effect area locations (left to right: #1 and #2).

Domains with two different actuators (hence, the volumes force area) location has considered (Fig. 10): #1 – directly before the separation zone, #2 – at the beginning of the separation zone.

Volume force and heat areas topology and channel quantity is similar to a validation problem "c" domain type with three discharge channels in Z-direction.

Solver type, boundary conditions, turbulence model and source terms magnitudes has been taken from a validation problem. Computational consists of 180 blocks and 4 million elements (Fig. 11).

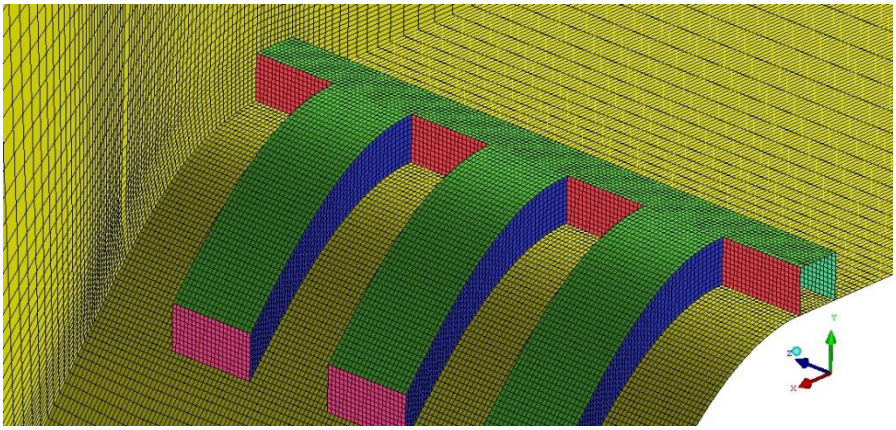


Figure 11: Computational mesh fragment for #2 volume force area location.

Z-direction volume force magnitude was the variable parameter. Duct pressure differential has been set at the value of 7825 Pa (101325 Pa and 93500 Pa, total inlet pressure and static outlet pressure respectively), inlet velocity has the value of 93 m/s under this conditions.

Velocity magnitude fields have presented on Fig.12. Volume force magnitude has been increased from zero (Fig.12, a) to a considerably large value, which affects flow separation zone (Fig.12, b, c).

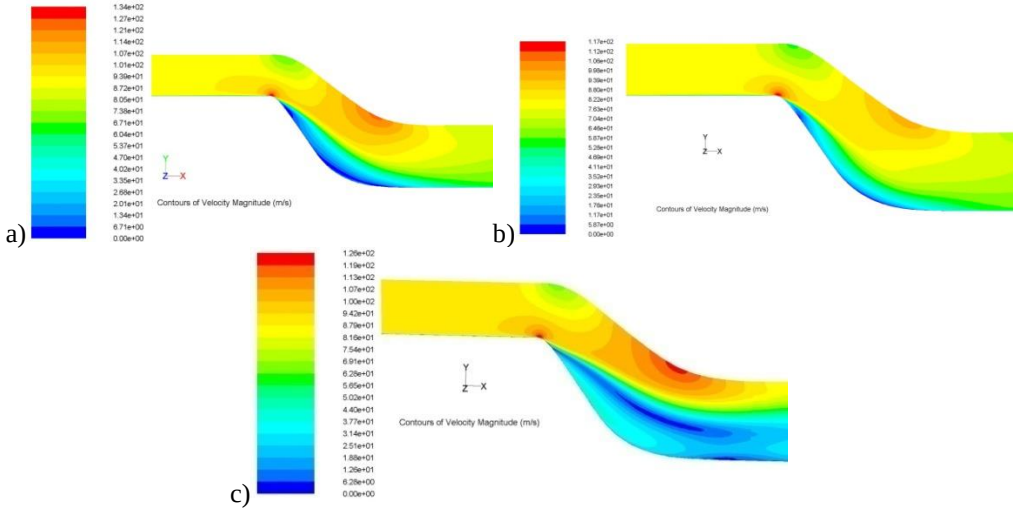


Figure 12: Velocity magnitude fields in longitudinal section: a) – zero volume force magnitude; b) – scheme #1 (Fig.10) volume force area location, c) – scheme #2 volume force area location.

The increased volume force magnitude value has been set at value of $2 \cdot 10^6$ N/m3.

DBD actuator #1 location has practically removed flow separation, except small region with reverse current (Fig.12, b). DBD actuator #2 location has partially promoted flow separation, because force (momentum) influence is located in separation zone mostly. It has to be mentioned that even a considerably large volume force magnitude value does not lead to a total flow separation zone removal in Z-direction (Fig. 13). It is presumably connected with pair vortex structures influence on the flow, which creates nonuniform flow structure.

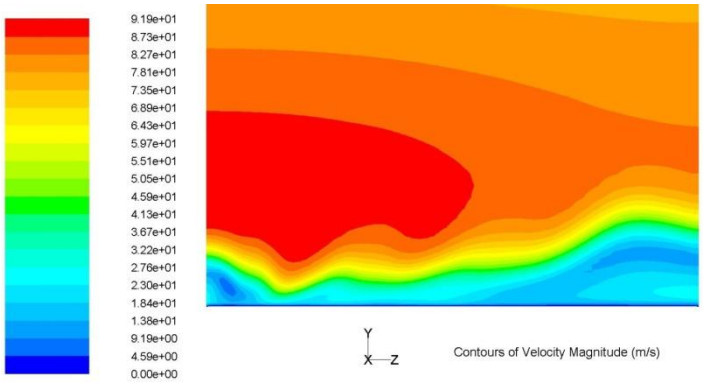


Figure 13: Velocity magnitude field in a cross section after DBD actuator location at the beginning of the diffuser part (scheme #1).

Constricted DBD flow control research results are similar to standard DBD ones. Significant effect on the flow separation zone could be reached only by applying large force (momentum) influence, which has not reached experimentally yet.

4 DPM APPROACH

Besides constant volume force regions approaches there are alternative methods connected with DBD actuator electrical potential calculation. These approaches also imply more or less detailed mathematical model of plasma formation and its interaction with the flow. Main advantage of this more complicated approach is achievement of continuous volume force field which corresponds to features of physical processes near actuator.

Computational domain was analogous to previous calculations (Fig.14). The only difference is additional area 2 mm depth (black boundaries on Fig.14) modeling the dielectric material of actuator. Distance between the electrodes in streamwise direction is 0.5 mm, electrodes length – 1 mm.

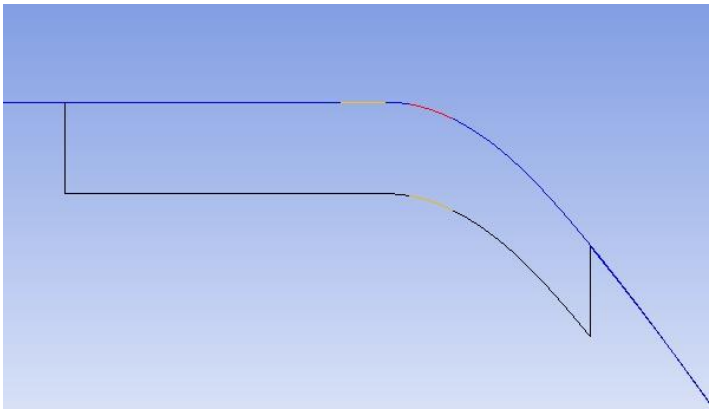


Figure 14: Boundary conditions in computational domain: orange color denote to electrodes, red color denote to area of variable charge density.

Computational 2D-domain consists of $5 \cdot 10^4$ elements. Dual-potential Method (DPM) was used for electrical characteristics (electrical potential φ and charge density ρ_C) calculations. This approach connected with solving a two-equation system for two components of electrical potential:

$$\varphi = \Psi + \Phi \tag{1}$$

Equations for both components are of the form:

$$\nabla \cdot (\varepsilon_r \nabla \varphi) = \frac{1}{\lambda_D^2} \varphi \tag{2}$$

Connection between charge density and electrical potential is of the form:

$$\rho_c = -\frac{\varepsilon_0}{\lambda_D^2} \varphi \tag{3}$$

All empirical constants and solution method has been taken from [5].

Boundary conditions for two scalar values are:

$\varphi = \varphi_0 sin(wt),\ \rho_C = 0$ - for exposed electrode,

$\varphi = 0$ - for buried electrode,

$\frac{\partial \varphi}{\partial \bar{n}} = 0,\ \rho_C = \rho_{C0} sin(wt)$ - for variable charge density area,

$\frac{\partial \varphi}{\partial \bar{n}} = 0,\ \frac{\partial \rho_C}{\partial \bar{n}} = 0$ - for the wall above the dielectric,

$\frac{\partial \varphi}{\partial \bar{n}} = 0, \rho_C = 0$ - for the rest boundaries.

Typical fields of electrical potential φ and charge density ρ_C near the electrodes is presented on Fig.15 and 16.

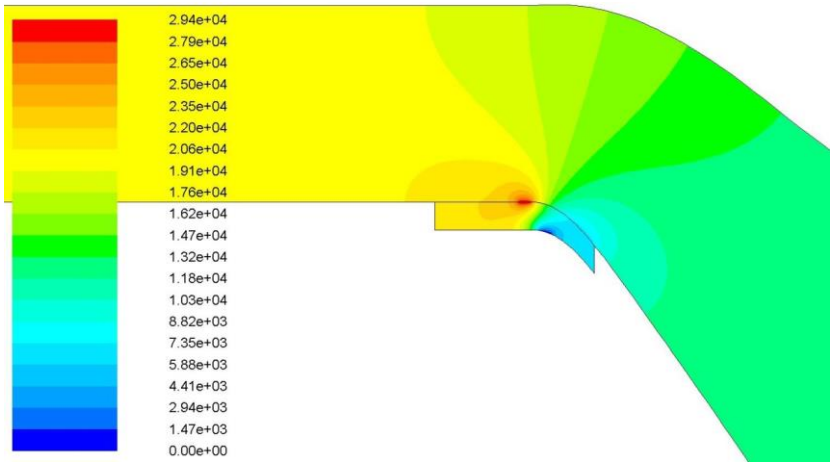


Figure 15: Typical instantaneous electrical potential φ (V) field.

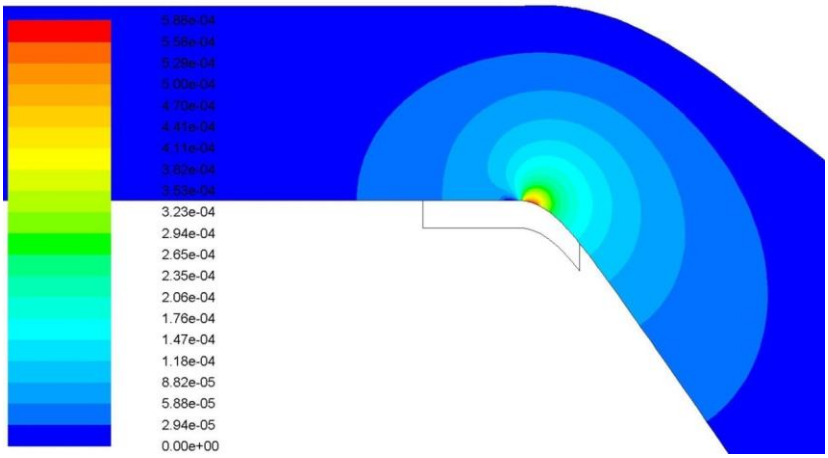


Figure 16: Typical instantaneous electrical potential ρ_C (C/m^3) field.

Using data of this calculation during one period of potential calculation of time-averaged volume force has been performed. The result time-averaged field of F_x and F_y force component for $\Delta\varphi$=50000 V, f=5000 Hz case is presented on Fig.17 and 18.

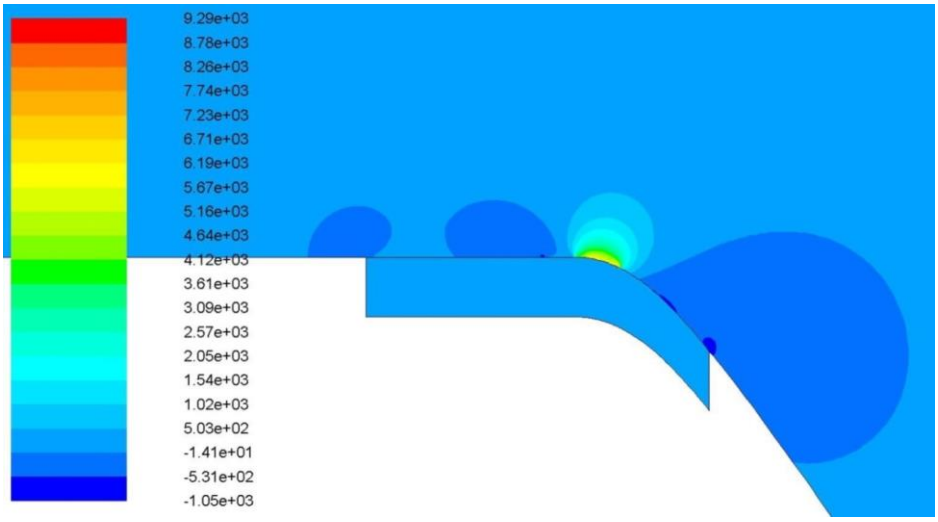


Figure 17: Time-averaged field of F_x force component for $\Delta\varphi$=50000 V, f=5000 Hz case (N/m^3).

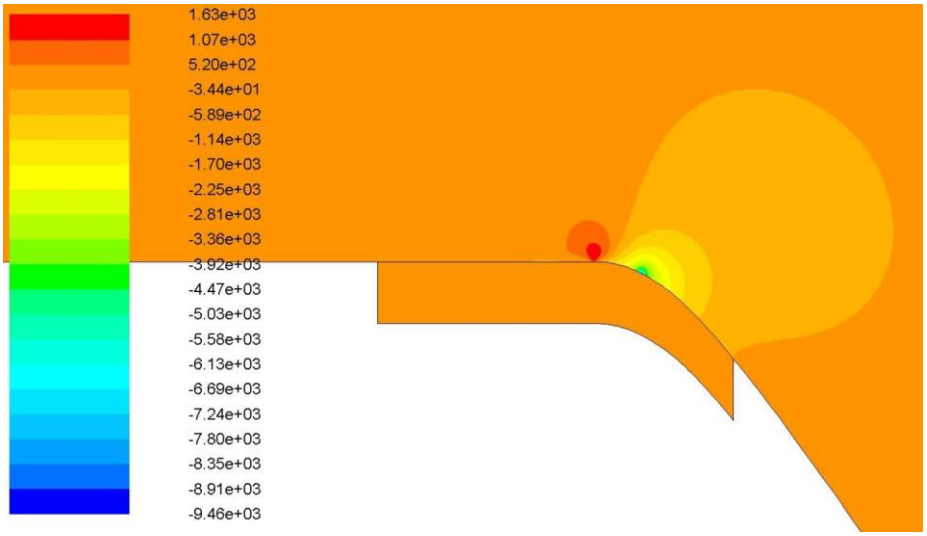


Figure 18: Time-averaged field of F_y force component for $\Delta\varphi$=50000 V, f=5000 Hz case (N/m^3).

First results of DPM approach application are in qualitative agreement with other authors results [6-7].

Maximum values of volume fields components for potential difference $\Delta\varphi$=50000 V, charge density amplitude ρ_{0C}=0.1 C/m^3, and voltage frequency 5000 Hz are: F_{xmax}=1.859*10^6 N/m^3, F_{ymax}=1.892*10^6 N/m^3.

Significant difference between piecewise constant volume regions approach means that it could be used mostly only as a first approximation.

5 CONCLUSIONS

Numerical procedure of DBD modelling has been developed and tested in present study for both the standard DBD and DBD working in constricted mode. Qualitative and approximate

quantitative correspondence between known experimental data and numerical results has been achieved. Typical volume force and heat magnitudes has been evaluated, these values could be used as estimation in steady approximation calculations.

Calculations have shown that experimental force (momentum) and heat influence intensity levels appears to be enough for a low velocity flow control (Mach number about 0.1 and lower).

However both DBD actuators models has shown up that required volume force magnitude has to be increased by 2-3 orders (in comparison with validation problems) to influence on flow separation in S-type engine ducts at relatively high inlet velocity (about 100 m/s).

Alternative method of volume force calculation using actuator electrical parameters (DPM) has been carried out. Continuous volume force components fields have been calculated as a function of DBD actuator potential difference and charge density amplitude.

REFERENCES

[1] Brandt A. Belson, Ronald E. Hanson, Denis Palmeiro, Phillip Lavoie, Katelyn Meidell, Clarence W. Rowley. Comparison of plasma actuators in simulations and experiments for control of bypass transition. *In:50th AIAA Aerospace Sciences Meeting*, 9–12 January 2012, Nashville, Tennessee.

[2] I. Moralev, S. Boytsov, P. Kazansky, V. Bityurin. Gas-dynamic disturbances created by surface dielectric barrier discharge in the constricted mode. *Exp Fluids 55:1747*, 2014.

[3] Y. B. Suzen, P. G. Huang, J. D. Jacob, D. E. Ashpis. 2005. Numerical Simulations of Plasma Based Flow Control Applications. *35th Fluid Dynamics Conference and Exhibit*. AIAA 2005-4633

[4] Arvind Santhanakrishnan, Daniel A. Reasor Jr.y, and Raymond P. LeBeauJr.z.2008. Unstructured Numerical Simulation of Experimental Linear Plasma Actuator Synthetic Jet Flows.*46th AIAA Aerospace Sciences Meeting and Exhibit, Jan.* 7-10, 2008, Reno, NV. DRAFT.

[5] B.E. Mertz, T.C. Corke, Time-dependent dielectric barrier discharge plasma actuator modeling. *47th AIAA Aerospace Sciences Meeting, Orlando, FL*, AIAA Paper, vol.1083, 2009

[6] Sheng-ji Dai , Yang Xia , Li-ming He, Tao Jin, Qian Zhang , Peng-hui Hou, Zi-chen Zhao. Film-cooling of cylindrical hole with downstream surface dielectric barrier discharge actuators. *International Journal of Heat and Mass Transfer 90* (2015) p.825–837

[7] E. Pescini, D.S. Martínez, M.G. De Giorgi, A. Ficarella. Optimization of micro single dielectric barrier discharge plasma actuator models based on experimental velocity and body force fields. *Acta Astronautica 116* (2015) p. 318–332

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

IMPLEMENTATION OF POD AND DMD METHODS IN APACHE SPARK FRAMEWORK FOR SIMULATION OF UNSTEADY TURBULENT FLOW IN THE MODEL COMBUSTOR

Michael D. Kalugin, Sergei V. Strijhak[1]

[1] Institute for System Programming of the Russian Academy of Sciences
109004, Russia, Moscow, Alexander Solzhenitsyn st., 25
e-mail: m.kalugin@ispras.ru, strijhak@yandex.ru

Keywords: Rearward facing step, Turbulent Flow, Finite volume method, Grid, Computational Fluid Dynamics, Vortices, Apache Spark, POD, DMD, snapshots, modes.

Abstract. *The paper is devoted to modelling and analysis of unsteady turbulent flow in a model combustor (channel) using LES (Large Eddy Simulation). Simulations were provided for 2D and 3D cases on different grids of a flow in a channel with rearward facing step. The calculation of a flow in a channel was performed on high performance cluster, using the new approach in Apache Spark framework and POD, DMD data processing algorithms. First 4 dynamic modes were defined.*

1 INTRODUCTION

Solution of Continuum Mechanics problems (hydrodynamics, turbulent combustion, aerodynamics and aeroacoustics) is an important direction for solving fundamental and applied tasks. Usually, numerical simulations are carried out with detailed grids (tens and hundreds of millions of cells) using high performance clusters. Scale turbulence resolution requirements impose certain restrictions on the cells size when using Hybrid RANS/LES and LES methods. As a result, calculation of unsteady flow generates large amounts of data (tens of terabytes), processing of which requires considerable time and computing resources.

2 PROBLEM STATEMENT

This paper considers the problem of incompressible subsonic flow simulation in a channel with a rearward facing step. This kind of flows occurs in various technical devices, including combustion chambers of aircraft engines and helicopters. The main feature of such flows is the formation of large-scale structures in the mixing layer behind a step, which is used as a flame stabilizer [1,2]. The extension of the mixing layer is determined by the dynamics of large vortices development, which are formed upstream and increase as the result of neighboring vortices merging and inclusions of viscous gases from the main flow (Fig.1). The mixing process inside of the large scale structures is determined by small-scale turbulence. The Reynolds number in our case is 22000.

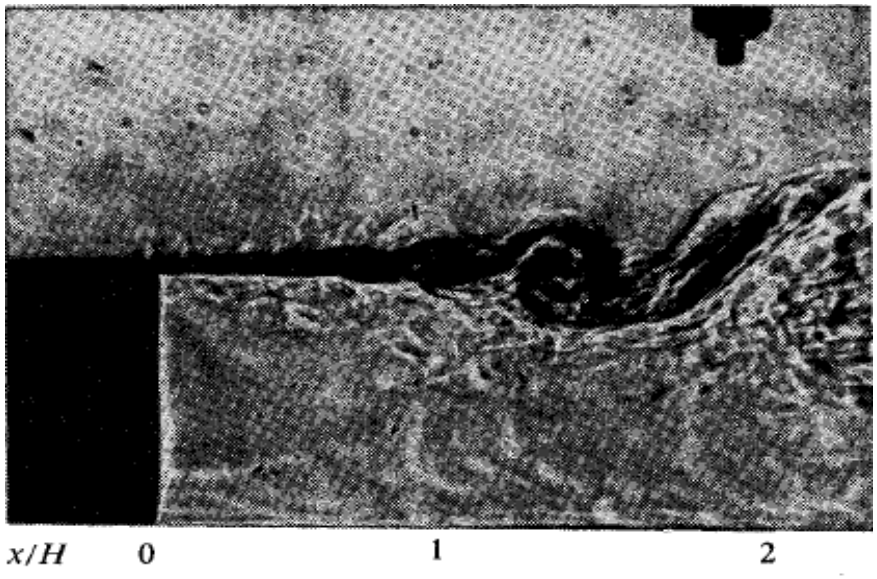


Figure 1: Schlieren photograph.

3 MATHEMATIC MODEL AND CALCULATION RESULTS

In this paper, the authors used open source package OpenFOAM, pisoFoam solver [3], for calculation of unsteady incompressible flow in a channel with a rearward facing step (Fig.2).

For the simulation of unsteady flow it is recommended to use Large Eddy Simulation (LES) model. Large-scale vortex structures are calculated by integrating the filtered Navier-Stokes equations [2,3]. To obtain the filtered equations the boxed filter is used. Small vortices, sized not exceeding the pitch of the computational grid, are simulated using the Smagorinsky model and model based on one differential equation for sub-grid kinetic energy. Finite volume method is used for discretization of the equations. The calculation time step is $dt = 10^{-5}$. The velocity at entrance were set at U = 10 m/s taking into account the overlapping of perturba-

tions for the simulation of a turbulent flow. The nonpermeability conditions were set on the walls. Continuation of the flow boundary condition was set at the exit. Calculation scheme has the second order of accuracy. The grid is based on hexahedrons. The block grid with 13 blocks was built-up for 2D and 3D calculations (Fig.3). The number of cells for 2D case ranged from 626 up to 2690 (Fig.4). For 3D case there were 244500 cells.

The resulting equation for the relation of velocity and pressure are solved by the iterative PISO algorithm. The system of linear equations was solved by the method of conjugate gradients with preconditioner.



Figure 2: Domain of computation Characteristic dimensions.

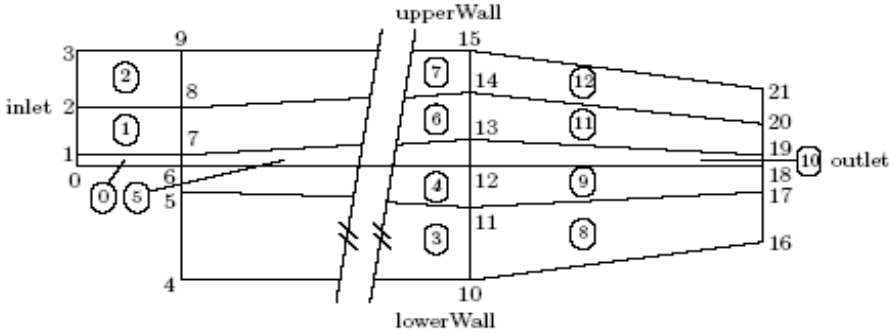


Figure 3: Domain of computation and blocks.

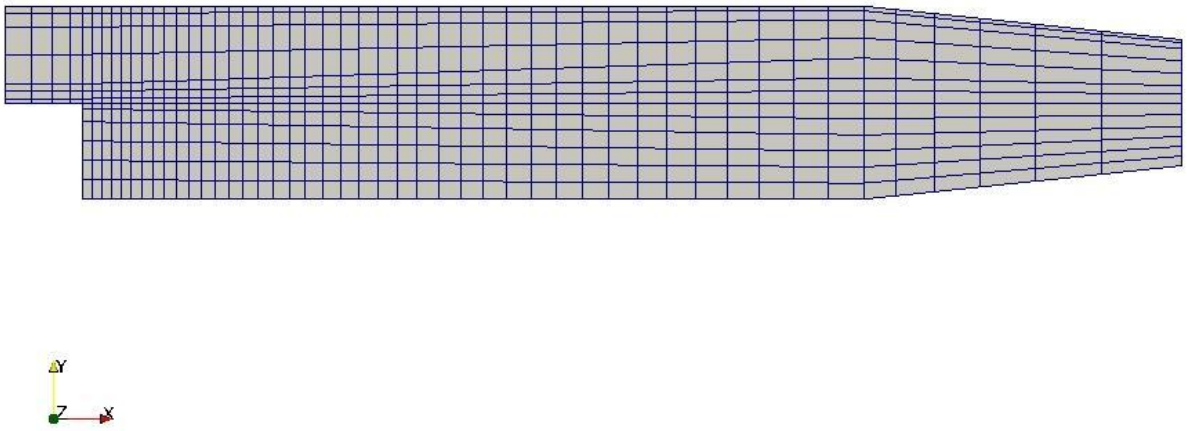


Figure 4: Coarse grid

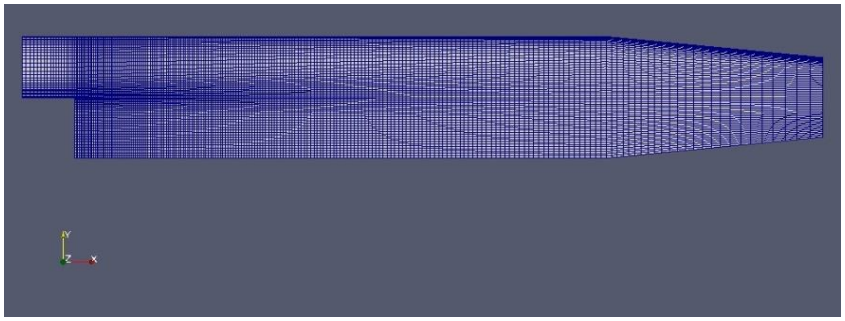


Figure 5: Detailed (Fine) grid.

The calculation results analysis by Paraview of the velocity magnitude value had revealed that the large-scale coherent structures, 3D vortex are formed behind a rearward facing step. Animated picture of the flow around the step has a pronounced unsteady flow character. Periodic self-quenching of the large-scale vortices were registered with a frequency of f=200 Hz. Figures 6-9 illustrate the results of the velocity magnitude calculation. The simulation cases were run with 24 cores on high performance cluster of ISP RAS.

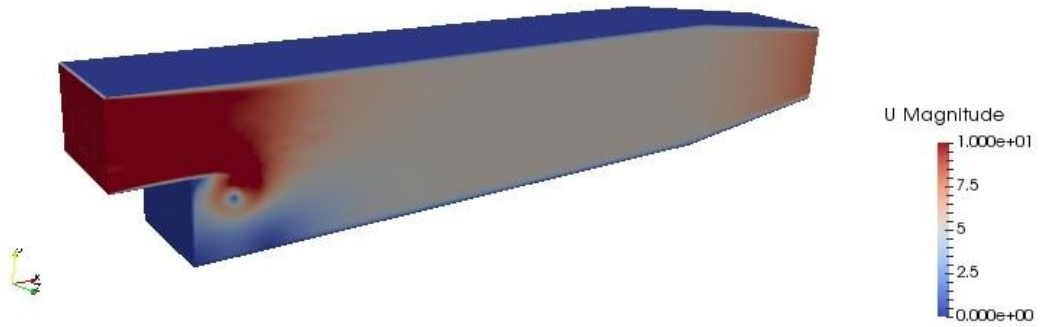


Figure 1: Field of velocity magnitude at the time t=0.005 s.

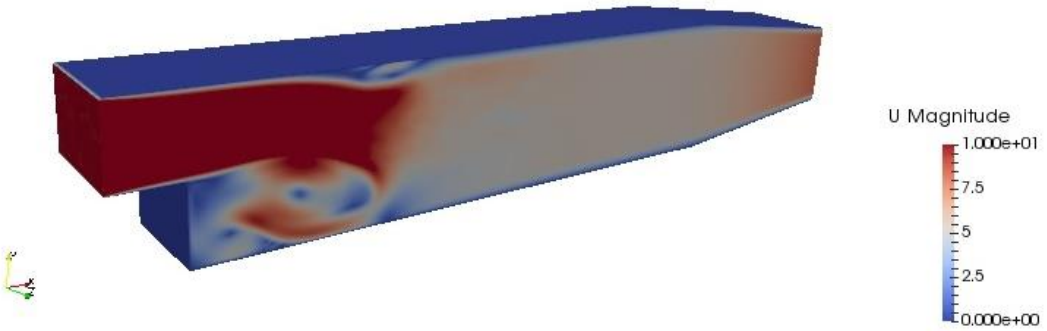


Figure 7: Field of velocity magnitude at the time *t*=0.02 s.

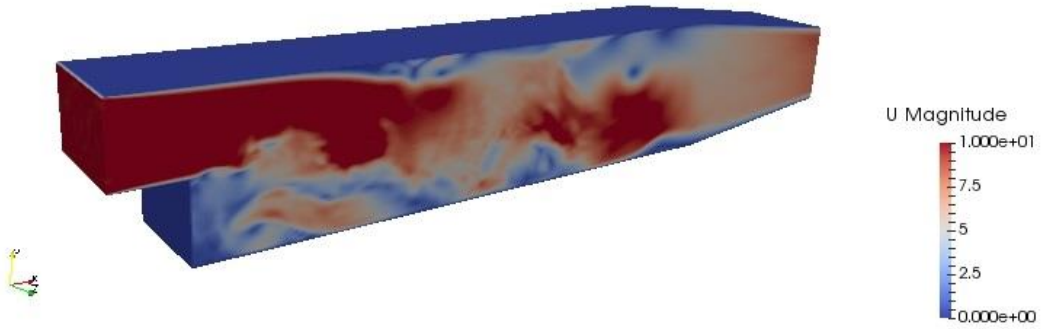


Figure 8: Field of velocity magnitude at the time t=0.05 s.

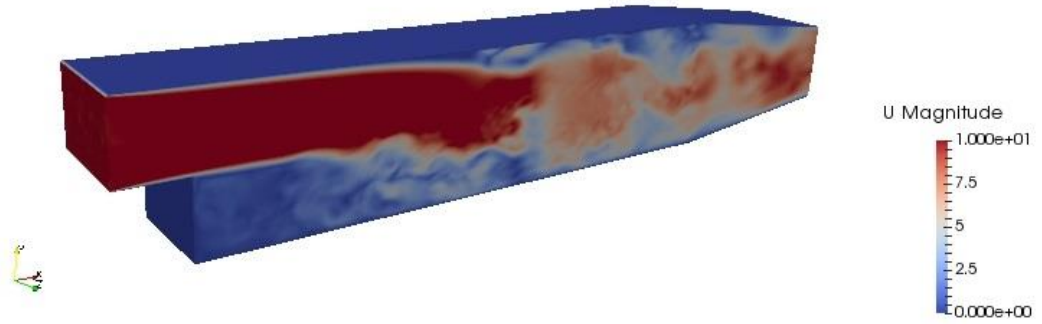


Figure 9: Field of velocity magnitude at the time t=0.1 s.

4 APACHE SPARK TECHNOLOGY AND POD AND DMD METHODS

Apache Spark technology allows to process large amounts of data in ram memory by distributing the load between the compute nodes, ensuring high fault tolerance (Fig. 10). When calculating the flow process, considerable amount of data in the form of separate time snapshots of the flow was obtained.

The POD and DMD algorithms allow to obtain new data on turbulence and coherent structures. Authors of this paper have applied POD and DMD algorithms using Apache Spark framework and LAPACK mathematical library.

The POD was first introduced by Lumley in the field of Computational Fluid Dynamics [4]. The present day analysis uses method of snapshots introduced by Sirovich [5]. To compute the POD using the Eq. 1 requires solving n × n eigenvalue problem.

$$\int_0^T R(t_1,t_2)\phi(t_2)\,dt_2 = \lambda\phi(t_1) \tag{1}$$

Where $R(t_1,t_2)$ is the auto-covariance of the fluid variable.

The main problem is the calculation of auto-covariance matrix $R(t_1,t_2)$. Snapshots method proposes that auto-covariance matrix $R(t_1,t_2)$ can be approximated by a summation of snapshots.

$$R_{ij} = \frac{1}{M}\sum_{n=1}^{M} x_i^n x_j^n \tag{2}$$

The snapshots are assumed to be distanced by a time or a special distance greater than the correlation time or distance.

The input data for DMD [6-8] have to be presented in the form of sequence of snapshorts and are set by a matrix V with the size n×m, where

$$V_1 = [v_1, v_2, v_3, ...v_{n-1}] \tag{3}$$

A linear mapping from one snapshort to another is assumed

$$V_1^N = \{v_1, Av_1, ..., A^{N-1}v_1\} \tag{4}$$

By the linear combination of available data fields, we have a standard Arnoldi iteration problem.

$$Av^{n-1} \approx v^{n-1}S \tag{5}$$

Utility written in C++, allowed transferring the data obtained during the hydrodynamic calculation from the OpenFOAM format into the library format, written in Scala.

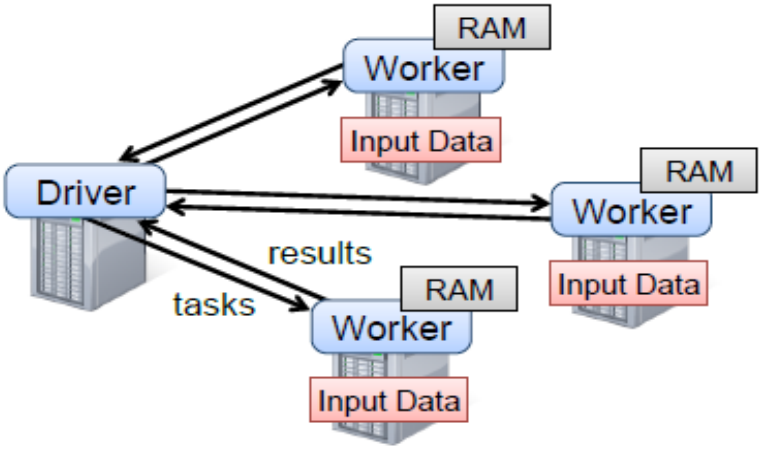


Figure 10: Architecture of Apache Spark technology.

To highlight the characteristic structures, authors propose to use POD (Proper Orthogonal Decomposition) method. For the data processing, 26 snapshots and calculation results of velocity magnitude field were used. As a result of POD and DMD algorithms implementation, the basic dynamic modes, characterizing the large-scale structure in the channel have been revealed. 4 most characteristic modes were selected (Fig.11-14). The arrows on figures display the velocity field taken at t = 0.25 s. Contribution of the other modes to the total turbulence energy is insignificant.

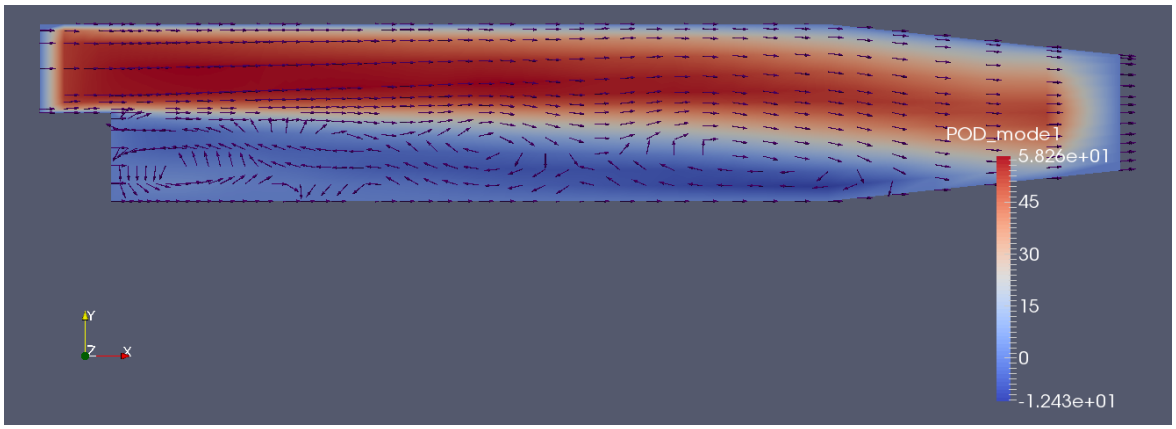


Figure 11: 1-st POD mode.

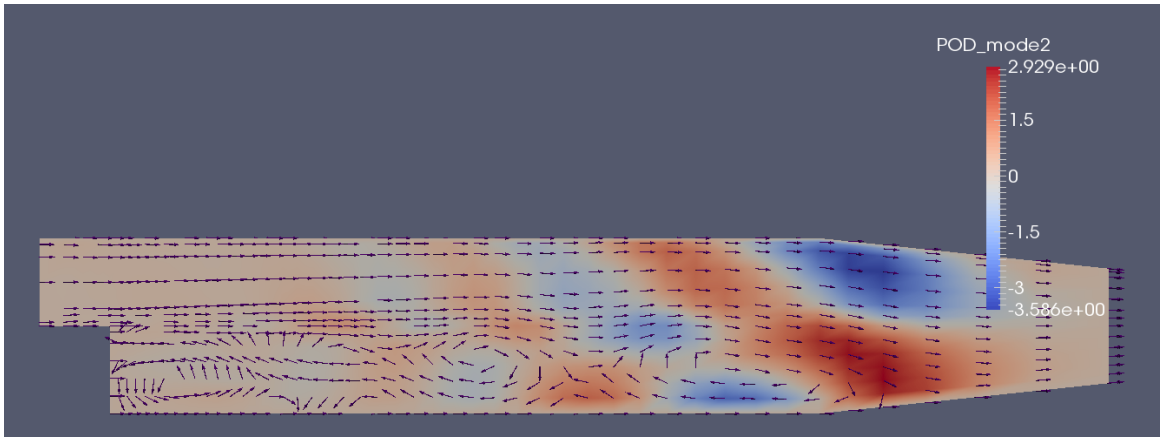


Figure 12: 2-nd POD mode.

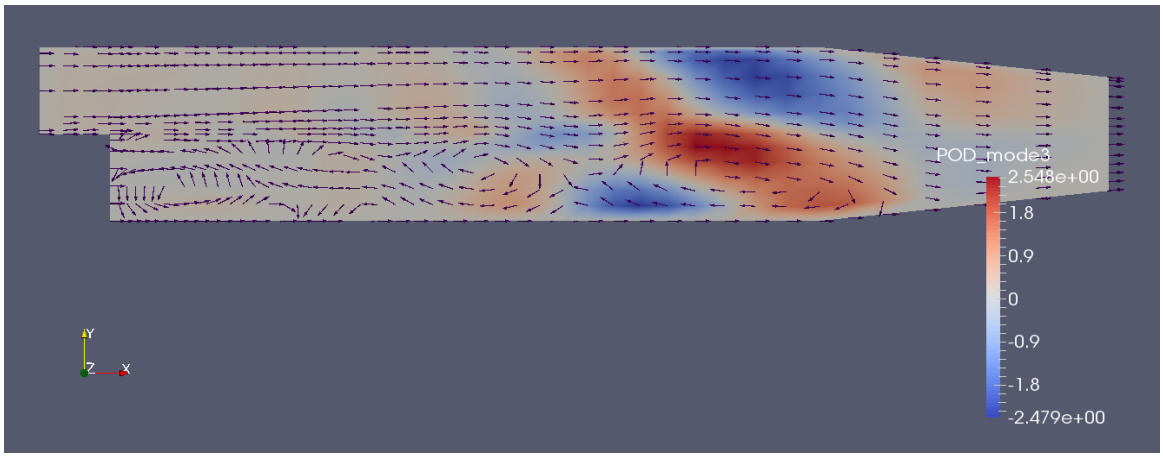


Figure 13: 3-rd POD mode.

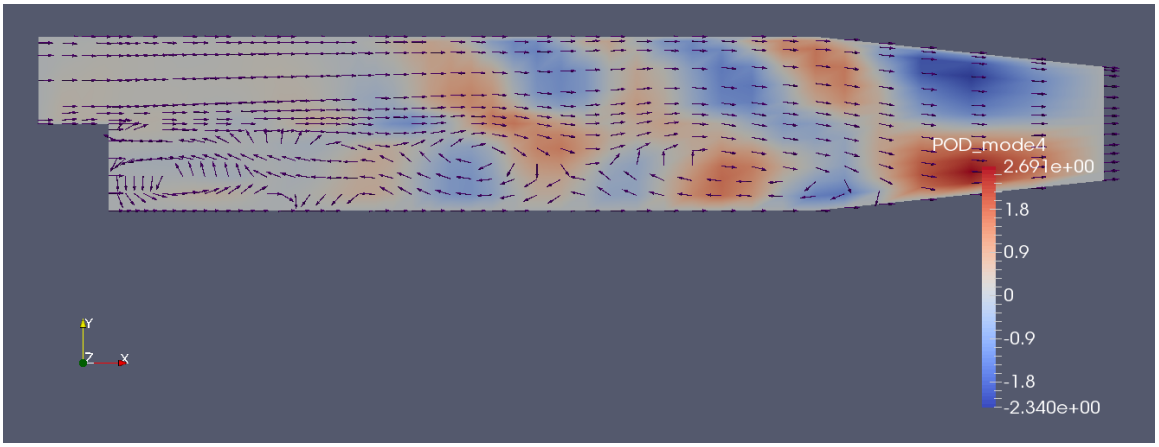


Figure 14: 4-th POD mode.

As a rule, the different modes correspond to the position of different vortices in channel. The four most significant singular values are shown in table 1. This values capture 87% of the kinetic energy.

Number of value	Absolute value	% of captured energy
1	24,14653	80,23%
2	0,897377	83,21%
3	0,716828	85,59%
4	0,690438	87,89%

Table 1: First 4 modes.

5 CONCLUSIONS

The calculation of the flow in the channel was performed, using the new approach in Apache Spark and POD, DMD algorithms, first 4 dynamic modes are selected. Application of this technology allows to solve more complex problems for investigation of turbulence flows and obtaining data on characteristics of coherent structures.

ACKNOWLEDGMENTS

This research is funded by The Ministry of Education and Science of Russia (RFMEFI60714X0090, grant number 14.607.21.0090).

REFERENCES

[1] R.W. Pitz, J.W. Daily. Experimental study of combustion in a turbulent free shear layer formed at a rearward facing step. *AIAA Journal,* **21**, 1565-1570, 1983.

[2] C. Fureby. Homogenization Based LES for Turbulent Combustion. *Flow, Turbulence and Combustion,* **84,** 459-480, 2010.

[3] H.G. Weller, G. Tabor, H. Jasak, C. Fureby. A tensorial approach to computational continuum mechanics using object oriented techniques, *Computers in Physics*, **12**, № 6, 620-631, 1998.

[4] G. Berkooz, P. Holmes, and J. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows, *Annual review of fluid mechanics*, **25,** no. 1, pp. 539–575, 1993.

[5] L. Sirovich. Turbulence and the Dynamics of Coherent Structures: Parts I, II, and III. *Quarterly of Applied Mathematics*, **45**, 561–590, 1987.

[6] P.J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics,* **656,** 5-28, 2010.

[7] P. J. Schmid, L. Li, M. P. Juniper, O. Pust. Applications of the dynamic mode decomposition. *Theor. Comp. Fluid Dyn.,* **25**, 249-259, 2011.

[8] C. Rowley, I. Mezic, S. Bagheri, P. Schlatter, D. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics,* **641**, 115–127, 2009.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL VALIDATION OF A FOUR PARAMETER LOGARITHMIC TURBULENCE MODEL

D. Cerroni[1], R. Da Vià[1*], S. Manservisi[1] and F. Menghini[1]

[1] DIN - Lab. of Montecuccolino, University of Bologna, Via dei Colli 16, 40136 Bologna, Italy
[*] e-mail: roberto.davia2@unibo.it

Keywords: Turbulent heat transfer, Turbulence modeling

Abstract. *Computational Fluid Dynamics codes are used in many industrial applications in order to evaluate interesting physical quantities, such as the heat transfer in turbulent flows. Commercial CFD codes use only turbulence models with an imposed constant turbulent Prandtl number Pr_t, which can give accurate results only for simulations when a strong similarity between the velocity field and the temperature field can be assumed. For fluids with a low Prandtl number, as for heavy liquid metals, a constant turbulent Prandtl number leads to an overestimation of the heat transfer, so experimental results and Direct Numerical Simulation cannot be reproduced. In this work we propose a new k-Ω-k_θ-Ω_θ turbulence model as an improvement of the k-ω-k_θ-ω_θ turbulence model, already validated by the authors, where Ω and Ω_θ are calculated as the natural logarithm of the variables ω and ω_θ. With this reformulation of the previous turbulence model we obtain some important advantages in numerical stability and robustness of the code. Results for the simulations of fully developed turbulent flows in two and three dimensional geometries are reported and compared with experimental correlations and DNS data, when available.*

1 Introduction

The IV generation of nuclear reactors is based on a design that combines increased safety, passive safety systems and low radioactive-waste production. Six new reactors have been proposed by Generation IV International Forum (GIF). Among them the Lead Fast Reactor and the Sodium Fast Reactor are cooled by a liquid metal. One of the main features of these reactors is the capability to produce fissile atoms by the transmutation of Uranium-238. In order to obtain this, a material with high atomic mass, such as liquid metals, has to be employed as core-coolant. Inside a nuclear reactor core the fluid flow is typically turbulent so a turbulence model is needed to simulate the flow with affordable computational cost.

The most common turbulence models currently employed for the study of turbulent flows are based on the Reynolds Average of the Navier Stokes equations (RANS models). In these models the similarity between the velocity and the thermal turbulent fields is usually assumed. An algebraic relation based on the eddy viscosity and the turbulent Prandtl number is used to compute the turbulent heat flux. This turbulence model is known as Simple Eddy Diffusivity (SED) model. In the SED model one computes the eddy viscosity ν_t by using two turbulent variables, such as the turbulent kinetic energy k and its dissipation rate ε. The eddy thermal diffusivity α_t is then computed as $\alpha_t = \nu_t/Pr_t$. This procedure can be used when turbulent flows with a molecular Prandtl number $Pr \simeq 1$ are simulated. Heavy liquid metals such as mercury, Lead-Bismuth Eutectic (LBE) and sodium-potassium alloys are characterized by a low molecular Prandtl number, varying from 0.01 to 0.05. In these fluids heat conduction is an important heat transfer mechanism and the similarity assumption between the velocity and the temperature fields does not hold. It is well known that when the standard SED model is adopted with low Prandtl numbers the obtained values of heat transfer are very different from the experimental ones. This problem can be solved by specifying the Pr_t for each geometry or using a more sophisticated turbulence model in order to define the proper thermal characteristics time scales [1, 2, 3, 4, 5].

In this paper we present a new logarithmic formulation of the four parameter k-ω-k_θ-ω_θ turbulence model for liquid metal flows. We briefly recall the RANS equations and the first formulation of a four parameter k-ε-k_θ-ε_θ, then we derive the k-ω formulation already validated by the authors [6, 7, 8, 9]. Finally the new logarithmic k-Ω-k_θ-Ω_θ model is presented. A discussion on the boundary conditions is reported to better motivate the derivation of this model. We report some numerical results obtained with the implementation of the logarithmic model in a finite element in-house code and compare them with experimental correlations and DNS data, when available.

2 Transport equations and turbulence models

Inside the nuclear reactor core the flow of liquid metal can be considered incompressible so the system of equations to be solved is based on the Reynolds Averaged incompressible Navier Stokes equations, namely

$$\boldsymbol{\nabla}\cdot\mathbf{u} = 0\,, \tag{1}$$

$$\rho\frac{\partial\mathbf{u}}{\partial t}+\rho(\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{u} = \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}-\boldsymbol{\nabla}\cdot\boldsymbol{\tau}^R+\rho\mathbf{g}\,, \tag{2}$$

$$\rho C_p\left(\frac{\partial T}{\partial t}+(\mathbf{u}\cdot\boldsymbol{\nabla})T\right) = \boldsymbol{\nabla}\cdot\mathbf{q}-\boldsymbol{\nabla}\cdot\mathbf{q}^R+Q\,, \tag{3}$$

where $\mathbf{u}$ is the averaged velocity of the fluid and T is the averaged temperature. The tensors $\boldsymbol{\sigma}$ and $\mathbf{q}$ are the usual viscous stress and heat flux and they are modeled using Navier-Stokes constitutive law for viscous fluids and Fourier law for heat conduction

$$\boldsymbol{\sigma} := -p\mathbf{I} + \mu\mathbf{D} \quad \text{with} \quad \mathbf{D} := \boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T \tag{4}$$

$$\mathbf{q} := -\lambda\boldsymbol{\nabla}T\,. \tag{5}$$

The transport equations for the averaged velocity and temperature are similar to the ones for the instantaneous quantities, except for the terms $\boldsymbol{\tau}^R$ and $\mathbf{q}^R$ which are introduced, as a result of the average procedure over time, in the following form

$$\boldsymbol{\tau}^R = \rho\overline{\mathbf{u}'\mathbf{u}'} \qquad \mathbf{q}^R = \rho C_p\overline{\mathbf{u}'T'}\,, \tag{6}$$

where the operator $\overline{(\cdot)}$ means the time average. These terms are known as the Reynolds stresses and the turbulent heat flux. They could be computed by solving the appropriate transport equations, which means a total number of 9 additional unknowns. This procedure is very expensive in terms of computational cost. Instead of solving these additional transport equations we approximate the terms $\overline{\mathbf{u}'\mathbf{u}'}$ and $\overline{\mathbf{u}'T'}$ by using the eddy diffusivities and the gradients of mean velocity and temperature in the following form

$$\boldsymbol{\tau}^R = -\nu_t\left(\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\right) + \frac{2k}{3}\mathbf{I}\,, \tag{7}$$

$$\mathbf{q}^R = -\alpha_t\,\boldsymbol{\nabla}T\,, \tag{8}$$

where the eddy diffusivity of momentum ν_t and the heat eddy diffusivity α_t must be properly defined in the turbulence model. The two eddy diffusivities are calculated as a function of the turbulent kinetic energy and of two characteristic time scales: τ_{lu} for the dynamical and $\tau_{l\theta}$ for the thermal turbulence. These time scales have to be defined appropriately in the turbulence model, for details one can see [3, 4, 5, 7].

2.1 The k-ε-k_θ-ε_θ turbulence model

We can define the turbulent kinetic energy k, its dissipation ε, the temperature fluctuations variance k_θ and its dissipation ε_θ as

$$k = \frac{1}{2}\overline{\mathbf{u}'^2}\,, \qquad \varepsilon = \nu\overline{\|\boldsymbol{\nabla}\mathbf{u}'\|^2}\,, \tag{9}$$

$$k_\theta = \frac{1}{2}\overline{T'^2}\,, \qquad \varepsilon_\theta = \alpha\overline{\|\boldsymbol{\nabla}T'\|^2}\,. \tag{10}$$

The transport equations for the variables in (9) and (10) are obtained by taking appropriate moments of the Navier-Stokes equations for the fluctuating quantities $\mathbf{u}'$ and T'. The equation for k is

$$\frac{\partial k}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})k = \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_k}\right)\boldsymbol{\nabla}k\right] + P_k - \varepsilon\,, \tag{11}$$

where

$$P_k := -\overline{u_i'u_j'}\,\frac{\partial u_i}{\partial x_j} = \frac{\nu_t}{2}\|\boldsymbol{\nabla}\mathbf{u} + \boldsymbol{\nabla}\mathbf{u}^T\|^2\,, \tag{12}$$

is the source term of the transport equation. The equation for ε has this form

$$\frac{\partial \varepsilon}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\varepsilon = \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla}\varepsilon\right] + C_{1\varepsilon}\,\frac{\varepsilon}{k}P_k - C_{2\varepsilon}\frac{\varepsilon^2}{k}f_\varepsilon\,. \tag{13}$$

$C_{1\varepsilon}$	$C_{2\varepsilon}$	C_μ	σ_k	σ_ε
1.5	1.9	0.09	1.4	1.4

Table 1: Values of the model constants for (11) and (13).

The values of the model constants that appear in (11) and (13) are reported in Table 1, while f_ε is a model function defined as

$$f_\varepsilon = (1 - \exp(-0.3226\, R_\delta))^2 \left(1 - 0.3 \exp(-0.0237\, R_t^2)\right) . \tag{14}$$

By defining the characteristic time $\tau_u = k/\varepsilon$ one can model the local dynamical time τ_{lu} in many ways [7, 8, 9].

For the thermal turbulence the two transport equations are

$$\frac{\partial k_\theta}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})k_\theta = \boldsymbol{\nabla}\cdot\left[(\alpha + \frac{\alpha_t}{\sigma_{k_\theta}})\boldsymbol{\nabla}k_\theta\right] + P_\theta - \varepsilon_\theta\,, \tag{15}$$

where

$$P_\theta := -\overline{\mathbf{u}'T'}\cdot\boldsymbol{\nabla}T = \alpha_t\,\|\boldsymbol{\nabla}T\|^2\,, \tag{16}$$

is the source term in the k_θ transport equation. The equation for ε_θ can be written as [7, 11]

$$\frac{\partial \varepsilon_\theta}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\varepsilon_\theta = \boldsymbol{\nabla}\cdot\left[(\alpha + \frac{\alpha_t}{\sigma_{\varepsilon_\theta}})\boldsymbol{\nabla}\varepsilon_\theta\right] + \frac{\varepsilon_\theta}{k_\theta}\left(C_{p1}\,P_\theta - C_{d1}\,\varepsilon_\theta\right) + \frac{\varepsilon_\theta}{k}\left(C_{p2}\,P_k - C_{d2}\,\varepsilon\right). \tag{17}$$

For heavy liquid metals with $Pr \approx 0.025$ we have used the coefficients defined in [3, 4], namely $C_{d1} = 0.9$, $C_{p2} = 0.9$. The coefficient C_{p1} has been set to 0.925 and

$$C_{d2} = 1.9\,(1 - 0.3\,\exp(-0.0237\, R_t^2))(1 - \exp(-0.0308\, R_\delta))^2\,. \tag{18}$$

We define a time scale τ_θ as the ratio between k_θ and ε_θ. For details on the time scales and functions to be used in the definition of the eddy viscosity and diffusivity one can see [5, 7, 8, 9, 10, 11].

2.2 The k-ω-k_θ-ω_θ turbulence model

The results obtained with the k-ε-k_θ-ε_θ turbulence model, in terms of the Nusselt number value for the evaluation of the global heat transfer, are more accurate than the ones obtained with the SED model when compared with the reference correlations for liquid metals [7, 8, 9]. The k-ε model better matches experimental results but shows poor numerical convergence due to the strong coupling of ε-ε_θ with k-k_θ on wall boundaries. In order to avoid this coupling between the state variables we develop a k-ω-k_θ-ω_θ model where ε-ε_θ are replaced by the turbulent kinetic energy specific dissipation ω and the specific dissipation of mean squared temperature fluctuations ω_θ defined as

$$\omega = \frac{\varepsilon}{C_\mu k} \qquad \omega_\theta = \frac{\varepsilon_\theta}{C_\mu k_\theta}\,. \tag{19}$$

$C_{1\varepsilon}$	$C_{2\varepsilon}$	C_μ	σ_k	σ_ε	c_{p1}	c_{p2}	c_{d1}	σ_θ
1.5	1.9	0.09	1.4	1.4	1.025	1.9	1.1	1.4

Table 2: Values of the model constants for (20) - (23).

This new turbulence model is obtained by substituting (19) into (11-13-15-17). This keeps the accuracy of the results and enhances the algorithm stability [6]. The new system of equations is

$$\frac{\partial k}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla} k = \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_k}\right)\nabla k\right] + P_k - C_\mu\, k\,\omega\,, \tag{20}$$

$$\frac{\partial \omega}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}\omega = \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla}\omega\right] + \frac{2}{k}\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla} k\cdot\boldsymbol{\nabla}\omega + + (c_{\varepsilon 1} - 1)\,\frac{\omega}{k} P_k - C_\mu\,(c_{\varepsilon 2} f_\varepsilon - 1)\,\omega^2\,, \tag{21}$$

$$\frac{\partial k_\theta}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla} k_\theta = \boldsymbol{\nabla}\cdot\left[\left(\alpha + \frac{\alpha_t}{\sigma_\theta}\right)\boldsymbol{\nabla} k_\theta\right] + P_\theta - C_\mu\, k_\theta\,\omega_\theta\,, \tag{22}$$

$$\frac{\partial \omega_\theta}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}\omega_\theta = \boldsymbol{\nabla}\cdot\left[\left(\alpha + \frac{\alpha_t}{\sigma_\theta}\right)\boldsymbol{\nabla}\omega_\theta\right] + \frac{2}{k_\theta}\left(\alpha + \frac{\alpha_t}{\sigma_\theta}\right)\boldsymbol{\nabla} k_\theta\cdot\boldsymbol{\nabla}\omega_\theta + + (c_{p1} - 1)\,\frac{\omega_\theta}{k_\theta} P_\theta + c_{p2}\frac{\omega_\theta}{k} P_k - (c_{d1} - 1)\, C_\mu \omega_\theta^2 - c_{d2} C_\mu \omega\omega_\theta\,. \tag{23}$$

The values of the model constants that appear in the system (20) - (23) are reported in Table 2.

2.2.1 Boundary conditions for the turbulence models

Since we do not use wall functions, the exact boundary conditions that must be imposed on wall boundaries can be obtained using a near wall Taylor series expansion for the turbulence variables. These variables are expanded based on their analytical formulation (9-10) and (19) under the constraints imposed by the velocity components on wall boundaries. For k, ε, k_θ and ε_θ we obtain the following asymptotic expressions

$$k_w = \frac{1}{2}\, a\,\delta^2, \qquad \varepsilon_w = \nu\frac{k_w}{\delta^2} = \frac{1}{2}\nu\, a\,, \tag{24}$$

$$k_{\theta w} = \frac{1}{2}\, a_\theta\,\delta^2, \qquad \varepsilon_{\theta w} = \alpha\frac{k_{\theta w}}{\delta^2} = \frac{1}{2}\alpha\, a_\theta\,. \tag{25}$$

From (24) and (25) we see that ε and ε_θ are constant and their values depend on a and a_θ which are related to the velocity and temperature fluctuations that are not known *a priori*. With the change of variables introduced in the k-ω turbulence model the boundary values for are

$$\omega = \frac{\varepsilon}{C_\mu k} = \frac{2\nu}{C_\mu \delta^2}\,, \qquad \omega_\theta = \frac{\varepsilon_\theta}{C_\mu k_\theta} = \frac{2\alpha}{C_\mu \delta^2}\,, \tag{26}$$

so their values depend only on the physical properties of the fluid (ν and α) and on the wall distance (δ^2). For the new variables ω and ω_θ it is possible to set exact Dirichlet b. c., while for k and k_θ we set the following Neumann b. c.

$$\frac{\partial k}{\partial n} = 2\frac{k}{\delta}\,, \qquad \frac{\partial k_\theta}{\partial n} = 2\frac{k_\theta}{\delta}\,, \tag{27}$$

where n indicates the direction normal to the wall surface. The latter condition in (27) can be applied when temperature fluctuations are assumed to be zero at the wall.

2.3 The logarithmic turbulence model k-Ω-k_θ-Ω_θ

The turbulence model we propose here consists in a further improvement of the previous models. The formulation of this new turbulence model is based on [12] in which the various k-ε, k-ω and k-τ models for dynamic turbulence are analyzed in their formulation with logarithmic variables. The introduction of logarithmic variables brings important advantages to the turbulence model, in particular the fact that the original variables are always kept positive because they are calculated as the exponential values of the new logarithmic variables, so this ensures an increased numerical stability. Another important feature is that the logarithmic variables have profiles that are smoother than the ones of the natural variables [12]. In this new turbulence model we use the logarithmic forms of the specific dissipations ω and ω_θ

$$\Omega = \ln(\omega)\,, \qquad \Omega_\theta = \ln(\omega_\theta)\,. \tag{28}$$

For these new variables we can set exact Dirichlet boundary conditions

$$\Omega_w = \ln(\omega_w) = \ln\left(\frac{2\nu}{C_\mu}\right) - 2\ln(\delta)\,, \tag{29}$$

$$\Omega_{\theta w} = \ln(\omega_{\theta w}) = \ln\left(\frac{2\alpha}{C_\mu}\right) - 2\ln(\delta)\,. \tag{30}$$

The new system of equations is

$$\frac{\partial k}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}k = \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_k}\right)\boldsymbol{\nabla}k\right] + P_k - C_\mu\, k\, e^{\Omega}\,, \tag{31}$$

$$\begin{aligned}\frac{\partial \Omega}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}\Omega = {}& \boldsymbol{\nabla}\cdot\left[\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla}\Omega\right] + \frac{2}{k}\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla}k\cdot\boldsymbol{\nabla}\Omega + \\ & + \left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\boldsymbol{\nabla}\Omega\cdot\boldsymbol{\nabla}\Omega + \frac{c_{\varepsilon 1}-1}{k}P_k - C_\mu\left(c_{\varepsilon 2} f_{exp} - 1\right)e^{\Omega}\,,\end{aligned} \tag{32}$$

$$\frac{\partial k_\theta}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}k_\theta = \boldsymbol{\nabla}\cdot\left[\left(\alpha + \frac{\alpha_t}{\sigma_\theta}\right)\boldsymbol{\nabla}k_\theta\right] + P_\theta - C_\mu\, k_\theta\, e^{\Omega_\theta}\,, \tag{33}$$

$$\begin{aligned}\frac{\partial \Omega_\theta}{\partial t} + \mathbf{u}\cdot\boldsymbol{\nabla}\Omega_\theta = {}& \boldsymbol{\nabla}\cdot\left[\left(\alpha + \frac{\alpha_t}{\sigma_{\varepsilon_\theta}}\right)\boldsymbol{\nabla}\Omega_\theta\right] + \frac{2}{k_\theta}\left(\alpha + \frac{\alpha_t}{\sigma_{\varepsilon_\theta}}\right)\boldsymbol{\nabla}k_\theta\cdot\boldsymbol{\nabla}\Omega_\theta + \\ & + \left(\alpha + \frac{\alpha_t}{\sigma_{\varepsilon_\theta}}\right)\boldsymbol{\nabla}\Omega_\theta\cdot\boldsymbol{\nabla}\Omega_\theta + \frac{c_{p1}-1}{k_\theta}P_\theta + \frac{c_{p2}}{k}P_k - \left(c_{d1}-1\right)e^{\Omega_\theta} - c_{d2}C_\mu e^{\Omega}\,.\end{aligned} \tag{34}$$

The eddy viscosity ν_t and the eddy thermal diffusivity α_t are modeled as

$$\nu_t = C_\mu k \tau_{lu}\,, \qquad \alpha_t = C_\theta k \tau_{l\theta}\,. \tag{35}$$

For the modeling of the characteristic time scales for dynamic and thermal turbulence one can see [7, 8, 9]. With this turbulence model one can correctly reproduce the near wall behavior of the turbulence variables, which means $k \propto \delta^2$, $\nu_t \propto \delta^3$, $k_\theta \propto \delta^2$ assuming zero temperature fluctuations, and $\alpha_t \propto \delta^3$.

3 Numerical results

In this section we present the results obtained from the simulations of fully developed turbulent flows of liquid metals in several geometries. The plane channel and cylindrical duct geometries are investigated for different values of Reynolds numbers. A complex 19-pin hexagonal

rod bundle is simulated for a specific pitch-to-diameter ratio and Reynolds number. In Table 3 the values of the physical properties are reported. In particular those values are representative of Lead-Bismuth-Eutectic (LBE) which has a molecular Prandtl number $Pr = 0.025$.

We implement the system (31-34) and the Navier-Stokes equations in an in-house finite element code. We employ Taylor-Hood finite elements for the system of Navier-Stokes in order to satisfy the inf-sup condition and this system is solved with a standard projection method. The two systems of turbulence equations are solved with standard quadratic finite elements.

3.1 Plane channel

The plane channel is one of the simplest type of geometry that can be simulated and many Direct Numerical Simulation are available for this kind of flow. In particular for fully developed turbulent flows results of the dynamical turbulence for the cases of friction Reynolds number $Re_\tau = 180$, 395, 640, 950, 2000 and 4200 are available [13, 14, 15], while results of fully developed thermal turbulent flow for a fluid with $Pr = 0.025$ are available only for the cases of friction Reynolds number $Re_\tau = 180$, 395 and 640 [13]. The domain consists of two plates located at a distance $L = 0.0605$ m, with infinite dimensions in the other directions. On the wall a uniform heat flux of 3.6×10^5 W/m^2 is applied. The condition of fully developed turbulent flow is imposed with periodic boundary conditions on the inlet and outlet of the channel.

We simulate seven test cases with Reynolds number $Re \simeq 5700$ (A), 14000 (B), 24000 (C), 37000 (D), 86000 (E), 198000 (F) and 325000 (G). They correspond to friction Reynolds numbers $Re_\tau \simeq 180$, 395, 640, 950, 2000, 4200 and 6600. All simulations are performed using a spatial discretization of the physical domain in order to have the first mesh point at $y^+ < 1$.

Property	Symbol	Value	Unit
Viscosity	μ	0.00184	Pa s
Density	ρ	10340	Kg/m^3
Thermal conductivity	λ	10.72	W/(m K)
Heat specific capacity	C_p	145.75	J/(Kg K)

Table 3: Physical parameters used in the CFD simulations.

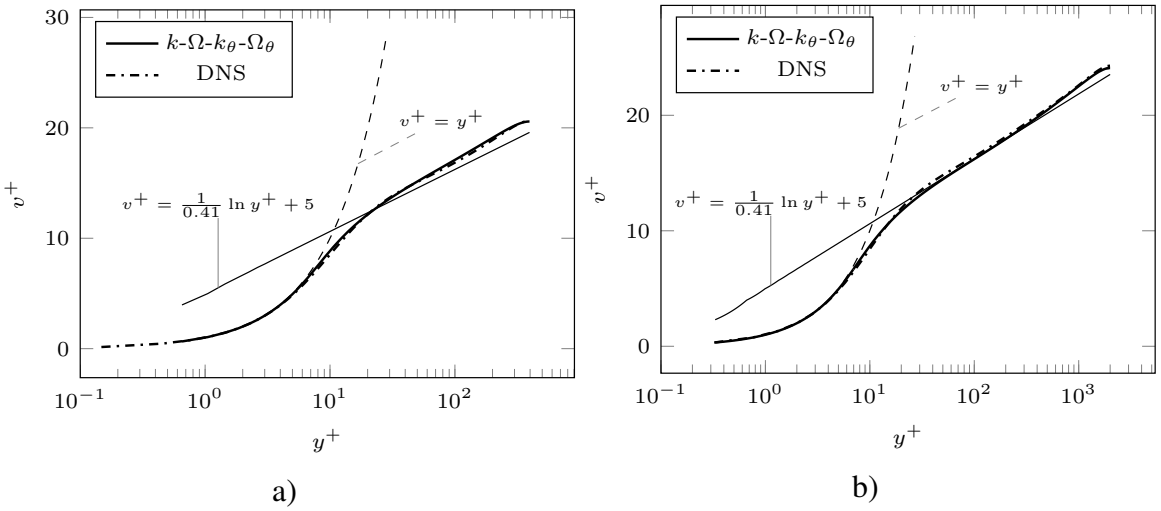


Figure 1: Non dimensional velocity profiles as a function of the non dimensional distance from the wall y^+. Cases $Re_\tau = 395$ (a) and $Re_\tau = 2000$ (b) obtained with the k-Ω turbulence model (continuous) and from DNS simulation data (dash-dotted).

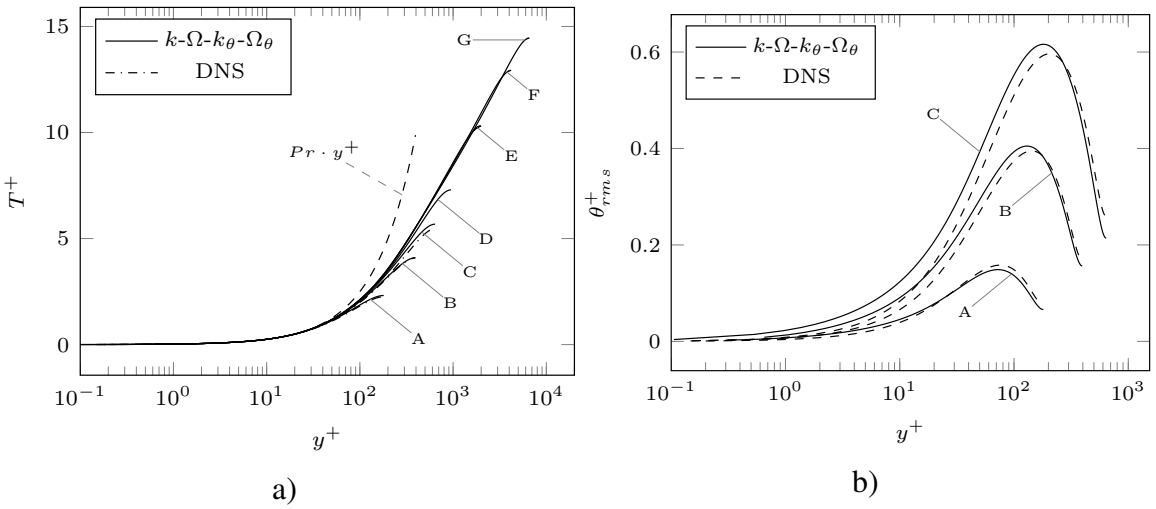


Figure 2: Profile of the non dimensional temperature T^+ (a) and of the root mean squared temperature fluctuations θ_{rms} (b) as a function of the non dimensional distance from the wall y^+. DNS results are reported for the cases $Re_\tau = 180$, 395 and 640 (dash-dotted).

In Figure 1 a comparison is shown between the results obtained with the k-Ω turbulence model and those obtained with DNS for the cases of $Re_\tau = 395$ and $Re_\tau = 2000$. In this Figure the velocity v has been non dimensionalized with the friction velocity v_τ which is calculated as $v_\tau = \sqrt{\tau_w/\rho}$, where τ_w is the wall shear stress, and plotted against the non dimensional wall distance $y^+ = yv_\tau/\nu$. The linear behavior $v^+ = y^+$ and the logarithmic one $v^+ = 1/0.41\ln(y^+) + 5$ are well reproduced in both cases.

The profiles of the non dimensional temperature are shown in Figure 2 (a), together with the results of DNS simulations for the cases where they are available. The temperature has been non dimensionalized with the friction temperature T_τ defined as $T_\tau = q/(v_\tau \rho C_p)$, where q is the heat flux applied on the wall. The linear behavior $T^+ = Pr\, y^+$ is well reproduced for every single case. In Figure 2 (b) the values of root mean squared temperature fluctuations $\theta_{rms} = \sqrt{2k_\theta}$ are compared with the ones obtained with DNS simulations. We can see that there is a good agreement between the results of the turbulence model and Direct Numerical Simulation data.

3.2 Cylindrical pipe

For the cylindrical pipe the DNS data for the dynamical turbulence are available for the cases $Re_\tau = 180$, 360, 550, and 1000 [16]. The results are thus validated using the experimental correlations for the Nusselt number. In particular we refer to the Kirillov correlation as the main reference for the cylindrical geometry [17]. This correlation is claimed to be valid for $10^4 < Re < 5 \cdot 10^6$. The Nusselt number is calculated as a function of the Peclet number $Pe = Pr \cdot Re$

$$Nu = 4.5 + 0.018\, Pe^{0.8}\,. \tag{36}$$

For the cylindrical case we simulate fully developed turbulent flows with $Re_\tau = 180$ (A), 360 (B), 550 (C), 1000 (D), 3580 (E), 5840 (F) and 6860 (G). The corresponding values of the Reynolds number for these cases are 5760, 12770, 20700, 41000, 165000, 286000 and 341000. As in the plane channel case, all the simulations for the cylindrical pipe are performed refining the mesh in order to have the first mesh point with $y^+ < 1$. In Figure 3 (a) the profiles of the non dimensional temperature T^+ are shown as functions of the non dimensional distance

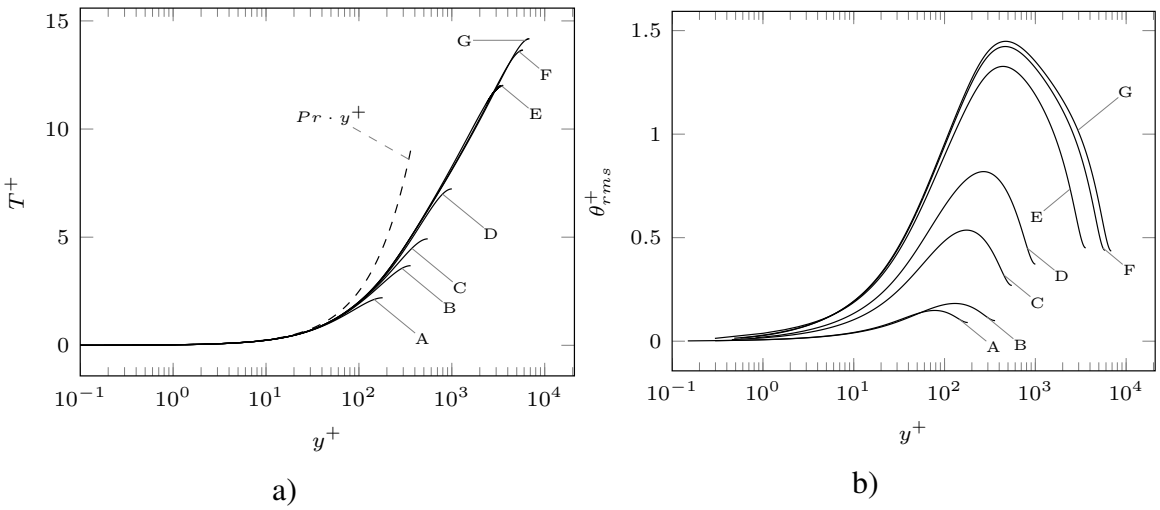


Figure 3: Non dimensional profiles of: mean temperature (a) and root mean squared values of temperature fluctuations (b) as functions of the non dimensional distance from the wall y^+ for all the simulated cases.

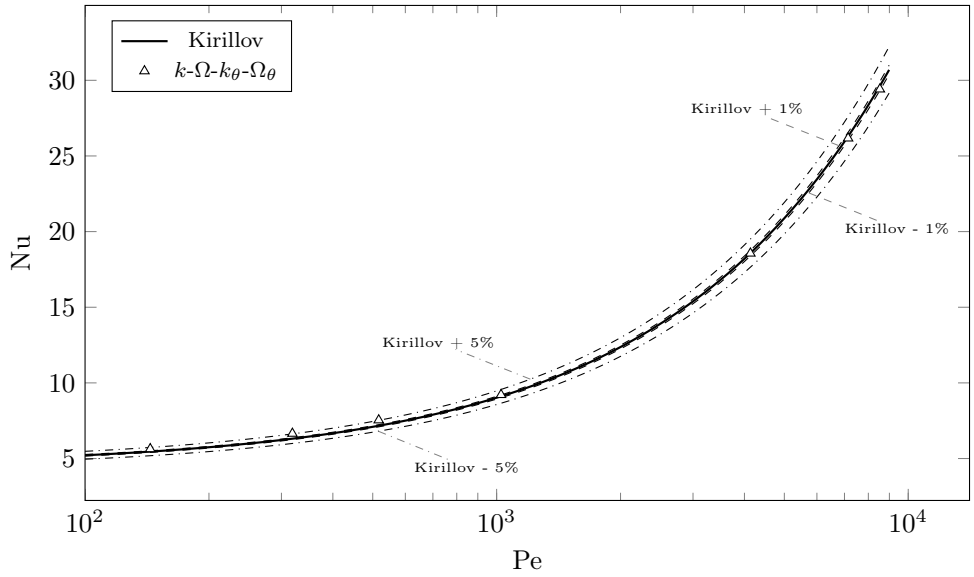


Figure 4: Comparison between the Kirillov correlation for the Nusselt number and the values of Nu calculated with the four parameters turbulence model.

from the wall y^+. In every case the linear behavior $T^+ = Pr\, y^+$ is well reproduced, as can be seen from the comparison with the dashed line. In Figure 3 (b) the non dimensional root mean squared values of temperature fluctuations $\theta^+_{rms} = \sqrt{2k_\theta}/T_\tau$ are presented for all the simulated cases.

Finally we calculate the Nusselt number values and compare them with the Kirillov correlation. This comparison is presented in Figure 4 where the values of the Nusselt number are plotted against the Peclet number. As it can be seen some differences between the correlation data and the values calculated with the turbulence model appear in the low velocity region (Pe $< 10^3$). The differences with the Kirillov correlation, in this region, are included within 5%. In the high velocity region with Pe $> 10^3$ our values are closer to the Kirillov correlation. As shown in the Figure the values are included within 1%.

3.3 Hexagonal rod bundle

Property	Symbol	Value	Unit
Number of pins	n	19	-
Pin diameter	d	0.0082	m
Pitch to diameter ratio	p/d	1.3	-
Height	z	2	m
Minimum pin-wall distance	l	0.00172	m
Hydraulic diameter	D_h	7.706×10^{-3}	m

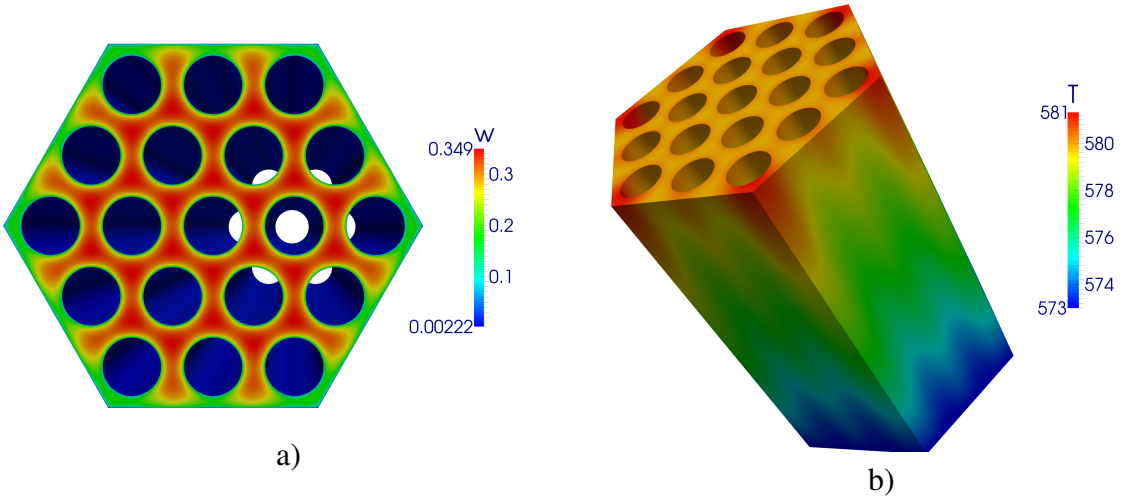
Table 4: Geometrical parameters of the simulated hexagonal bundle.

In this Section we present the results of a simulation of a 19-pins hexagonal rod bundle. In Figures 5 (a) and (b) a geometrical representation of the whole bundle is given. The bundle is bounded by adiabatic solid walls. This particular configuration is used for experimental analysis [18]. Due to the symmetry of the geometry and boundary conditions it is possible to simulate only one twelfth of the entire bundle.

A transverse section of the computational domain can be seen in Figures 6-7 while the geometrical parameters of the bundle are reported in Table 4. In the simulated case the velocity and the dynamical turbulence are fully developed, so we impose periodic boundary conditions at the inlet and outlet of the bundle. On the wall boundaries we set

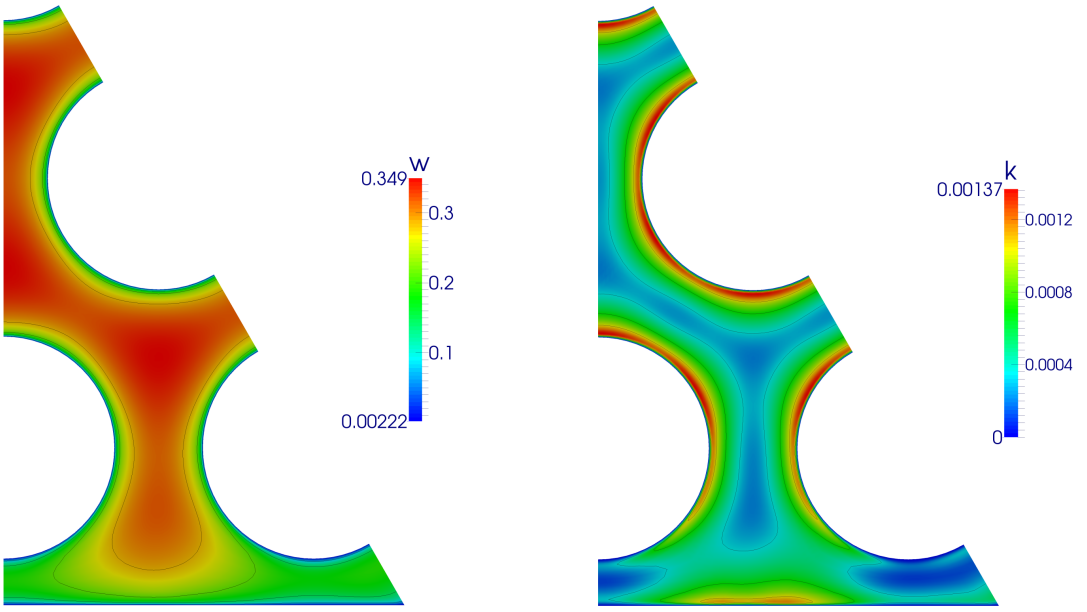
$$u = v = 0\,, \quad \frac{\partial w}{\partial n} = \frac{\mu}{\delta}\,, \quad \frac{\partial k}{\partial n} = 2\frac{k}{\delta}\,, \quad \Omega = \ln\left(\frac{2\nu}{C_\mu \delta^2}\right), \tag{37}$$

where n indicates the direction normal to the wall and δ the wall distance. On the other boundaries we impose a symmetry condition. For the thermal field we impose a fixed uniform temperature of 573 K at the inlet of the bundle, so k_θ is set to zero and Ω_θ is set to the same value that it has on the wall boundaries. On the pin walls we impose a uniform heat flux of 3600 W/m^2, while the external wall is considered adiabatic, $\partial T/\partial n = 0$. For the thermal turbulent variables



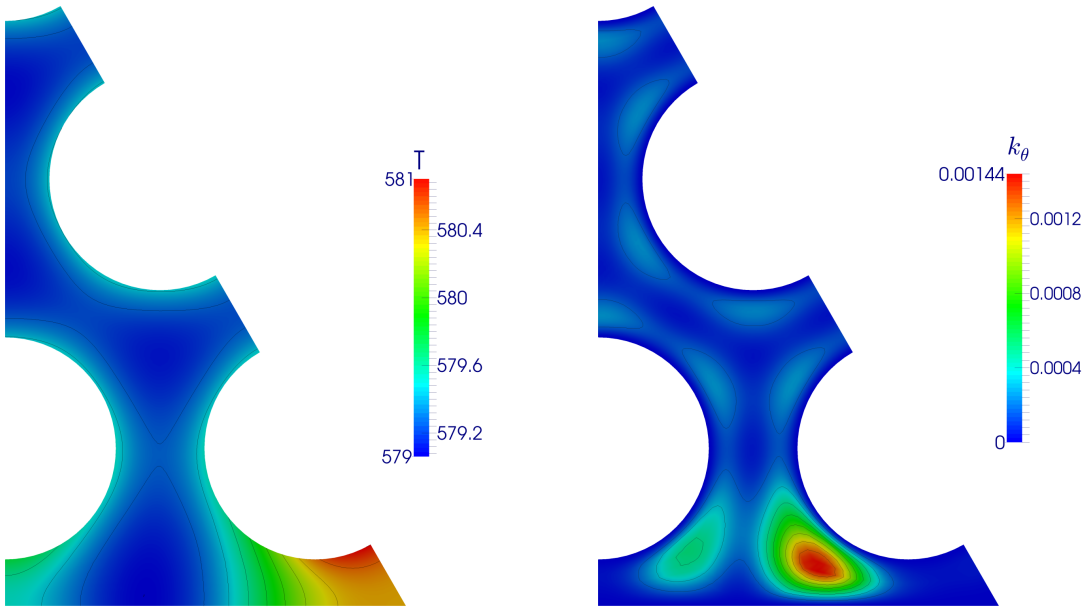
a) b)

Figure 5: Front view of the hexagonal bundle with the velocity field (a) and a scaled view of the whole hexagonal bundle with the thermal field (b).



a) b)

Figure 6: Distribution of the axial component w of the velocity field (a) and of the turbulent kinetic energy k (b) on a transverse section of the bundle.



a) b)

Figure 7: Distribution of the temperature T field (a) and of the mean squared temperature fluctuations k_θ (b) on a transverse section of the bundle.

on every wall we impose the following boundary conditions

$$\frac{\partial k_\theta}{n} = 2\frac{k_\theta}{\delta}\,, \quad \Omega_\theta = \ln\left(\frac{2\alpha}{C_\mu \delta^2}\right). \tag{38}$$

At the outlet of the bundle we consider that no heat exchange occurs, so we impose a null derivative in the direction normal to the outlet surface for T and for k_θ and Ω_θ. The physical properties of the fluid are the same shown in Table 3. The mean axial velocity is $w_{mean} = 0.23$

m/s, so the Reynolds number is $Re = 10070$.

In Figure 5 (a) the axial component of the velocity field is shown at the bundle outlet while in Figure 5 (b) the temperature field is presented in a scaled view of the whole bundle. In Figure 6 (a) the axial velocity is reported on a transverse section of the bundle. In the region near the adiabatic wall the velocity is smaller than in the remaining part of the domain. Due to this phenomenon we have some areas characterized by a low turbulence in the same region, as it can be seen from Figure 6 (b), where the field of turbulent kinetic energy is shown. We recall that we simulate the condition of fully developed turbulent flow, so the distributions shown in Figure 6 (a) and in Figure 6 (b) are the same for every transverse section along the axial coordinate z.

The temperature distribution for the transverse section at $z = 1.8$ m is shown in Figure 7 (a). As we can see the temperature distribution is almost uniform except in the region near the adiabatic wall where a stagnation point appears. In particular the temperature distribution reaches its maximum value in the corner region, as it can be better seen from Figure 5 (b). In the region near the adiabatic wall the temperature changes sharply. This fact reflects the distribution of the mean squared value of the temperature fluctuations, which is reported in Figure 7 (b). The transverse section is at the same height as Figure 7 (a).

4 Conclusion

In this work we have proposed a new four parameter turbulence model with logarithmic variables. The model has been tested in simulations of fully developed turbulent flows of low Prandtl number fluids ($Pr = 0.025$) and compared with reference results available in literature. For the case of plane channel, where more DNS data are available for both dynamical and thermal turbulence, the results of the new four parameter turbulence model show a good agreement with DNS results. For the case of cylindrical pipe we focused our attention on the values of the Nusselt number obtained for all the simulated cases because a few DNS data for the thermal field are available in literature. The obtained results are very close to the Kirillov correlation, which has been taken as the main reference for the Nusselt number prediction. Finally some preliminary results of the application of the turbulence model on a complex three-dimensional geometry have been reported, showing the robustness of the model also in thermally developing three-dimensional flows. The new turbulence model can thus be used as a reliable tool for the study of turbulent flows characterized by low Prandtl numbers.

REFERENCES

[1] X. Cheng and N. Tak, *Investigation on turbulent heat transfer to lead-bismuth eutectic flows in circular tubes for nuclear applications*, Nuclear Engineering and Design, Vol.236, pp. 385-393, 2006.

[2] X. Cheng and N.I. Tak, *CFD analysis of thermal-hydraulic behavior of heavy liquid metals in sub-channels*, Nuclear Engineering and Design, Vol. 236, pp. 1874-1885, 2006.

[3] Y. Nagano and M. Shimada, *Development of a two equation heat transfer model based on direct simulations of turbulent flows with different Prandtl numbers*, Physics of Fluids, Vol. 8, pp. 3379-3402, 1996.

[4] K. Abe, T. Kondoh and Y. Nagano, *A new turbulence model for predicting fluid flow and heat transfer in separating and reattaching flows II. Thermal field calculations*, International Journal of Heat Mass Transfer, Vol. 38 (8), pp. 1467-1481, 1995.

[5] B. E. Launder, G. J. Reece and W. Rodi, *Progress in the development of a Reynolds-stress turbulence closure*, Journal of Fluid Mechanics, Vol. 37 (3), pp. 537-566, 1975.

[6] D. Cerroni, R. Da Vià, S. Manservisi, F. Menghini, G. Pozzetti and R. Scardovelli, *Numerical validation of a k-ω-k_θ-ω_θ heat transfer turbulence model for heavy liquid metals*, Journal of Physics: Conference Series, Vol. 655, 012046, 2015.

[7] S. Manservisi and F. Menghini, *A CFD Four Parameter Heat Transfer Turbulence Model for Engineering Applications in Heavy Liquid Metals in rod bundle geometries*, International Journal of Heat and Mass Transfer, Vol. 69, pp. 312-326, 2014.

[8] S. Manservisi and F. Menghini, *Triangular rod bundle simulations of a CFD k-ϵ-k_θ-ϵ_θ Heat Transfer Turbulence Model for Heavy Liquid Metals*, Nuclear Engineering and Design, Vol. 273, pp. 251-270, 2014.

[9] S. Manservisi and F. Menghini, *CFD simulations in heavy liquid metal flows for square lattice bare rod bundle geometries with a four parameter heat transfer turbulence model*, Nuclear Engineering and Design, Vol. 295, pp. 251-260, 2015.

[10] K. Abe, T. Kondoh and Y. Nagano, *A two-equation heat transfer model reflecting second-moment closures for wall and free turbulent flows*, International Journal of Heat and Fluid Flow, Vol. 17, pp. 228-237, 1996.

[11] H. Hattori, Y. Nagano and M. Tagawa, *Analysis of turbulent heat transfer under various thermal conditions with two-equation models*, Engineering Turbulence Modeling and Experiments 2 (Edited by W. Rodi and F. Martelli), pp. 43-52, 1993.

[12] F. Ilinca, J. F. Hetu, and D. Pelletier, *A unified finite element algorithm for two-equation models of turbulence*, Computers & Fluids, Vol. 27 (3), pp. 291-310, 1998.

[13] H Kawamura, H Abe and Y Matsuo, *DNS of turbulent heat transfer in channel flow with respect to Reynolds and Prandtl number effects.*, International Journal of Heat and Fluid Flow, Vol. 20, pp. 196-207, 1999.

[14] S. Hoyas and J. Jiménez, *Reynolds number effects on the reynolds-stress budgets in turbulent channels*, Physics of Fluids, Vol. 20 (10), 101511, 2008.

[15] A. Lozano-Durn and J. Jiménez, *Effect of the computational domain on direct simulations of turbulent channels up to re_τ = 4200*, Physics of Fluids, Vol. 26 (1), 011702, 2014.

[16] G.K. El Khoury, P. Schlatter, A. Noorani, P.F. Fischer, G. Brethouwer and A.V. Johansson, *Direct numerical simulation of turbulent pipe flow at moderately high Reynolds numbers*, Flow, Turbulence and Combustion, Vol. 91 (3), pp. 475-495, 2013.

[17] P. Kirillov and P.A. Ushakov, *Heat transfer to liquid metals: specific features, methods of investigation, and main relationships*, Thermal Engineering, Vol. 48 (1), pp. 50-59, 2001.

[18] J. Pacio, K. Litfin, A. Batta, M. Viellieber, A. Class, H. Doolaard, F. Roelofs, S. Manservisi, F. Menghini and M. Böttcher, *Heat transfer to liquid metals in a hexagonal rod bundle with grid spacers: Experimental and simulation results*, Nuclear Engineering and Design, Vol. 290, pp. 27-39, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

LARGE-EDDY SIMULATION AND FAR FIELD ACOUSTICS OF A SUBSONIC HOT JET

O. Labbé

ONERA - The French Aerospace Lab
F-92322 Châtillon, France
odile.labbe@onera.fr

Keywords: LES, Hot Turbulent Jet, Ffowcs Williams and Hawkings.

Abstract. *The present work deals with the numerical simulation of a hot subsonic turbulent jet using LES. The main difficulty is to obtain a turbulent jet flow right from the exhaust of the nozzle, as in real jets for Reynolds numbers over 10^5. Injecting disturbances in the boundary layer inside the nozzle to force the transition from a laminar state to a turbulent one, may introduce spurious noise which can pollute the acoustic field. Experimental measurements on mean velocity and near and far field pressure were carried out in the Onera CEPRA19 anechoic wind tunnel on a nozzle of diameter D= 80 mm, called Φ80. These measurements constitute a database to validate numerical simulations. The aerodynamic solver is based on the compressible Navier–Stokes equations expressed in conservative form with a spatial discretization method based on the cell-centered Finite Volume methodology on structured grid. The influence of mesh refinement is studied, thanks to a comparison between a fine and very fine mesh including respectively 30 and 240 millions elements on the same jet configuration. The noise radiation is performed using the Ffowcs Williams and Hawkings surface formulation, that computes time pressure histories at any observer location by integration of the flow field solution on a control surface surrounding the jet and containing all the noise sources. The near and far fields are compared to experimental measurements. Discrepancies remain close to the nozzle exit which lead to an overestimation of the pressure levels in both near and far fields, especially near the 90◦ angular sector.*

1 INTRODUCTION

With the increase of computational capacities in the last decades, it became possible to perform unsteady flow simulations to reproduce and investigate noise generation [1]. If direct numerical simulations (DNS) are still limited to the simulation of low Reynolds, academic configurations [2][3], the use of large-eddy simulations (LES) allows to compute higher Reynolds number, turbulent flows [4]-[8] more representative of the industrial concerns. However, performing a Large Eddy Simulation (LES) of a turbulent jet able to reproduce the experimental behavior is still a challenge despite the progresses made over the recent years. One of the difficulties is to obtain a turbulent jet flow right from the exhaust of the nozzle, as in real jets [9] for Reynolds numbers over 10^5. A very fine resolution of the boundary layers inside the nozzle is then necessary, which leads to prohibitive computational resources for experimental boundary layers thickness. So, the numerical jets tend to be laminar at the nozzle exit and as a consequence, strong vortex pairings appear in the shear layer, what usually does not occur in the experiments at higher Reynolds numbers. An additional noise source is then observed in the numerical far-field pressure spectra in the medium frequencies when compared to the measured ones [10]-[13]. Downstream of the laminar part of the flow, the transition towards the turbulent regime is then abrupt, as seen through shear layer properties (faster spreading rate), leading to an underestimation of the potential core length. A key issue of the success of such simulations is to provide boundary conditions representative of the anechoic facilities where jet noise experiments are performed. Several approaches have been developed to allow the turbulent transition of the jets. It is a difficult problem because the jet forcing must generate a minimum spurious noise, in order to avoid to contaminate the acoustic field. The jet development is strongly linked to the initial turbulent or laminar shear layer state, as observed by Zaman [9], the latter leading to additional noise. Non-reflecting conditions are required at lateral and outflow boundaries to model the free-field conditions and to avoid spurious noise contamination of the physical sound field [14]. For a detailed review of available numerical boundary conditions for aeroacoustic simulations, the reader may refer to Tam [15].

Several authors such as Ukeiley and Ponton [16], Suzuki and Colonius [17], Muller *et al.* [19], Fayard *et al.* [20], or Hall *et al.* [21] highlighted the importance of the low order azimuthal modes particularly for position close to the potential core end. Previously, Lorteau *et al.* [22] performed an analysis of the near field pressure from an experimental database of an isothermal and a hot subsonic jet. They established the existence of a sector linked to an acoustic behaviour in the near field in which the low frequencies around St = 0.2 and the axisymmetric mode are dominant.

In the present work, a numerical simulation using the LES methodology is carried out on a hot subsonic jet in order to analyze its near pressure field and its links to the jet flow and to the acoustic far field. The influence of mesh resolution on jet flow development and noise radiation is investigated using LES combined with the Ffowcs Williams & Hawkings surface integral formulation [23] to simulate the far-field noise. No turbulence seeding is made in the boundary layers. Several experimental [19][24] and numerical investigations [10][18]-[20][22][25][26] in the past were carried out on this jet configuration and experimental data are available for comparison with simulation results. The paper is organized as follows. Details of the simulated configuration are given in section 2, where the geometric configuration, numerical parameters for flow and noise simulations and the computational grids are also described. Unsteady and steady flow fields and radiated noise are then presented in Section 3 through comparison with experimental data acquired on the same jet configuration. Mean flow and turbulence evolutions are especially detailed upstream and downstream of the nozzle. Several effects of the meshing on the aerodynamic and acoustic fields are also presented in

this section. Near and far fields are analyzed and different control surfaces are used for the FW-H formulation and compared.

2 JET CONFIGURATION

The geometry considered is a round nozzle with an exhaust diameter D of 80 mm. The jet static temperature is T_j = 830 K and its Mach number is M_j = 0.7, which corresponds to a jet velocity U_j of 410 m/s. The Reynolds number based on nozzle diameter is $Re_D = U_j D/\nu = 4 \times 10^5$, where ν for the kinematic viscosity. The ambient pressure and temperature are p_0 = 101,325 Pa and T_0 = 280 K, respectively. Aerodynamic and acoustic measurements, performed in the CEPRA19 anechoic facility of Onera [19] [24] are available for comparisons with simulations. The characteristics of this single jet are provided in Table 1. They lead to a Reynolds number of 400,000 based on the nozzle diameter and the jet velocity. Because of the high temperature of the jet, the jet velocity U_j at the nozzle exit is lower than the local sound velocity c_j but higher than the sound velocity c_∞ outside the jet. So, the local Mach number $M_j=U_j/c_j$ is subsonic whereas the convective Mach number $M_c=(U_j+c_j)/(c_j+c_\infty)$ is supersonic. The following table summarizes the jet characteristics with subscripts j, tot and ∞, respectively, representing jet exit conditions, jet total stagnation characteristics and ambient conditions.

U_j (m.s−1)	M_j	T_j/T_∞	U_j/c_∞	T_{tot}/T_∞	T_∞ (K)	Re_D	p_{tot}/p_∞
410	0.7	2.96	1.2	3.2	280	4×10^5	1.4

Table 1: jet characteristics with subscripts j, tot and ∞, respectively, representing jet exit conditions, jet total stagnation characteristics and ambient conditions.

The computational domain is cylindrical and extends from X/D = -29 to 100 and has a radius of 80D. An illustration of the nozzle with a detail of the nozzle exit is given in Figure 1. Inside the nozzle, the inflow condition is located 7.15 diameters upstream of the exhaust. A first contraction is present for -6 < X/D < -5.5, followed by a slowly converging shape up to the nozzle exit where a second contraction is visible (-0.16 X/D). The nozzle fairing starts with a radius of 2.5 diameters at X/D =-29 to -18.5 and converges up to X/D = -13 with an angle of 4.3◦ and then up to X/D = -0.5 with an angle of 6.9° . At the exhaust, an angle of 16.4° is visible.

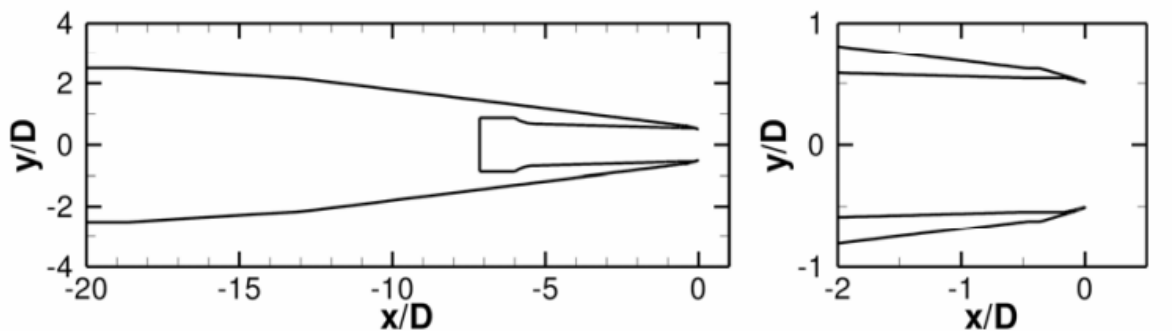


Figure 1: Illustrations of the nozzle shape.

2.1 Grid parameters

Two meshes have been used in this study. The first one called 'Fine' is issued from Huet [10] and consists in $30x10^6$ cells, while the second one called 'Very Fine' contains 240 $x10^6$

cells and was provided by Lorteau *et al.* [22]. The grid of both meshes is axi-symmetrical and structured and includes a O-type treatment on the jet axis to ensure that hexahedral cells have an homogeneous size in the core region. A refined zone which extends downstream of the nozzle exit from (X/D = 0 ; r/D = 2) to (X/D = 25 ; r/D = 5) and upstream of the nozzle exit up to (X/D = -2 ; r/D = 2.4) has been imposed for a good resolution of the jet flow development. Bogey *et al.* [12] tested different azimuthal discretizations and showed the importance of this grid parameter on the shear layer development and jet properties: the 'Very Fine' grid has four times more cells (480) in the azimuthal direction than the 'Fine' grid (120). The grid is stretched from a refined zone, in which the flow is accurately calculated, to the simulated domain boundaries in order to damp acoustic waves before they reach the borders and thus avoid spurious reflections. The details of mesh refinement can be found in the respective papers [10] [22]. The main characteristics of both grids are summarized in Table 2.

Grid name	$\Delta x/D$ (%)	$\Delta r_j/D$ (%)	$r_j.\Delta\theta/D$(%)	St_{cut}	Number of Cells$\times 10^6$	$T.U_j/D$
Fine	0.25	0.25	2.6	0.5	30	512
Very Fine	0.06	0.03	0.6	1	240	449

Table 2: Characteristics of both grids, $\Delta x/D$ and $\Delta r_j/D$ at X/D = 0 and $r_j.\Delta\theta/D$ at r_j = D /2, the number of cells ($\times 10^6$), and non-dimensional simulation time.

2.2 Numerical methods and boundary conditions

The aerodynamic solver FUNk [27] used in this study, is based on the compressible Navier–Stokes equations expressed in conservative form. The spatial discretization method is based on the cell-centered Finite Volume methodology (FVM) on structured grid. An upwind biased scheme, with a third-order MUSCL interpolation scheme of AUSM+(P) family is used for the convective terms. A second-order-accurate centered scheme is used for viscous fluxes. The time integration is carried out by means of a third-order compact Runge–Kutta scheme. Two approaches are commonly used to perform LES with classical FV methods: explicit subgrid stress models are used to represent the effect of the unresolved scales of motion on the large scales and the second method consists in using the Monotonic Integrated Large-Eddy Simulations (MILES) proposed by Boris *et al.* [28]. The latter is based on the assumption that the intrinsic dissipation of an upwind scheme is able to mimic the dissipative behavior of the unresolved turbulent scales, and that when using such a scheme, no subgrid model is needed. This approach has been used in the present study.

At the entrance of the nozzle, uniform stagnation pressure, temperature profiles and velocity direction are imposed, thus, the boundary layers develop freely. Outside the nozzle, static pressure p_∞ = 101 325 Pa is imposed at the outflow boundary. On the lateral and upstream boundaries, non-reflecting boundary conditions are imposed using uniform external flow (static temperature T_∞ = 280 K and velocity U_∞ = 5 ms^{-1}). All the walls are assumed to be adiabatic. The time step of the Runge-Kutta scheme is $\Delta t = 2\times 10^{-8}$ s for Very Fine mesh (Fine: $\Delta t = 2\times 10^{-7}$ s).

The simulation is divided in two parts:

A first run (~ 100ms) is carried out to reach an established state for the turbulent jet mixing layer and for the acoustic field.

Then, the computation is performed for a physical time of about 100ms (500 D/U_j) for the aerodynamic and acoustic analyses. During this second run, time-averaged quantities are computed and instantaneous fields on the FW-H control surfaces are stored for acoustic analysis in far field. This time step storage for the FW-H calculations is small enough to correctly take into account the temporal scales of the acoustic phenomena.

Time histories of flow variables are also stored in the near field, on the jet axis, around the potential core end, and in the shear layer. Most of the signals are located inside the refined mesh zone. A common time step of $\Delta t_{sto} = 10^{-5}$ s (i.e., $\Delta t_{sto} \sim 0.05D/U_j$) has been used for all the storages.

2.3 Acoustic computation

The noise radiation is performed using the Ffowcs Williams and Hawkings surface formulation available in the code KIM [29] developed at ONERA. This formulation computes time pressure histories at any observer location by integration of the flow field solution on a control surface surrounding the jet and containing all the noise sources. For the Very Fine grid, two closed control surfaces extend over the entire length of the refined mesh zone, i.e., from X/D = -2 up to X/D = 25, while for the Fine mesh the same surfaces are not closed. These surfaces differ by their radial extent. The far field microphones are located at 75D from the nozzle exit for the directions between 20° and 150°, as in the experiments. Due to the spurious noise generated in aeroacoustics computations by the flow turbulence passing through a control surface, additional surface terms proposed by Rahier [30] are added in order to reduce this spurious noise. The results using different configurations of the control surface together with these additional surface terms is analyzed in the present paper.

3 AERODYNAMIC AND ACOUSTICS RESULTS

In this section, aerodynamic and acoustic results of the simulations with comparisons to measurements are presented and discussed in order to see how they reproduce the jet configuration. The mesh effects will be studied through the results comparison between both grid "Fine" and "Very Fine" with experimental data. First of all, the shear layer development and its initial state are studied. Then, the jet development is examined and how the near and far pressure fields match the experimental data.

3.1 Jet development

Figure 2 reproduces unsteady vorticity norm $|\omega|$ at the nozzle exit for both grids (left: Fine, right: Very Fine). It appears that, for the fine simulation, a part of the shear layer is laminar and exhibits pairings downstream of the nozzle exit. On the other hand, the transition from a laminar state to a turbulent state appears sooner for the Very Fine simulation (around X/D = 0.02) than for the other simulation (around X/D = 0.05). Moreover, the vorticity levels are higher in the Very Fine simulation, especially in the shear layer. As can be seen on axial vorticity snapshots in the plane Z/D=0 presented in Figure 3, the Very Fine simulation shows a vorticity field richer in small structures more tridimensional with more intense vortical structures than the Fine simulation. These results are confirmed by the snapshots of instantaneous axial vorticity for $-20\ U_j/D \leq \omega_x \leq 20 U_j/D$ in the three planes (from left to right) X/D=0.25, 0.5 and 1 for both meshes (top: Fine, bottom: Very Fine). Both simulations have similar vorticity field despite having different grid resolutions. An enlargement of the latter is given in Figure 4 in the plane X/D=0.25 for both meshes (left: Fine, right: Very Fine) and allows to understand the discrepancies between both simulations.

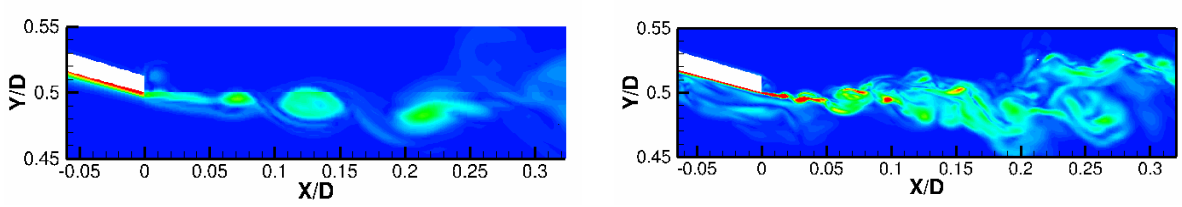


Figure 2: Instantaneous vorticity norm for $0 \leq |\omega| \leq 200Uj\ /D$ in the plane Z/D=0 downstream of the nozzle lip for both meshes (left: Fine, right: Very Fine).

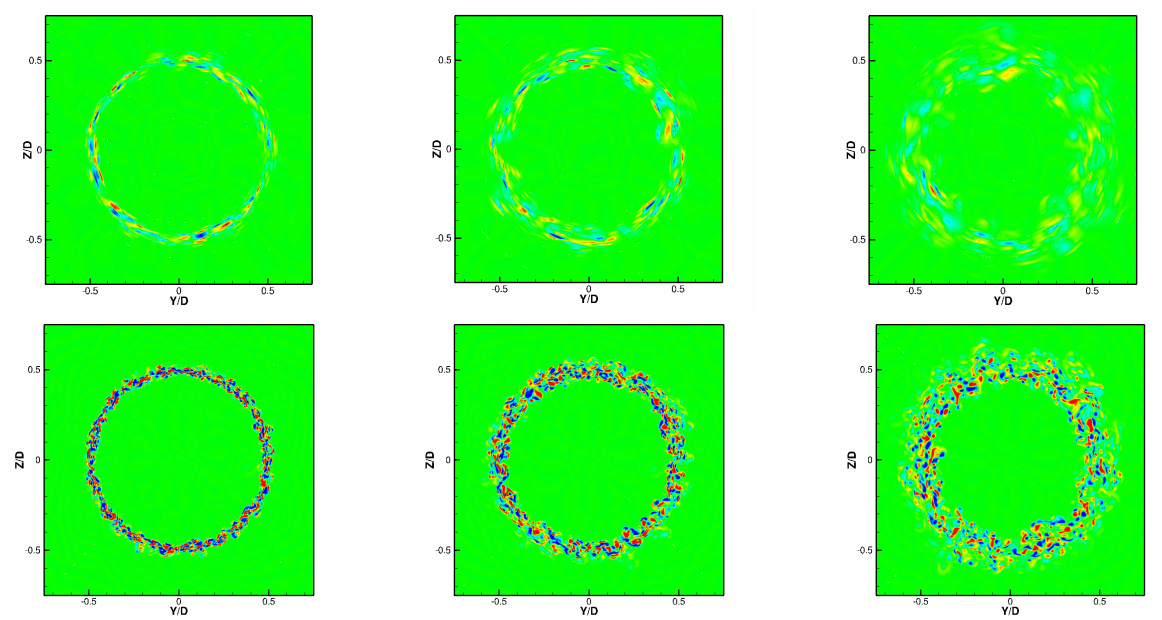


Figure 3: Instantaneous axial vorticity for $-20\ U_j\ /D \leq \omega_x \leq 20U_j\ /D$ in the three planes (from left to right) X/D=0.25, 0.5 and 1 for both meshes (top: Fine, bottom: Very Fine).

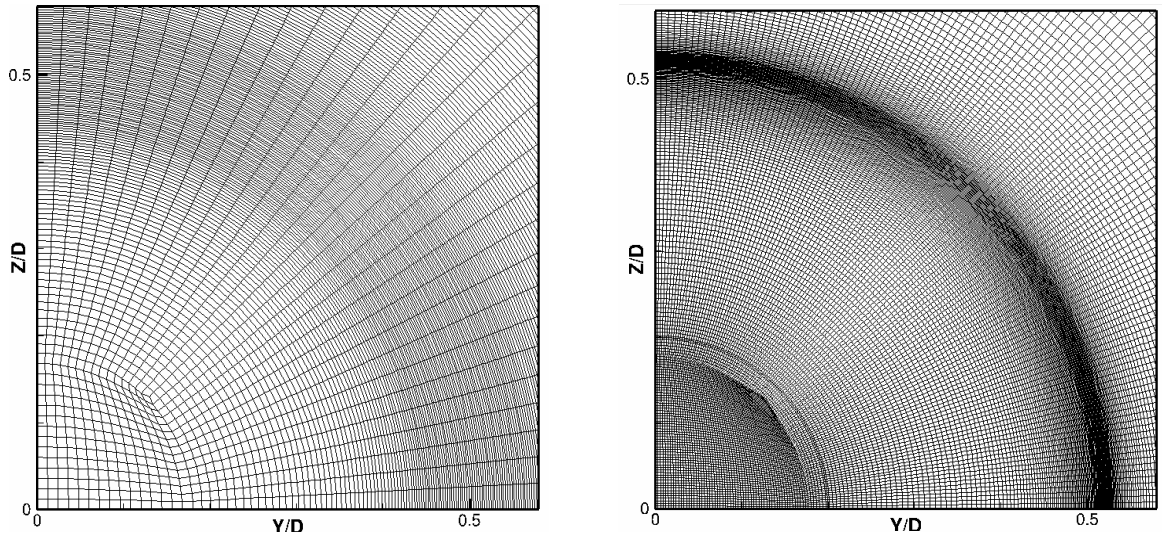


Figure 4: Enlargement of the grid in the plane X/D=0.25 for both meshes (left: Fine, right: Very Fine).

The shear layer state presented previously has an impact on the jet development. As can be seen in Figure 5, both simulations have different evolutions of the mean and rms axial velocity on the jet axis. The potential core $L_c\ /D$ corresponds to U_{mean}/U_j=0.95. Compared to experimental data, the Fine simulation has a shorter potential core length, around $L_c\ /D$ = 4.6,

while the Very Fine simulation has a larger one with L_c /D = 6.3. The axial location of the peak of rms for the Very Fine simulation is shifted downstream compared with the one obtained experimentally. Both simulations have a lowest peak of rms velocity compared to the experiments, due to the rms level (0.03) already present at X/D=0 in experimental measurements. The mesh refinement shows an impact on the rms peak location, but not on its level.

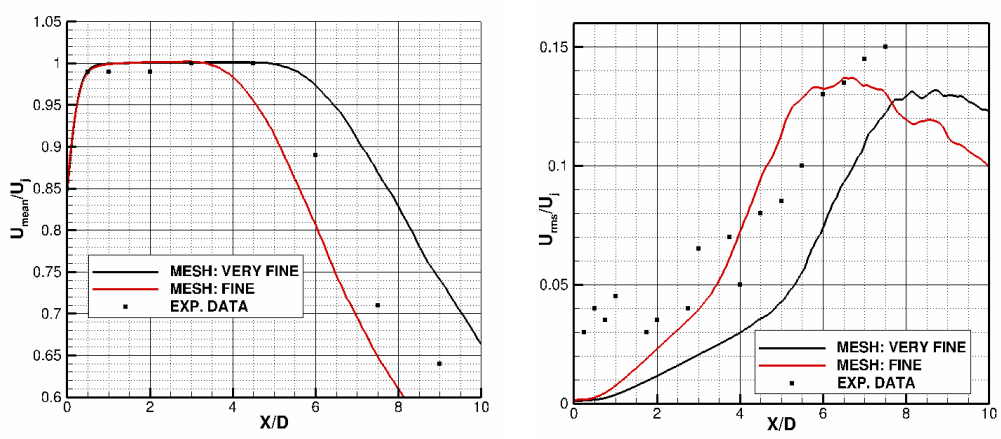


Figure 5: Comparison of the axial evolution on the jet axis of the mean and rms axial velocity for both simulations with experimental data.

Figure 6 shows rms values of the axial velocity normalized by the jet velocity as well as both control surfaces (S1, S2) for the Very Fine simulation. The highest levels of turbulence are located in the shear layers; they start at the vicinity of the nozzle exit and end downstream of the potential core, for X/D ≈ 10. This region of high turbulent fluctuations represents the major part of the acoustic sources, that radiate to the far-field. For axial positions downstream of X/D = 10, where U_{rms} /Uj = 13% close to the jet axis, the fluctuation level decreases with increasing axial distance. The turbulence level falls for instance to 5.5% of the jet velocity at X/D = 25 and to 4.1% at X/D = 30. An enlargement of U_{rms} /Uj in the vicinity of nozzle exit displayed in Figure 7 shows the flow laminarity at this location and confirms the beginning of turbulence around X/D = 0.02 found in the instantaneous vorticity field in Figure 2.

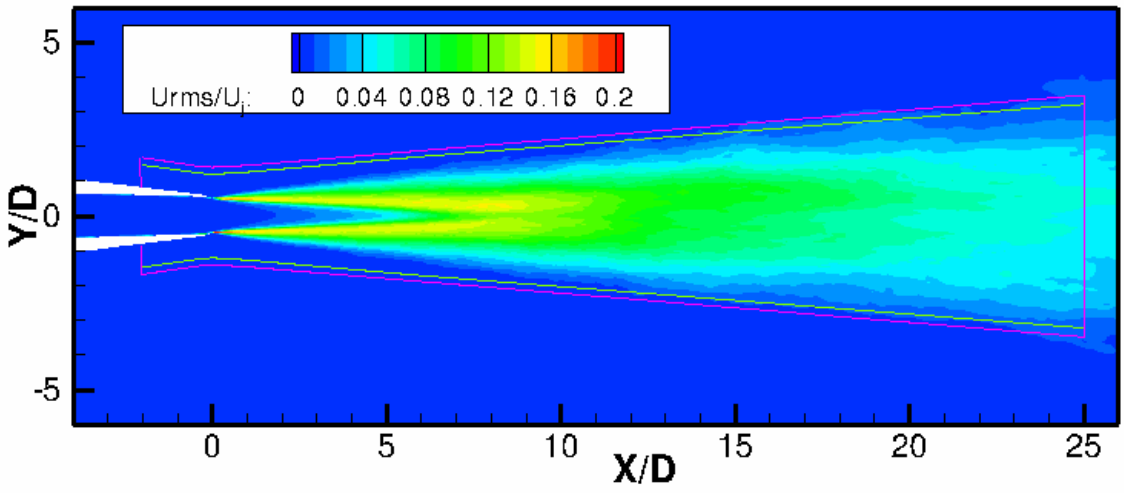


Figure 6: Iso-contours of rms axial velocity in the plane Z/D = 0 and integration surfaces (S1:blue and S2: purple) for acoustic calculations.

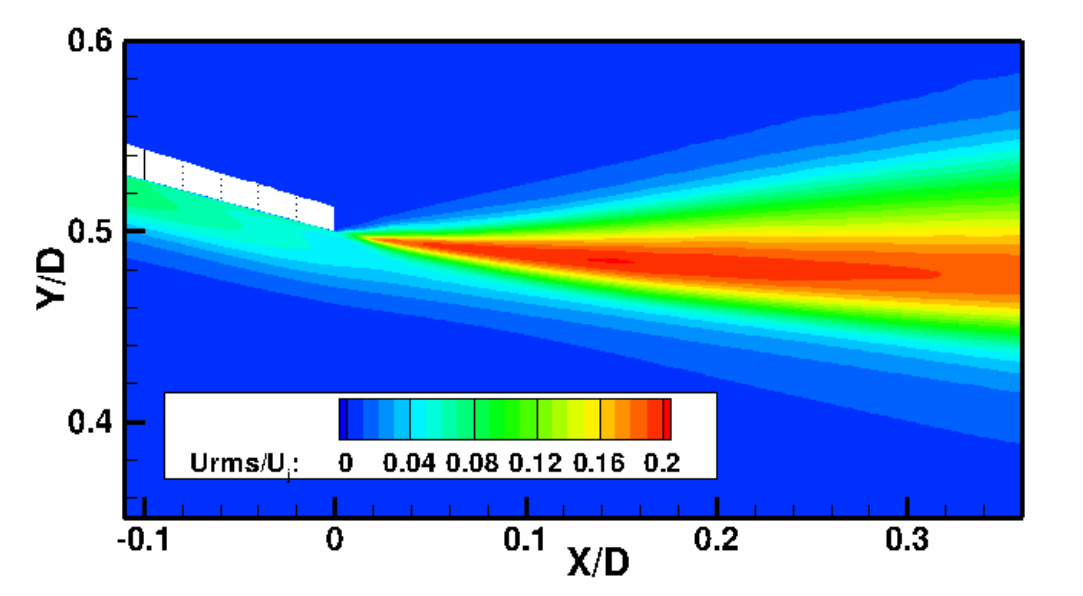
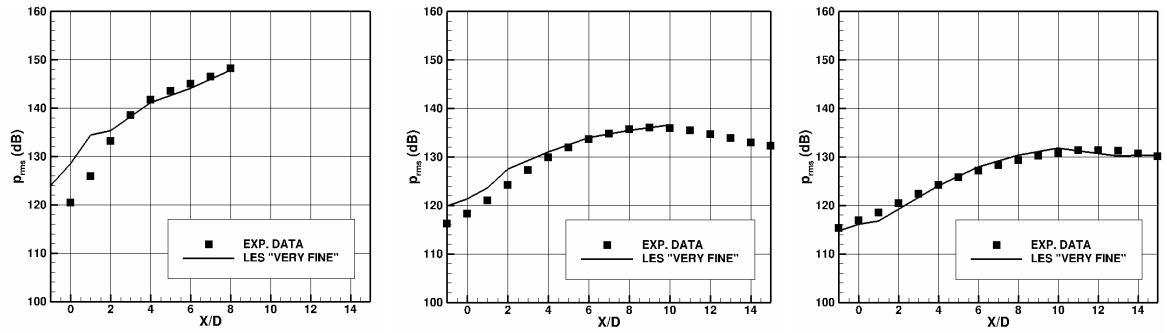


Figure 7: Iso-contours of rms axial velocity in the vicinity of nozzle exit.

3.2 Near field pressure

Thanks to pressure time histories stored in the near field, Figure 8 represents the axial evolution of the rms pressure levels, azimuthally averaged, at three different radial positions (r /D $\in$ {1.5; 3; 5}) for the Very Fine simulation and experimental data. For the three radial positions, a good agreement is observed between numerical and experimental data for X/D $\geqslant$ 4. However, for axial positions near the nozzle exit, an overestimation can be observed, especially for radial positions close to the jet (r/D = 1.5). At the location r/D=5, there is a good agreement between the experimental data and the simulation.



r/D=1.5 r/D=3 r/D=5

Figure 8: Comparison of the longitudinal evolution of the rms pressure at radial positions r /D $\in$ {1.5; 3; 5} between the Very Fine simulation and experimental data.

Hall *et al.* [31], Coiffet [32] and Muller [18] among other authors studied the azimuthal structure of near field pressure and observed that low wave number modes are dominant and the axisymmetric one (m=0) has a downstream growing contribution. The modes |m| < 5 include the most part of energy in the near field. The Strouhal number St is defined by St = f .D /U$_j$ where f corresponds to the frequency. The power spectral density (PSD) of the pressure signal of the axisymmetric mode are presented in Figure 9 for three different axial positions X/D=0, 4 and 8 and for two radial locations r/D=1.5 and 5. The overestimation observed for rms pressure levels at the location X/D=0 and r/D=1.5 is also visible on the power spectral

density (PSD) of the pressure signal for St $\in$ [0.4; 4]. A better agreement is observed when the axial distance is increased, which means that the sound radiated downstream is captured more accurately than the sideline radiated noise, because the turbulence fine scales require an higher mesh refinement. The effect of numerical dissipation on the PSD is also noticeable in Figure 9 at r/D = 5 for St $\geqslant$ 1. Apart from the overestimation visible for positions close to the nozzle exit, a good agreement between numerical and experimental data is found for axial positions X/D $\geqslant$ 4. The near field pressure is thus considered well reproduced by the simulation except for a zone around the nozzle exit.

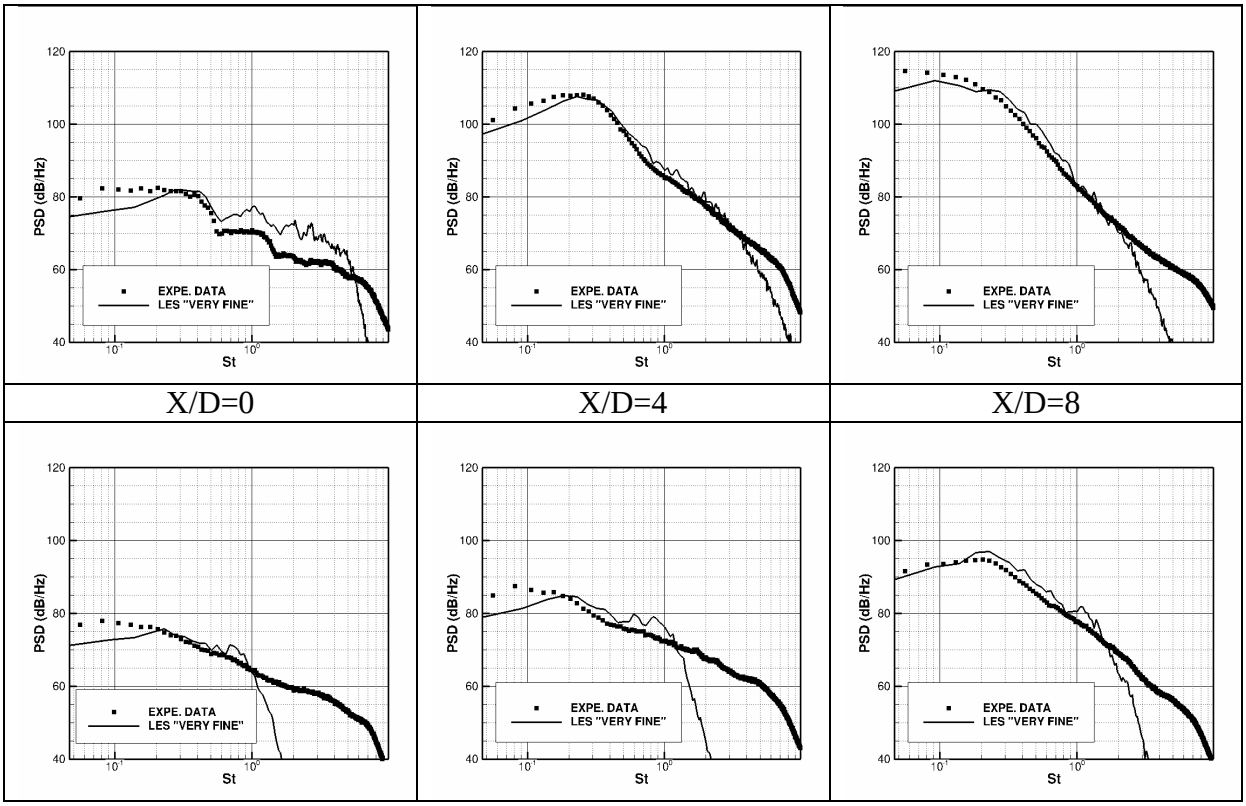


Figure 9: PSDs of the pressure field in the near field for different axial and radial positions : r /D = 1.5 (top), r /D = 5 (bottom) for m=0.

3.3 Far field pressure

The integration of the Ffowcs Williams and Hawkings (FW-H) equation on a porous surface allows determining the pressure radiated by a volume of perturbed fluid knowing the perturbation field on a closed control surface surrounding this volume. It is often used for noise prediction of isolated jets for which a control surface enclosing the mixing layer should contain all the noise sources. In theory, the control surface must be closed, but open surfaces can be used in practice if their extent is sufficiently large to enclose the noise sources adequately. However, if the control surface is closed, that reduces its extent and thereby the volume of data to be stored. In the present study, the FW-H method is used to analyze the acoustic content of the LES fields through the radiated pressure obtained in the far field. These radiation calculations are performed with two control surfaces S1 and S2 drawn in Figure 6, that axial and radial expanses are given below:

- Surface S1: from X_{min}/D = -2 to X_{max}/D = 25, r_{min}/D = 1.69, r_{max}/D = 4.52
- Surface S2: from X_{min}/D = -2 to X_{max}/D = 25, r_{min}/D = 1.98, r_{max}/D = 4.91

These surfaces are open for the "Fine" simulation and closed at X/D = -2 and X/D = 25 for the "Very Fine" simulation, but for this last case the surfaces can be analyzed open.

The signals obtained for a far field observation point, using the FW–H surface integral on the surfaces S1 and S2 open at both ends are compared in Figure 10. The predicted time histories are very similar and quite identical for both surfaces. This suggests firstly that the volume sources outside the control surface S1, not taken into account in the present acoustic analysis, make a very weak contribution to noise radiation. This result validates the choice of these control surfaces for the FW–H formulation. These similar signatures also demonstrate that the perturbation fields have been correctly transported from S1 to S2 by the aerodynamic calculation and therefore the CFD grid is sufficiently dense for a correct propagation of perturbations to the control surfaces. In the following analysis, only the surface S1 for "Very Fine" simulation is presented.

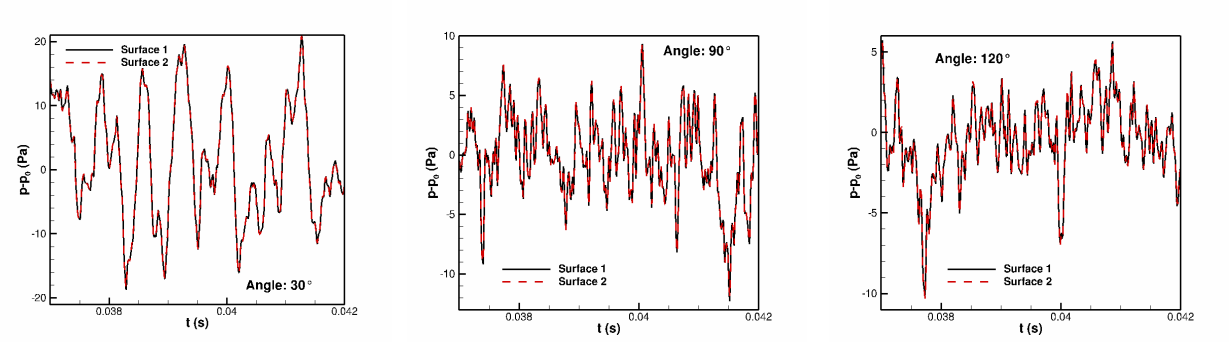


Figure 10: Time history of the pressure predicted in the far field (30°, 60° and 120°) using FW–H formulation on both integration surfaces for "Very Fine" simulation.

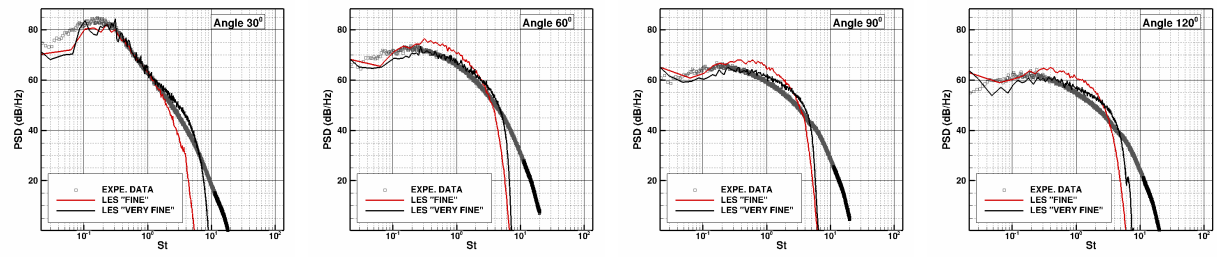


Figure 11 : Pressure Spectral Densities for four observer locations at r = 75D for both simulations compared to the experimental data.

The acoustic analysis is performed for 37 observation points uniformly distributed on a circle located in far field (75D from the nozzle exit) and centered on the nozzle exit, for comparison with experimental results. Their angular position θ is defined with respect to the jet axis and ranges from 0° (downstream) to 180° (upstream). The radiated pressures at the observer points are exploited only on 78 ms because of non reliable values in the beginning and in the end of the signals (incomplete contribution of the control surface because of the acoustic propagation delays). These radiated pressures are described also using a time step of 10^{-5} s. Fourier Transforms are applied on these signals and the levels are averaged over eight frequencies in order to smooth the spectra. These spectra are plotted for the range [0.2–20 kHz].

The comparison of the PSD between both simulations with open control surfaces and experimental data presented in Figure 11 for different angles 30, 60, 90 and 120° shows globally a better agreement for the Very fine simulation than for the Fine simulation with experimental

data. Underestimations are found for low frequencies St $\leqslant$ 0.1 and a slight overestimation for angles $\geqslant$90° for frequencies St $\geqslant$ 1.

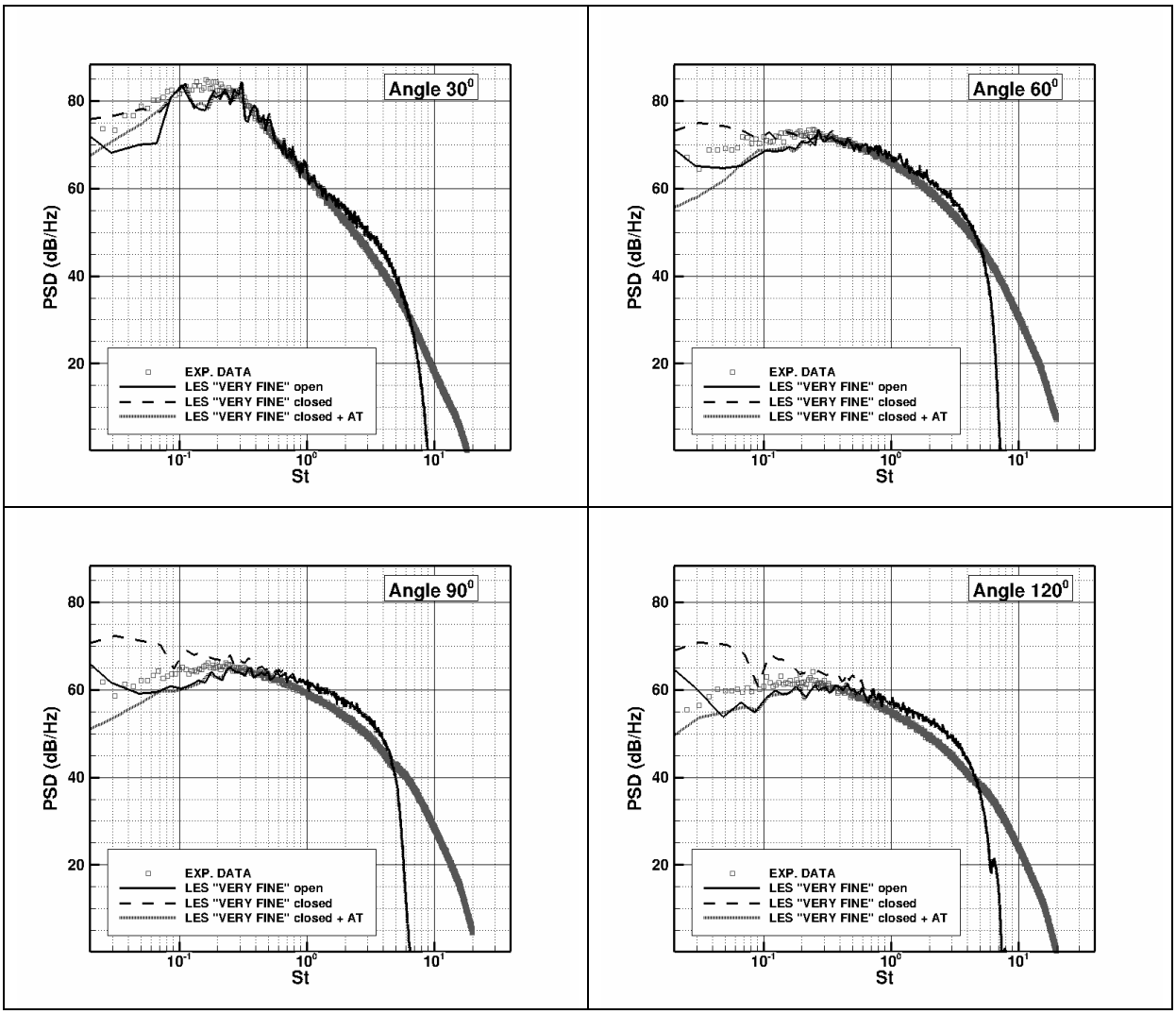


Figure 12: Pressure Spectral Densities for four observer locations at r = 75D for the Very Fine simulation (open, closed, closed + additional terms) compared to experimental data.

In [30], Rahier *et al.* proposed additional surface terms in order to reduce the spurious noise generated in aeroacoustic computations by the flow turbulence passing through a Ffowcs Williams and Hawkings (FW-H) control surface. This spurious noise is due to the fact that some volume sources are not taken into account in the calculations limited by surface integrations. An expression of additional surface terms representing an approximation of the missing volume integral has been defined. These additional terms include two separate Doppler amplification effects, respectively related to the flow velocity and to the control surface velocity and can be applied to any fixed or moving, rigid or deformable control surface. Compared to the use of open control surfaces, closed surfaces have the advantage of filtering any acoustic wave coming from the boundaries of the CFD grid. Another advantage for jet noise predictions is the ability to use shorter control surfaces that are less penalizing for the CFD grid refinement and for the storage of aerodynamic data for acoustic calculations. Short control surfaces have however the drawback of not taking correctly into account the refraction effects for the low angles of observation.

In the following, three configurations are tested and compared on the Very Fine simulation on the surface S1:

- LES Very Fine open : calculations without additional flux terms with the control surface S1 open at both ends.
- LES Very Fine closed : calculations without additional flux terms with the control surface S1 closed at both ends.
- LES Very Fine closed +AT : same as previously with additional flux terms.

Figure 12 presents the spectra comparison obtained for four observation angles, 30°, 60°, 90° and 120° for the three previous cases. Globally, for medium and high frequencies the PSD levels do not depend on the control surface and are in fairly good agreement with the experimental results. In contrast, at low frequencies the levels vary a lot with the control surface choice and appear to be largely overestimated for closed surface without additional terms. The more the angle is large, the more the overestimation concerns a wider band of frequencies. The stability of the results for the dominant frequencies of the actual acoustic radiation of the jet confirms that the main noise sources are contained inside the surface S1, because this surface extends very far downstream. It also confirms that open control surfaces commonly used for jet noise predictions can provide reliable results except for low frequencies.

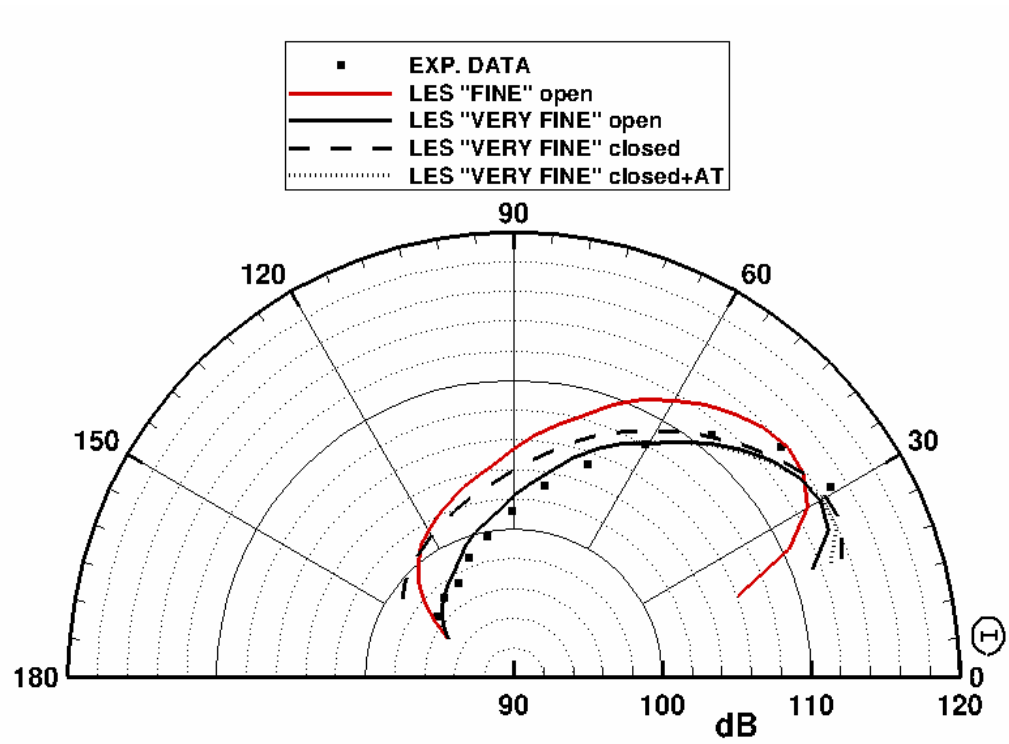


Figure 13: the overall sound pressure level (OASPL) directivity in the far field
for Fine and Very Fine (open, closed, closed +AT) simulations and experimental data.

The Overall Sound Pressure Levels (OASPL) for the directivities are calculated in the range [0.2–20 kHz] i.e. St $\in$ [0.4; 4] in accordance with the experimental data and presented in Figure 13. First of all, the comparison between Fine and Very Fine simulations with open surface shows clearly the improvement of the directivities due to the mesh refinement in all directions, especially in the azimuthal one, so as to obtain isotropic mesh cells with a grid cut-off frequency sufficiently high. The overall sound pressure level (OASPL) directivity in the far field is well predicted by the Very Fine simulation within a ±1 dB margin. However, overestimation due to pairings noise still remains, especially in the direction normal to the jet axis.

The acoustic directivities confirm that the use of closed control surfaces without additional flux terms is not suitable. Using closed control surfaces with additional flux terms seems to provide the same acoustic directivities as with an open control surface, because in this study this surface is long enough, except for low angles. Both acoustic directivity shapes are very close to the experimental data.

The overestimation for frequencies around St = 1, especially visible at 90◦ in the far field, is linked to the one observed in the near field for axial positions close to the nozzle exit (i.e., for X/D $\leqslant$ 4). Nevertheless, the difference is less visible in the far field. Indeed, at X/D = 0 and r/D = 1.5, there is an overestimation of around 6-7 dB/Hz, whereas for the signals in the far field at 90◦, the overestimation is about 4 dB/Hz. The high overestimation in the near field may be attributed to the hydrodynamic pressure, which amplitude decreases according to an exponential law, while the acoustic pressure amplitude decreases according to a $1/r^2$ law. The lower overestimation of 5 dB/Hz observed at X/D = 0 and r/D = 5 can be viewed as a consequence of the dissipation of the hydrodynamic part.

4 CONCLUSIONS

The present work deals with the numerical simulation of a hot subsonic turbulent jet using LES. The influence of the axial, radial, and azimuthal mesh refinements has been studied, thanks to a comparison between the present simulation (Very Fine) and a previous one (Fine) on the same jet configuration. The development of the shear layer depends on the mesh refinements. The vorticity field in the Very Fine simulation is richer in small structures and more tridimensional with more intense vortical structures than in the Fine one.

Despite having a good development, the shear layer remains laminar at the nozzle exit. Compared to the experimental data, the Fine simulation has a shorter potential core length, while the Very Fine simulation has a larger one. The axial location of the peak of rms for the Very Fine simulation is shifted downstream compared to the experimental one. At the nozzle exit, there is no turbulence in the jet center, so both simulations have a lowest rms velocity peak compared to the experiments.

The azimuthal structure of near field pressure confirms that axisymmetric one (m=0) has a downstream growing contribution. The overestimation observed for rms pressure levels at the location X/D=0 and r/D=1.5 is also visible on the power spectral density (PSD) of the pressure signal for St $\in$ [0.4; 4]. Due to the turbulence fine scales, that requires an higher mesh refinement, the sound radiated downstream is captured more accurately than the sideline radiated noise. For St $\geqslant$ 1 at r/D = 5, the numerical dissipation effect on the PSD is also noticeable. The near field pressure is thus considered well reproduced by the simulation except for a zone around the nozzle exit.

The Ffowcs Williams and Hawkings (FW-H) formulation is used for acoustic analysis of aerodynamic numerical simulations. This acoustic analysis is performed for observation points located in far field (75D from the nozzle exit) and centered on the nozzle exit for comparison with experimental results. Globally, the mesh refinement contributes to improve the Pressure Spectral Densities for different angles compared with experiments, although discrepancies are found for low frequencies (St $\leqslant$ 0.1) and for angles $\geqslant$90° for Strouhal St $\geqslant$ 1. Without additional flux terms, the turbulence passing through the downstream closing disk generates relatively high levels of spurious radiated pressure, which affects a wide band of frequencies, including the low frequencies. The use of closed control surface without additional flux terms has therefore no real interest for jet noise analysis. The results obtained using closed control surfaces with additional flux terms lead to the almost same results as the ones

obtained with open surface because the surface extent used in this study is long enough to include the main acoustic sources.

REFERENCES

[1] M. Wang, J. B. Freund, and S. K. Lele, Computational prediction of flow-generated sound, *Annual Review of Fluid Mechanics*, **38**, 483–512, 2006.

[2] J. B. Freund, Noise sources in a low-Reynolds-number turbulent jet at Mach 0.9, *Journal of Fluid Mechanics*, **438**, 277–305, 2001.

[3] X. Jiang, E. J. Avital, and K. H. Luo, Sound generation by vortex pairing in subsonic axisymmetric jets, *AIAA Journal*, **42**, 241–248, 2004.

[4] N. Andersson, L.-E. Eriksson, and L. Davidson, Large-eddy simulation of subsonic turbulent jets and their radiated sound, *AIAA Journal*, **43**, 1899–1912, 2005.

[5] D. J. Bodony, and S. K. Lele, On using large-eddy simulation for the prediction of noise from cold and heated turbulent jets, *Physics of Fluids*, **17**, 2005.

[6] C. Bogey, and C. Bailly, Computation of a high Reynolds number jet and its radiated noise using large eddy simulation based on explicit filtering, *Computers and fluids*, **35**, 1344–1358, 2006.

[7] S. R. Koh, W. Schröder, and M. Meinke, Turbulence and heat excited noise sources in single and coaxial jets, *Journal of Sound and Vibration*, **329**, 786–803, 2010.

[8] F. Vuillot, N. Lupoglazoff, and G. Rahier, Double-stream nozzles flow and noise computations and comparisons to experiments, *Proceedings of the 46th AIAA Aerospace Sciences Meeting*, No. AIAA 2008-9, 2008.

[9] K. B. M. Q. Zaman, Effect of Initial Condition on Subsonic Jet Noise, *AIAA Journal*, **23**, 1370–1373, 1985.

[10] M. Huet, Influence of boundary layers resolution on heated, subsonic, high Reynolds number jet flow and noise, *AIAA Paper*, 2013-2141, 2013.

[11] C. Bogey, O. Marsden, and C. Bailly, Influence of initial turbulence level on the flow and sound fields of a subsonic jet at a diameter-based Reynolds number of 10^5, *J. Fluid Mech.*, **701**, 352–385, 2012.

[12] C. Bogey, O. Marsden, and C. Bailly, Large-eddy simulation of the flow and acoustics fields of a Reynolds number 10^5 subsonic jet with tripped exit boundary layers, *Physics of Fluids*, **23**, 035104, 2011.

[13] K. B. M. Q. Zaman, Far-field noise of a subsonic jet under controlled excitation, *J. Fluid Mech.*, **152**, 83–111, 1985.

[14] C. K. W. Tam, Computational aeroacoustics: issues and methods, *AIAA Journal*, **33**, 1788–1796, 1995.

[15] C. K. W. Tam, Advances in numerical boundary conditions for computational aeroacoustics, *Journal of Computational Acoustics*, **6**, 377–402, 1998.

[16] L. S. Ukeiley and M. K. Ponton, On the near field pressure of transonic axisymmetric jet, *Int. J. Aeroacoustics*, **3**, 43–66, 2004.

[17] T. Suzuki and T. Colonius, Instability waves in a round jet detected using near-field phased microphone array, *J. Fluid Mech.*, **565**, 197–226, 2006.

[18] F. Muller, F. Vuillot, G. Rahier, and G. Casalis, Modal analysis of a subsonic hot jet LES with comparison to the linear stability, *Proceedings of the 11th AIAA/CEAS Aeroacoustics Conference*, No. AIAA 2005-2886, 2005.

[19] F. Muller, F. Vuillot, G. Rahier, G. Casalis, and E. Piot, Experimental and numerical investigation of the near field pressure of a high subsonic hot jet, *AIAA Paper* 2006-2535, 2006.

[20] B. Fayard, G. Rahier, and F. Vuillot, Modal analysis of jet flow from a coaxial nozzle with central plug, *AIAA Paper*, 2009-3355, 2009.

[21] J. W. Hall, J. T. Pinier, A. M. Hall, and M. N. Glauser, A spatio-temporal decomposition of the acoustic source in a Mach 0.85 jet, *AIAA Paper*, 2007-442, 2007.

[22] M. Lorteau, F. Cléro, and F. Vuillot, Analysis of noise radiation mechanisms in hot subsonic jet from a validated large eddy simulation solution, *Physics of Fluids*, **27**, 075108, 2015.

[23] J. E. Ffowcs Williams, and D. L Hawkings, Sound Generation by Turbulence and Surfaces in Arbitrary Motion, *Philosophical Transactions of the Royal Society of London*, **264**, A1151, 321-342, 1969.

[24] O. Piccin, CEPRA19: the ONERA large anechoic facility, *Proceedings of the 15th AIAA/CEAS Aeroacoustics Conference*, AIAA 2009-3303, 2009.

[25] O. Labbé, C. Peyret, G. Rahier, and M. Huet, A CFD/CAA coupling method applied to jet noise prediction. *Computers & Fluids*, **86,** 1-13, 2013.

[26] A. Biancherin, N. Lupoglazoff, F. Vuillot, and G. Rahier, Comprehensive 3D unsteady simulations of subsonic and supersonic hot jet flow-fields. Part 1: Acoustics analysis, *AIAA Paper*, 2002-2600, 2002.

[27] C. Laurent, I. Mary, V. Gleize, A. Lerat, and D. Arnal, DNS database of a transitional separation bubble on a flat plate and application to RANS modeling validation, *Computers & Fluids*, **61**, 21-30, 2012.

[28] J. P. Boris, F. F. Grinstein, E. S. Oran, R. L. Kolbe, New insights into large eddy simulation. *Fluid Dyn Res*,**10**,199–228, 1992.

[29] G. Rahier, J. Prieur, F. Vuillot, N. Lupoglazoff, and A. Biancherin, Investigation of integral surface formulations for acoustic post-processing of unsteady aerodynamic jet simulations, *Aerosp. Sci. Technol.*, **8**, 453–467, 2004.

[30] G. Rahier, M. Huet, J. Prieur, Additional terms for the use of Ffowcs Williams and Hawkings surface integrals in turbulent flows, *Computers & Fluids* **120,** 158-172, 2015.

[31] J. W. Hall, J. T. Pinier, A. M. Hall, and M. N. Glauser, A spatio-temporal decomposition of acoustic source in a Mach 0.85 jet, AIAA paper, **442**, 2007.

[32] F. Coiffet, P. Jordan, J. Delville, Y. Gervais, F. Ricaud, Coherent structures in subsonic jets : a quasi-irrotational source mechanism? *International Journal of Aeroacoustics*, Multi-Science Publishing, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON THE APPLICATION OF MODERN TURBULENCE MODELS IN THE FLOW LAMINARIZATION PROBLEMS

Sergey L. Chernyshev[1], Alexander I. Ivanov[1], Andrey Ph. Kiselev[1], Dmitry S. Sboev[1], Leonid L. Teperin[1], Petr P. Vorotnikov[1], Valery V. Vozhdaev[1]

[1]Central Aerohydrodynamic Institute n.a. prof. N.E. Zhukovsky (TsAGI),
1, Zhukovsky st., 140180, Zhukovsky, Moscow reg., Russia
a-ph-kiselev@ya.ru

Keywords: Boundary Layer, Laminar-turbulent Transition, Turbulence Models, Experimental Investigation, Free-stream Turbulence, Boundary-layer Suction.

Abstract. *Analyzed are the results of experimental and computational studies of influence of the incoming flow turbulence level and boundary layer suction on the laminar-turbulent transition characteristics on straight and swept wings. The experimental investigations were carried out in the low-turbulence TsAGI T-124 wind tunnel. CFD-calculations were made basing on the numerical solution of the Reynolds-averaged Navier–Stokes equations (RANS). Turbulence models allowing predict the laminar-turbulent transition were used for closing the system of equations. The comparison is made of the predicted laminar-turbulent transition characteristics obtained using different turbulence models and the test data. The results of application of the semi-empirical e^N-method and simple empirical criteria for the transition prediction are also analyzed.*

1 INTRODUCTION

Numerical methods, based on the solutions of the Reynolds-averaged Navier–Stokes equations (RANS), are widely used nowadays. These equations are closed using semi-empirical turbulence models, with some models enabling in principle to calculate the laminar-turbulent transition.

The problems concerning the influence of the free stream turbulence and the extrapolation of the experimental data obtained in wind tunnels become very urgent in the process of the laminar airplane development. Various active methods of laminar-turbulent transition delay can be used, such as boundary layer suction, surface cooling or local heating, etc. Therefore it is necessary to have the numerical methods permitting calculate laminar-turbulent transition with acceptable accuracy both for wind tunnel and flight conditions, at different boundary conditions on streamlined surface. The validation of the transition prediction methods by comparison of experimental and calculated data obtained under the simulated influence of the above-mentioned factors is of great importance.

For this purpose the detailed experimental investigations of the transition zone on the straight and swept wings were carried out in the low-turbulence T-124 TsAGI wind tunnel. The influences of the incoming flow turbulence and acoustic disturbances were studied on the straight and swept wings models with laminar profile LV6 [1-3]. To increase the level of the flow turbulence, the turbulence grids were used. The effect of local boundary layer suction on the laminar-turbulent transition was studied on the model of low-sweep wing with the symmetric profile SR-0012 [4]. The air suction was realized through the slots in the model surface. The laminar-turbulent transition was detected using three independent methods: the china clay coating, the Stanton probes and the single-wire constant-temperature anemometer. The time-averaged and pulsating flow components in the boundary layer were measured by the hot-wire method.

The possibilities of application of different methods for prediction of the laminar-turbulent transition and for investigation of the increased free stream disturbances effect on the transition characteristics on the straight and swept wings models with laminar profile LV6 were analyzed earlier in conference papers [5, 6]. In particular, the measured characteristics of the boundary layer on straight wing were compared with the results of calculations based on the turbulence models of Langtry–Menter [7] and Walters–Cokljat [8]. The results of using the semiempirical e^N-method and some simple empiric criteria for prediction of laminar-turbulent transition both for straight and swept wings with LV6 profile also were analyzed.

In the present paper the application is studied of modern turbulence model [9] based on the transport equation for the intermittency, to prediction of the laminar-turbulent transition on the swept wings surfaces for both the increased turbulence level of the incoming flow and the presence of the boundary layer suction. Comparison is made of the predicted laminar-turbulent transition characteristics for straight and swept wings with the LV6 airfoil. The comparative analysis of the computed boundary layer parameters and the experimental data is fulfilled [1-4].

2 AERODYNAMIC FACILITY

The experimental investigations of the effects of the incoming flow turbulence and local boundary layer suction on the laminar-turbulent transition on straight and swept wings were carried out in the low-turbulence T-124 TsAGI wind tunnel [10].

TsAGI T-124 subsonic wind tunnel (velocity range – $U_0 = 2\ldots100$ m/s) is a closed-circuit compressor wind tunnel with test section dimensions of 1 m×1 m×4 m. Wind tunnel T-124 is remarkable for sufficiently low initial turbulence level and low noise. These characteristics

were achieved by means of applying high contraction ratio in the nozzle (17.6), high accuracy of maintaining the fan speed, using deturbulizing meshes in the settling chamber with the inner cell size of 0.7 mm, applying diffuser with a small opening angle, thorough finishing of the inner surfaces of the facility's channels and making of wood almost all basic elements of the wind tunnel with the exception of the test and fan sections. So, for the flow speed of less than 80 m/s the level of initial turbulence appears to be only 0.05%. It creates favorable conditions for researches of the boundary layers development leading, in the end, to the laminar-turbulent transition.

3 INVESTIGATIONS OF THE INFLUENCE OF INITIAL FLOW TURBULENCE LEVEL ON THE LAMINAR-TURBULENT TRANSITION ON STRAIGHT AND SWEPT WINGS WITH LAMINAR PROFILE LV6

3.1 Description of the models and experimental techniques

The tests described in this section are the continuation of the researches completed in TsAGI in the framework of the TELFONA project (6^{th} European FP) [1], which are devoted to investigations of the possibilities of implementation of different calculation methods to the predictions of the laminar-turbulent transition and the influence of higher disturbance background level on its characteristics.

The LV6 airfoil model is a rectangular wing with the chord of 1000 mm and the span of 998 mm. This airfoil was designed by DLR as part of the TELFONA Pathfinder Wing design activity. The shape of the laminarized airfoil is close to symmetric one. The model is placed in the closed test section of the wind tunnel T-124 from wall to wall at the equal distances from the floor and from the ceiling. Since such wing position permits to avoid the formation of tip vortices, the modeling of the flow over the "infinite span" wing is performed in the best way and the flow in the middle section of the model may be considered as two-dimensional. The photographs of the 2D LV6 airfoil model before and after installation in the test section of the T-124 wind tunnel are shown at Figure 1. The right photo shows also the *XZ*-traverse sting, the sting extender and *Y*-traverse gear.

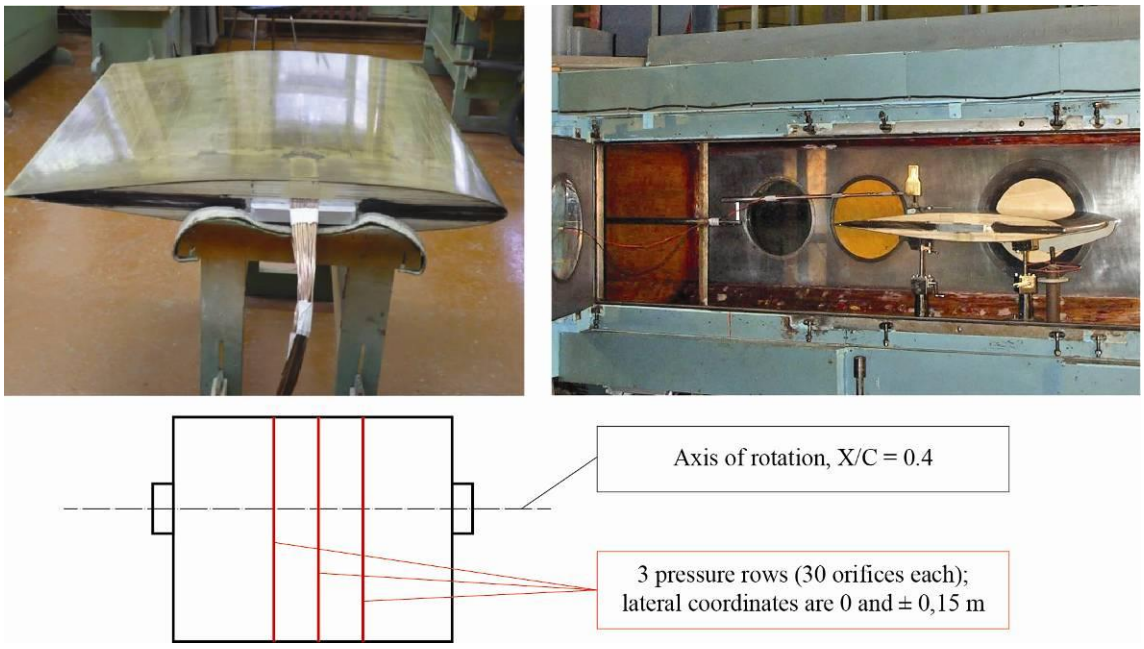


Figure 1: 2D LV6 airfoil model before and after installation in the test section of the T-124 tunnel

The second model has been tested is the wing section model with the sweep angle $\chi = 35°$, with the chord length along the free stream direction $C = 1000$ mm and the span of 998 mm. This model has the same shape of the LV6 laminar airfoil in the section along free stream direction, with the relative thickness of 11%.

In the closed test section of the T-124 wind tunnel the swept wing model was also placed at the equal distances from the floor and from the ceiling from wall to wall. The overview of 2.5D swept wing model with 3D traverse gear in T-124 wind tunnel test section is presented in Figure 2. In order to prevent the spreading additional disturbances along leading edge from the junction area of the wing and right side (in stream direction) test section wall, the model was equipped with the fence. The fence was fixed at distance of 120 mm from the right wall of the test section.

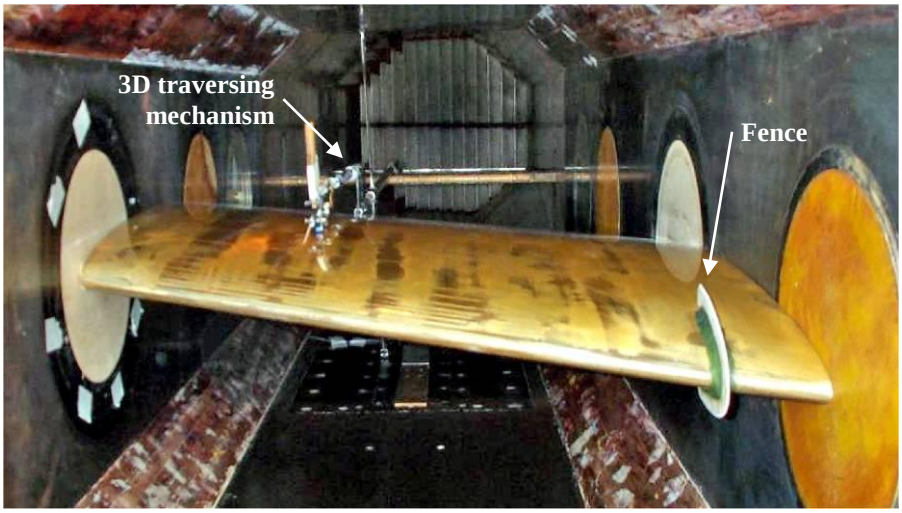


Figure 2: 2D LV6 airfoil model before and after installation in the test section of the T-124 tunnel

Regimes at angles of attack $\alpha = -2°$ and 0 were chosen for further investigations in swept wing boundary layer. Both models have the relative thickness of 11%. The experiments were conducted at free stream velocity $U_0 = 80$ m/s that correspond to Reynolds number 5.5×10^6 based on full chord of the model and Mach number 0.24. The system of coordinates was used with the longitudinal axis X directed perpendicularly to the leading edge of the wing, and the transversal axis Z directed along the leading edge.

The objective of the test was to study the transition mechanism associated with the Tollmien-Schlichting waves or cross-flow instability development in the 2D flow or infinite swept wing condition. Before the main tests, it was necessary to make certain that in the central part of the models there are areas in which the flow parameters do not depend on the lateral coordinate Z. For this purpose both models were equipped with three rows of pressure taps (Figure 1). Also before the main measurements the Y-traverse influence on the averaged flow was investigated by means of the model pressure distributions comparison for two test conditions – with and without the Y-traverse gear. Data analyses showed that the traverse gear influence on the pressure distributions was rather small [1, 2].

For both 2D and 3D cases at $\alpha = 0$ the pressure distributions have the following specific features: after the domain of flow acceleration ($X/C < 0.15$) there is a region practically without the pressure gradient. In this zero pressure gradient region the measured boundary layer velocity profiles coincide very well with Blasius solution for 2D case. On swept wing model there is favorable pressure gradient region at $\alpha = -2°$, $0 < X/C < 0.5$.

The hardware component of the measuring system used consists of the single hot-wire with CTA DISA 55D01, a container for 32 pressure transducers, and an automatized traverse gear, which are connected to the PC through the input/output subsystem.

The experiments were carried out at three flow regimes: at the natural conditions of the wing tunnel test section and with the elevated free-stream turbulence level generated by two grids. Grids were installed at the point of entry to the test section. Turbulence level Tu_u (based on longitudinal component of pulsations u_{rms}) at the model leading edge estimated by the values of 0.7% for grid № 1 and 1.1% for grid № 2. Turbulence level Tu (based on three components of pulsations) at the model leading edge estimated by the values of 0.61% for grid № 1 and 0.91% for grid № 2 due to some degree of anisotropy of generated turbulence (v_{rms}/u_{rms} = 0.73…0.78). For natural conditions Tu = 0.064%. The Taylor micro-scale of turbulence for the longitudinal pulsations appeared to be equal to 5–6 mm for both grids. Longitudinal integral turbulence scale was made by means of integration of the autocorrelation function. For grid № 1 the value of this scale was 21.3 mm and 44.7 mm for grid № 2.

In all experiments the region of laminar flow destruction was determined by studying the intermittency γ in the boundary layer. The following model [11] for γ was used in the transition region

$$\gamma = 1 - \exp\left[-\frac{n\sigma}{U_0}(X - X_t)^2 \right],$$

where X_t is the coordinate of the transition inception, n is the rate of turbulent spots generation in the domain of their emergence, and σ is a kinematic parameter depending on the velocity and angle of propagation of turbulent spots. It was demonstrated [11, 12] that this dependence can be used to describe transition regions in various flows. In particular, in the majority of flows, the function

$$F = \sqrt{-\ln(1-\gamma)}$$

in the transition region could be approximated by a straight lines. The crossings of these lines with X axis were chosen as X_t positions, while the value γ = 0.99 (F = 2.14) was chosen to specify the location of transition completion X_T. In order to characterize transition location with single parameter the position where γ = 0.5 is referred below as $X_{0.5}$. $\Delta X = X_T - X_t$ gives the length of transition region.

All intermittency measurements were performed at the height in the boundary layer corresponding to U/U_e = 0.5, i.e. in the middle of boundary layer (U_e is local external velocity). It allowed conducting the measurements in 3D boundary layer with single-wire probe, because the angles between direction of U_0 and local streamline at this height in the measurements region didn't exceed 2–4° so resulting error didn't exceed 2%. The additional details of experimental techniques could be found in [2, 3].

3.2 Modelling of laminar-turbulent transition on the basis of numerical solutions of the Reynolds-averaged Navier–Stokes equations

Calculations were performed on the basis of numerical solutions of the Reynolds-averaged Navier–Stokes equations, implemented in the software package ANSYS CFX (TsAGI's License №501024).

The mathematical model of the test section, as well as the geometry and the positions of the wings in the wind tunnel fully comply with the conditions of the experiment.

Turbulence models are used to close the Navier -Stokes equations which enable in principle to calculate the laminar-turbulent transition [7-9] .

Model $\gamma - \mathrm{Re}_\theta$ [7] was developed by F. Menter and R. Langtry on the basis of the shear stress transport (SST) equation model. The transition characteristics are defined using the intermittency coefficient γ and number $\tilde{\mathrm{Re}}_{\theta_t} = \rho\theta U / \mu$, corresponding to Reynolds number at the transition beginning, defined using the local velocity U and momentum thickness θ. Both parameters, γ and $\tilde{\mathrm{Re}}_{\theta_t}$ are defined from the solution of the corresponding differential transport equations.

The results of the CFD-studies of the laminar-turbulent transition on the straight wing completed using $\gamma - \mathrm{Re}_\theta$ Langtry and Menter model and another well-known model of the laminar-turbulent transition by Walters and Cokljat [8], are presented in paper [5]. Compared with the Walters and Cokljat transition model based on the transport equations for laminar and turbulent kinetic energy, Langtry and Menter model demonstrates more reliable results.

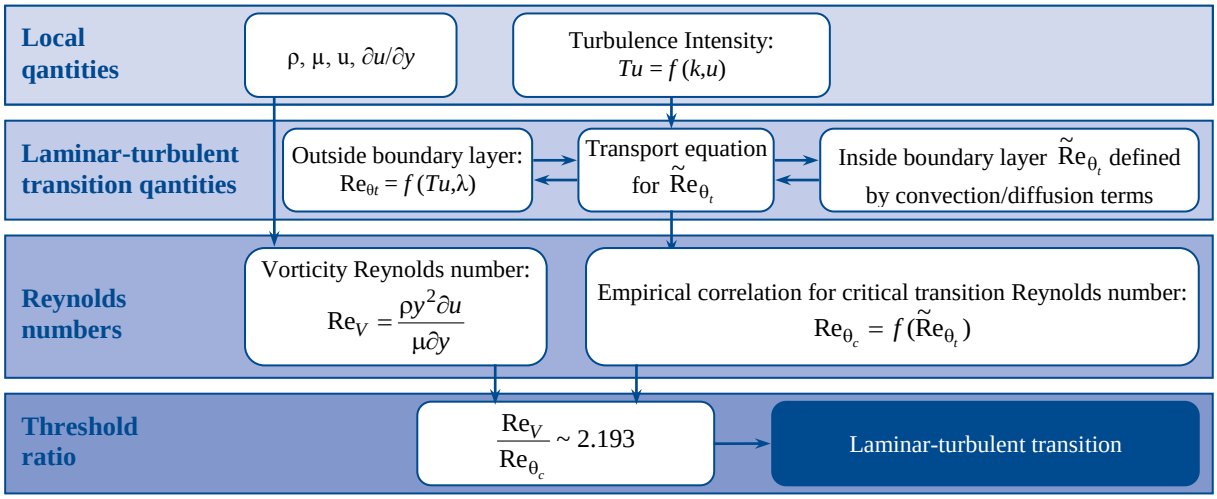


Figure 3: The block - scheme of calculating the position of laminar - turbulent transition on the basis of the model Langtry–Menter [7]

For calculations by means of the ANSYS CFX software the unstructured finite-difference mesh was generated. The total number of nodes in the computational domain reached 11 million and the nodes were combined into volume elements (tetrahedra and prisms). The total number of volume elements is 26.5 million, the total number of tetrahedra being 6.8 million. An additional thickening of mesh was made on the upper surface of the wing, as a result of which the maximum size of an element on the upper surface did not exceed 3 mm.

The turbulence intensity is defined experimentally, as a rule, in the aerodynamic model installation zone. In the numerical methods the turbulence intensity is specified at the inlet to the computational domain. There may be some turbulence attenuation depending on the distance from the inlet to the computational domain x^* to the model.

In modern computational models the turbulence intensity decay is calculated as follows:

$$Tu = Tu_0 \left\{ 1 + \frac{3\rho U \beta Tu_0^2}{2\mu R_T} \cdot x^* \right\}^{-\left(0.5\frac{\beta^*}{\beta}\right)}. \qquad (1)$$

Here U – mean convective velocity, Tu_0 – initial turbulence intensity; ρ – density, β and β^* – constants of the SST (Shear Stress Transport) model of turbulence.

Relative turbulent viscosity R_T is defined as:

$$R_T = \frac{\rho k}{\mu \omega},$$

were ω – the specific rate of dissipation of turbulent energy; k – turbulent kinetic energy.

In the experiments of [1-3] the turbulence intensity was measured in the vicinity of the wing leading edge and it was 0.0637% under natural conditions. On the basis of formula (1) it is possible to obtain the required turbulence intensity at a given distance from the inlet to the computational domain (Figure 4).

It should be noted that in calculation of the wing flow characteristics, the relative turbulent viscosity was taken equal to 1. The values of constants in the SST turbulence model β and β^* were taken 0.09 and 0.0828, respectively. As a model of laminar-turbulent transition the $\gamma - \mathrm{Re}_\theta$ – model of Langtry and Menter was used. The following computational results were obtained using standard values of the model constants.

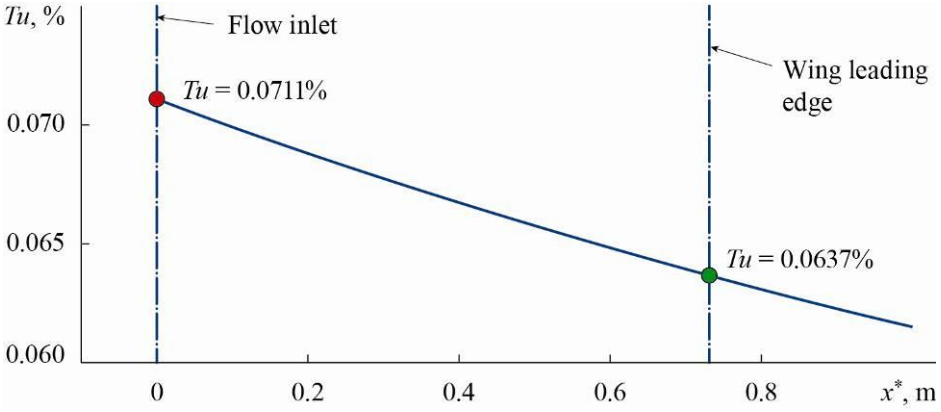


Figure 4: Decay of turbulence intensity depending on the distance downstream

Reynolds-averaged Navier–Stokes equations are closed using the semi-empirical turbulence models. However only few of turbulence models allow predict the laminar-turbulent transition (Figure 5).

Transition on the swept wing is essentially triggered by the instability of the cross-flow, and the presence of both positive and negative pressure gradients contributes to the instability. In the turbulence models of Langtry and Menter [7] and Walters and Cokljat [8] the cross-flow instability is not taken into account.

Further development of $\gamma - \mathrm{Re}_\theta$ turbulence model became new Local-Correlation-Based Transition Modelling (γ-model) model developed by Menter and Smirnov [9], in which the possibility of taking into account the crossflow instability was realized.

It only solves one equation for the intermittency γ and is again based strictly on local variables. Significant effort was invested in ensuring a simple formulation with a limited number of user accessible constants, which allow the fine-tuning of the model for specific applications. Finally, an additional indicator was developed, which allows the detection of crossflow instabilities and allows its connection to the Arnal C1-criterion [13].

It should be noted that since γ-model of turbulence [9] does not involve the transport equation for the transitional Reynolds number, this gives a reduction in calculation time compared with the time of the calculations based on the $\gamma - \mathrm{Re}_\theta$ model of Langtry and Menter applied to calculate the laminar-turbulent transition characteristics on a straight wing.

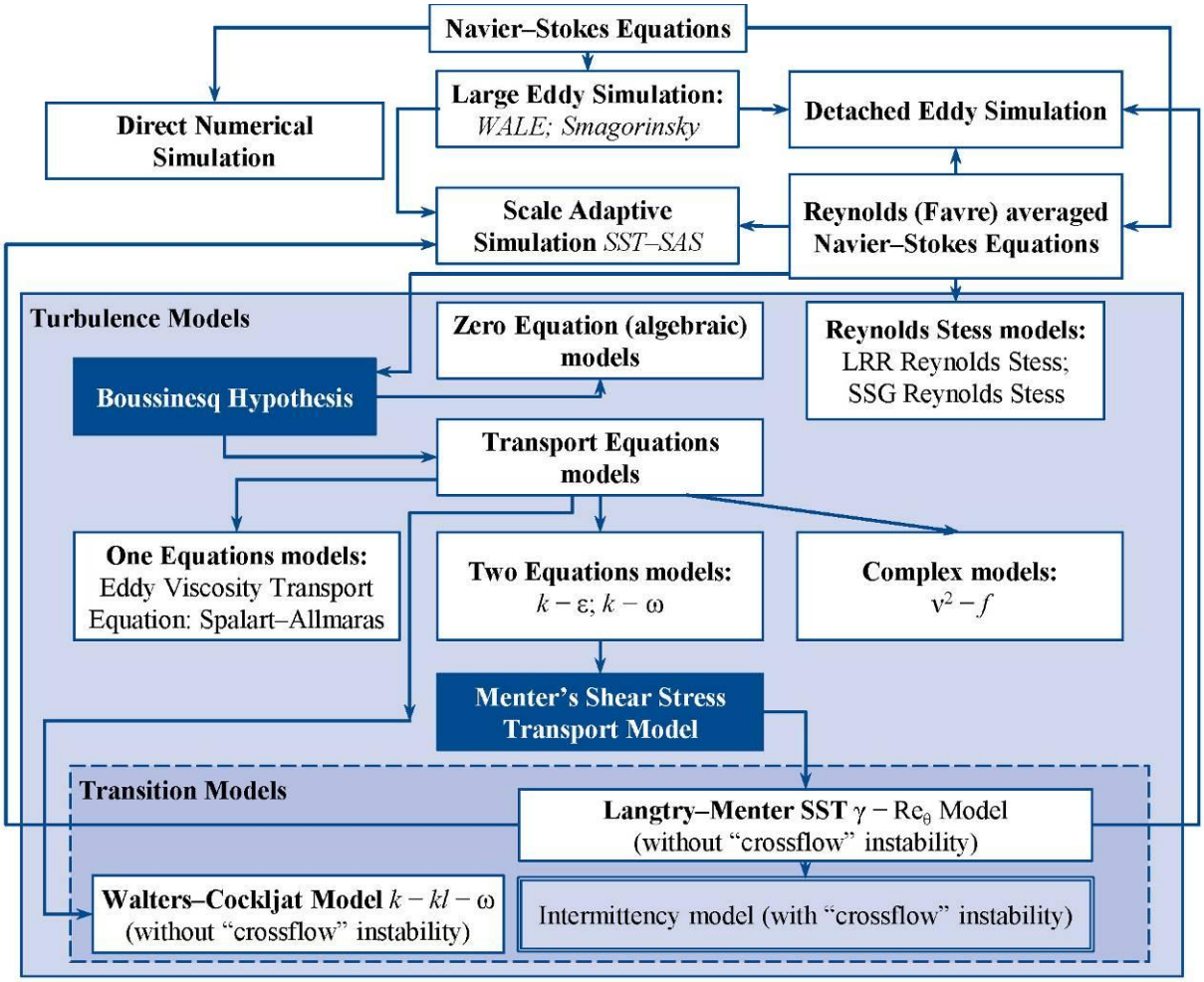


Figure 5: Classification advanced turbulence models

Comparison of experimental and calculated pressure distribution in the central section of the upper surface of the wing at zero angle of attack is shown in Figure 6. The dashed-dotted line shows the conversion of the pressure distribution on the straight wing to the oblique wing with the same flow chord according to the infinite swept wing rule of the linear theory.

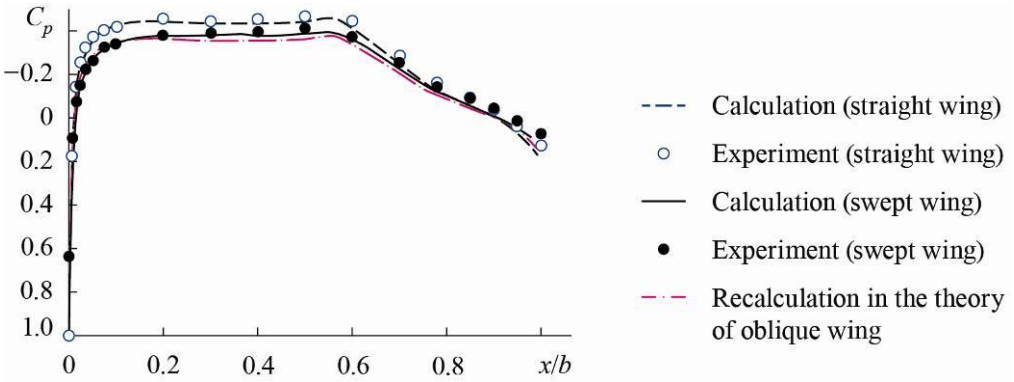


Figure 6: The pressure distributions on the upper surface of wings at $\alpha = 0$

Experimental and calculated intermittency distributions at natural disturbances background in the central section of swept and straight wings, indicating the beginning of the laminar-turbulent transition in the boundary layer, the length of the transitional area and the beginning of fully tur-

bulent flow, are shown in Figure 7. It should be noted that the use of γ-model in the calculation of the laminar-turbulent transition on swept wings for this type of flow provides fairly good accuracy.

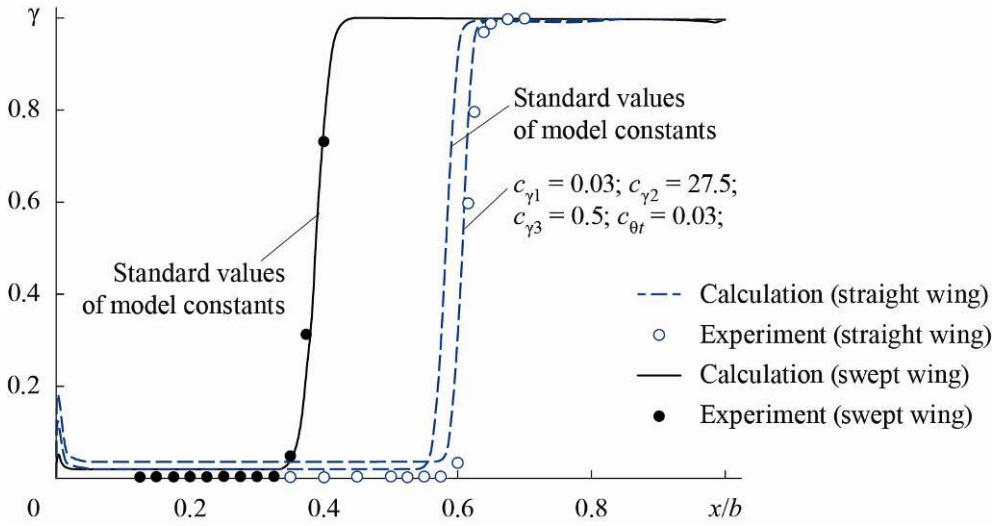


Figure 7: Distribution of intermittency γ on the upper surface of the wings at α = 0 and *Tu* = 0.064%

On the straight wing with the standard values of the model constants of the $\gamma - \mathrm{Re}_\theta$ model there are some differences between the calculated and experimental data (Figure 7). However, conducted computational studies on the effect of constants of the $\gamma - \mathrm{Re}_\theta$ model used to calculate the laminar-turbulent transition on a straight wing, allowed set the values of model constants that provide a good agreement between the calculated and experimental data

In experimental studies [1-3] on the effect of free-stream disturbances on the boundary layer transition the initial turbulence was varied by installation turbulizing grids at the entrance to the test section of WT. According to the measurements of pulsations in the WT without models at a distance of 2.4 m from the grid, it was found that in the vicinity of the model leading edge the turbulence increases to 0.61% in case of installation of the grid No. 1, and increases to 0.91% in case of installation of the grid No. 2.

As already mentioned above, the degree of turbulence defined at the inlet to the computational domain attenuated too fast compared with the experimental data. In this regard, the inlet to the computational domain was placed closer to the wing model. The turbulence intensity in the boundary condition at the flow inlet to the computational domain was set according to the formula (1) so that in the wing area the degree of turbulence corresponded to the experimental data. In this regard, the inlet in the computational domain was moved closer to the model of the wing. The calculations used γ–model of turbulence with standard values of the constants.

Figure 8 shows the intermittency distribution near the surface of the wing in its middle section. This restructure of the computational domain significantly affected the results of the calculation at the natural background of flow disturbances (*Tu* = 0.064%). Comparison of the calculated and experimental results shows that in the absence of turbulizing grid γ-model predicts earlier laminar-turbulent transition compared with the experimental data.

With increased degree of turbulence corresponding to the turbulizing grid No. 1, the CFD prediction demonstrates sharper increase of intermittency in the laminar-turbulent transition zone, but what concerned the position of the point of the transition beginning, calculated and experimental data are in satisfactory agreement (Figure 8). The computational prediction of the character of the intermittency changes corresponding to the presence of grid No. 2 are in good agreement with the experimental data.

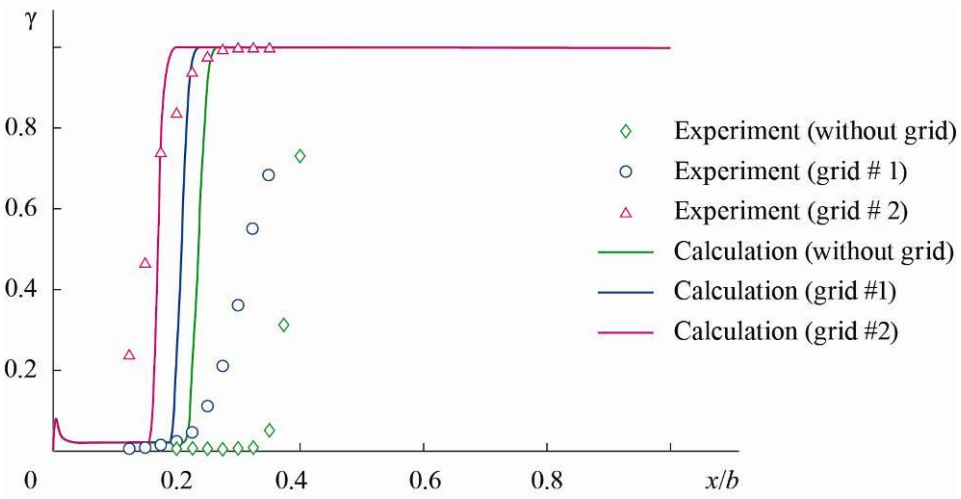


Figure 8: Distribution of intermittency in the central section on the upper surface of the swept wing at zero angle of attack

At the angle of incidence $\alpha = -2°$ without turbulizing grids calculation gives more upstream position of the laminar-turbulent transition compared with the experimental one (Figure 9). At the presence of both turbulizing grids the position of the laminar-turbulent transition shifts forward approximately at the same distance (Figure 9). The experiment shows similar effect. Visualization of the skin friction coefficient, demonstrating the calculated position of the laminar-turbulent transition, is shown in Figure 10.

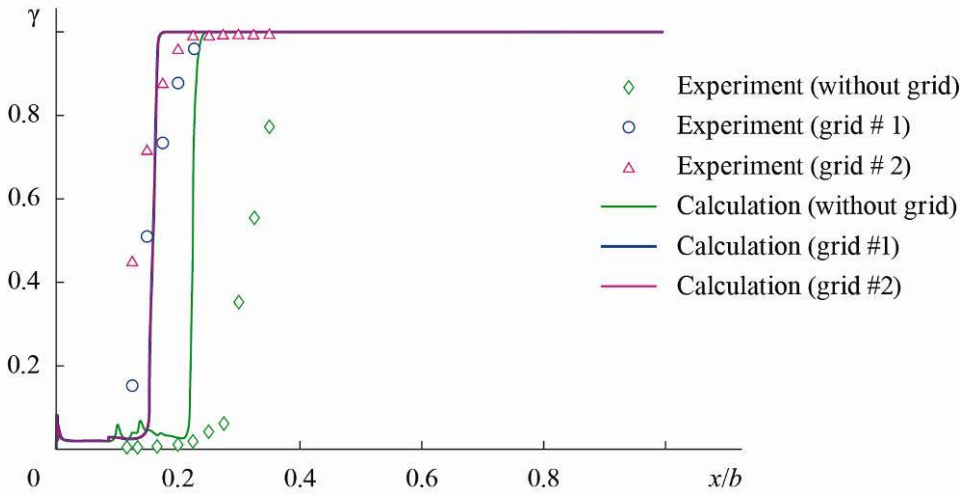


Figure 9: Distribution of intermittency in the central section on the upper surface of the swept wing at $\alpha = -2°$

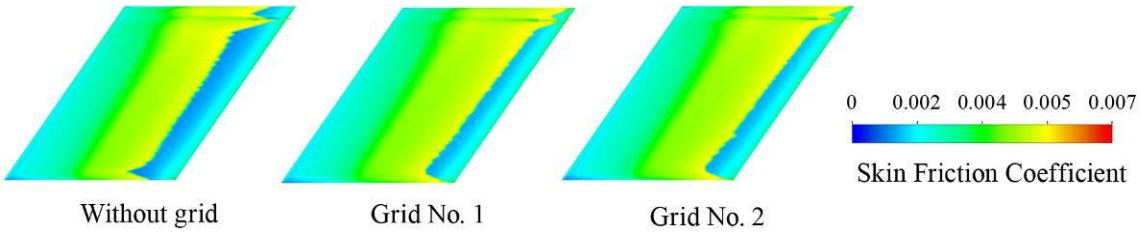


Figure 10: Distributions of skin friction coefficient on the upper surface of the swept wing at $\alpha = -2°$

3.3 The results of using of semiempirical e^N-method and some simple empiric criteria

Regardless of a significant progress in the development of numerical methods, empirical and semi-empirical models of the transition are still widely used in engineering practice.

The results of using of some simple empiric criteria and semiempirical e^N-method for prediction of laminar-turbulent transition both for straight and swept wings with LV6 profile were analyzed in the papers [5, 6] and [14].

For 2D boundary layer at the elevated free-stream-tyrbulence levels the transition length was calculated using models by Mayle [12], Solomon *et al.* [15] and Roberts & Yaras [16].

Also in 3D case for low turbulence environment three simple empirical transition criteria were applied, namely, Arnal C1 [13], Brown [17], and Barinov–Lutovinov [18]. Finally, semi-empirical model recently developed by Ustinov [19] for free-stream-induced transition was tried. It based on transient growth idea with subsequent secondary instability of streaks.

For calculation of hydrodynamic stability characteristics of 2D and 3D three-dimensional boundary layers on the LV6 straight end swept wings general numerical matrix method developed in TsAGI was used [20]. The computation of eigenvalues is performed for real counterpart of initial complex matrix with the use of QR-algorithm. For 2D flow the *N*-factors were computed using envelope method.

For definition of the laminar-turbulent transition location in the range $0.1\% \le Tu \le 1\%$ L.M. Mack [21] proposed the empirical dependence of *N*-factor vs. *Tu*.

The calculated values of the *N*-factor for 2D boundary layer are compared with the Mack's relation in Figure 11. In the case Grid No. 2, correlation overpredicts the *N*-factor of the transition. At $Tu = 0.61\%$, the criterion yields feasible results, especially if the value corresponding to the $X_{0.5}/C$ is treated as the transition point. In Grid No. 1 and Grid No. 2 cases, the *N*-factors corresponding to X_t/C are equal to zero, i.e., the transition started in these situations before the boundary layer lost its stability (by-pass transition). For natural conditions ($Tu = 0.064\%$) the laminar separation with turbulent reattachment were observed with the value of $N = 13.5$ at separation.

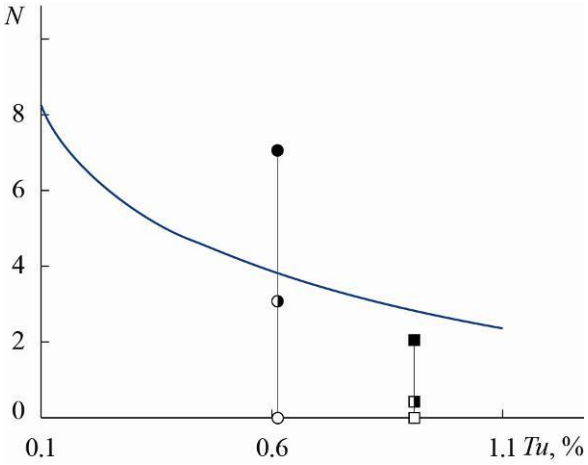


Figure 11: Dependence of the transition *N*-factor on the free-stream turbulence level for 2D flow. The solid line shows the Mack relation [21], open symbols corresponds to the X_t, solid symbols to the X_T, the half-open symbols refers to the $X_{0.5}$

Because in the 3D boundary layer the external turbulence influence can't be reliably taken into account nowadays using the e^N-method, for comparison the experimental data, which were obtained at the low-turbulence external flow should be used. In these conditions the transition line on the swept wings has the serrated form, with critical *N*-factor values for other

values of spanwize co-ordinate possibly being slightly different from the obtained ones. However, the dimension of turbulent wedges in the longitudinal direction usually do not exceed 10% of the chord, giving rise to the error in definition of the critical *N*-factor values no more than ±1.

The calculated longitudinal distributions of the *N*-factor for the cases of angles of attack of $\alpha = 0$ and $-2°$ are shown in Figure 12.

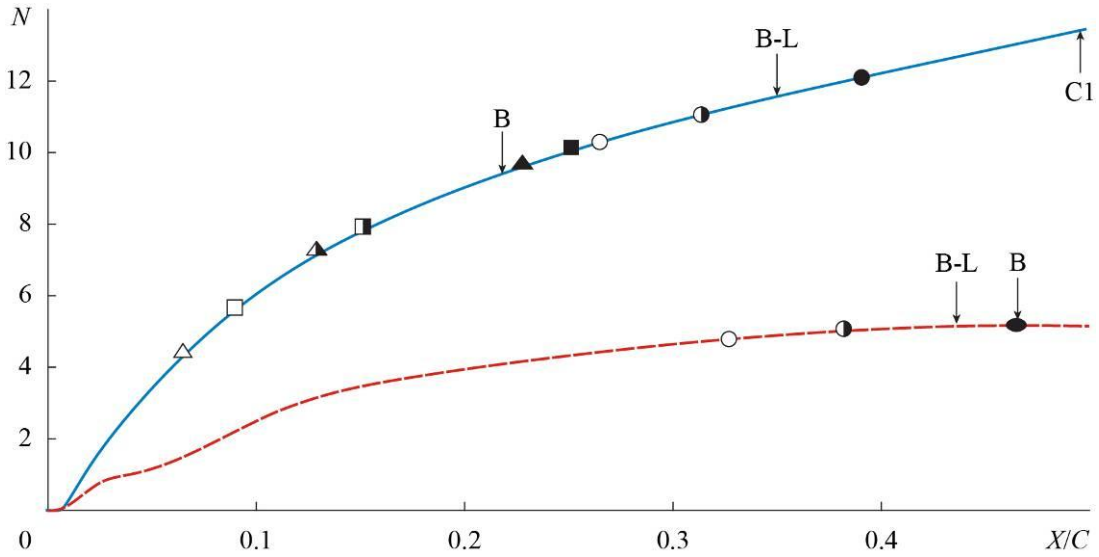


Figure 12: N-factor dependence from the longitudinal co-ordinate: solid line – $\alpha' = -2°$, dashed line – $\alpha' = 0$, ○ – natural condition, $Tu = 0.064\%$, □ – $Tu = 0.61\%$, △ – $Tu = 0.91\%$. Open symbols show the transition onset, filled symbols – points, where $\gamma = 0.99$, semi-filled – position $\gamma = 0.5$. Shooters showed the transition location defined by means of criteria C1, Brown (B) and Barinov-Lutovinov (B-L).

The specific points obtained at different experimental conditions are also marked on these curves. Figure 12 demonstrates that at zero angle of attack the calculated *N*-factor values are substantially lower than at $\alpha = -2°$. It is explained by the fact that according to calculations the cross-flow intensity at $\alpha = 0$ is approximately two times lower than at $\alpha = -2°$. At $X/C = 0.2$ typical values of the maximum cross-flow velocity in the co-ordinate system associated with the inviscid flow streamline are equal to $0.023U_0$ and $0.042U_0$ respectively for these two cases. *N*-factor value corresponding to $\gamma = 0.5$ for low-turbulence external flow in the case of $\alpha' = -2°$ was 11.1. This value practically coincides with the well-known data [22] for the tests in the low-turbulence wind tunnels in the presence of cross-flow instability. However, for the same external conditions at $\alpha = 0$ the *N*-factor value corresponding to $\gamma = 0.5$ appeared to be only 5. This value is substantially lower than *N*-factors, obtained in the condition of the increased turbulence at $\alpha = -2°$, where they were 7.9 for $Tu = 0.61\%$ and 7.1 for $Tu = 0.91\%$ (see Figure 12). It should be noted especially that in all calculation cases the cross-flow instability waves has become the most quickly rising ones. In both the experiment and calculations the evidence of the longitudinal instability (like the Tollmien-Schlichting waves) was unimportant. So, it may be concluded that contrary to the 2D flows, for the swept wing boundary layer the semi-empirical e^N-method (in the variant of the envelope method) gives substantial scatter of the critical *N*-factor values and does not permit to correlate reliably the transition data. In the papers [22, 23] similar conclusions were drawn also for other variants of e^N-method. The reason of this phenomenon is dominating of the non-linear effects in the transition processes in the 3D boundary layer even at low level of the external turbulence.

The results of implementation of the criteria [13, 17,18] for definition of the transition location in the conditions of the described experiments are also shown in Figure 12. It is shown, that C1 criterion [13] at $\alpha = -2°$ gives shifted downstream transition location, and at $\alpha = 0$ this criterion predicts absence of transition up to the separation line in the adverse pressure gradient zone. Brown criterion [17] has scatter of about 10% of the chord length relative to $\gamma = 0.5$ point in both positive and negative. The best results were demonstrated by the criterion of Barinov and Lutovinov [18], which in both cases predicts the transition location shifted downstream from the point $\gamma = 0.5$ by approximately 5% of the chord, i.e. corresponding to $\gamma \approx 0.8$.

The important findings of the analyses [5, 6] and [14] can be briefly summarized as following:

- The empirical and semi-empirical models, allowing estimate the transition location basing only on the averaged flow characteristics are still very robust and effective means for laminar-turbulent transition prediction at elevated free-stream turbulence level, at least in 2D boundary layers. Some of empirical criteria permit to evaluate the transition location on the swept wings, but they can't be applied at high free-stream turbulence level.
- The e^N-method rather successfully correlates with laminar-turbulent transition for 2D boundary layer; however, the values of *N*-factor corresponding to transition are substantially different for various external conditions. Contrary to the 2D flows, for the swept wing boundary layer the semi-empirical e^N -method gives substantial scatter of the critical *N*-factor values and does not permit to correlate reliably the transition data. The reason of this phenomenon is dominating of the non-linear effects in the transition processes in the 3D boundary layer even at low level of the external turbulence.
- In 3D boundary layers at some specific combinations of the pressure distribution and free-stream turbulence-level the very uncommon ways to transition could be existed. So, it was found that the transition location on the swept wing at high free-stream turbulence shifts downstream compared with the straight wing tests. This fact contrasts sharply with the available data concerning the sweep angle influence on the transition location at the low-turbulence external flow. This fact must be accounted for careful planning of experiments in industrial wind tunnels.

4 STUDIES OF SUCTION EFFECT ON THE LAMINAR-TURBULENT TRANSITION ON THE LOW-SWEPT WINGS WITH SR-0012 PROFILE

4.1 Experimental investigations

The experiments [4] were conducted on a wing at zero angle of attack and values of freestream speed from 30 to 90 m/s. For this speed range the free-stream turbulence level changes from 0.03 to a 0.07%.

The influence of suction on flow regime in the boundary layer was investigated on the upper surface of a wing with the following characteristics: symmetrical airfoil section, sweep angle $\chi = 15°$, thickness-to-chord ratio 12%, chord C= 800 mm, span L = 1 m (equal to the test section width). Three slots were made on the wing upper surface with length and width l_0 = 535 mm and $h \sim 0.15 \div 0.2$ mm, respectively. The Reynolds number of the slot is defined as $\mathrm{Re}_{i,s} = v_w h/\nu$, where v_w – average speed of flowing in a slot, ν – kinematic viscosity coefficient, and, the full air expense through a slot $Q_i = v_w h l_0$, i – number of slot.

The view of the model in the test section of T-124 WT is presented at Figure 13. The scheme of the suction slots location on the wing is shown at Figure 14.



Figure 13: The photo of model in test section of T-124 WT

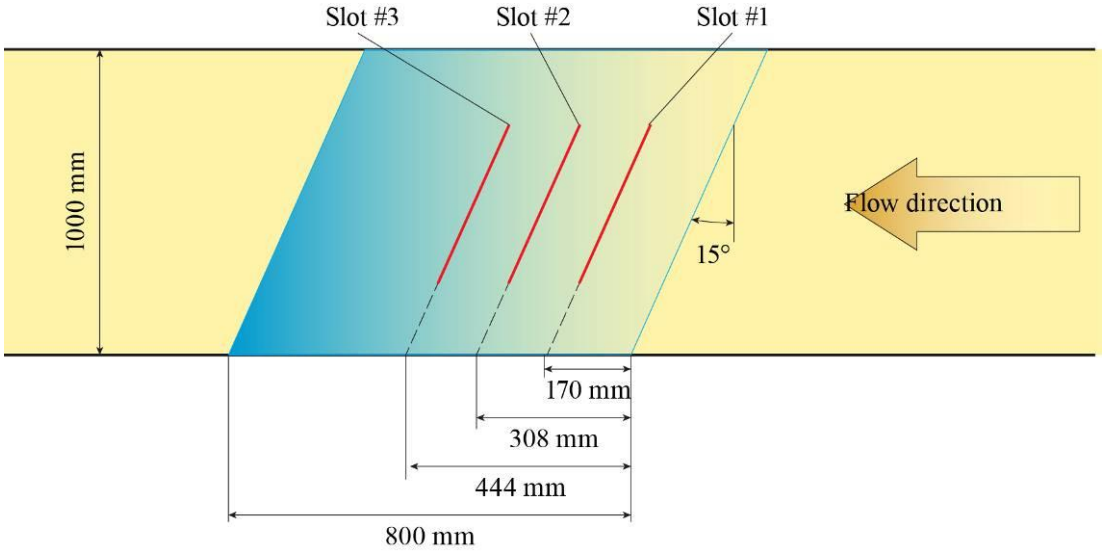


Figure 14: Scheme of the suction slots location on the wing

The wing surface was carefully polished. The wing's construction provided autonomous and spanwise uniform air suction through any of the slots.

Positioning of laminar-turbulent transition at the upper surface of the wing was performed at the same speed V_∞ = 40 m/s. Identification of flow regimes in the boundary layer on the model was carried out with the help of three independent methods: with kaolin coating on the lover surface (i.e., in the absence of suction), and on the opposite surface using a modified razor-blade technique and by measuring the variation of the intensity of the longitudinal velocity fluctuations ε near wall with a hot-wire anemometer and remotely-controlled traversing gear.

The razor-blade technique turned out to be effective in determining transition zone as well. The coordinates of the beginning of the transition zone, $\overline{X}_t = X_t/C$, obtained with hot-wire anemometer are presented in Figure 15.

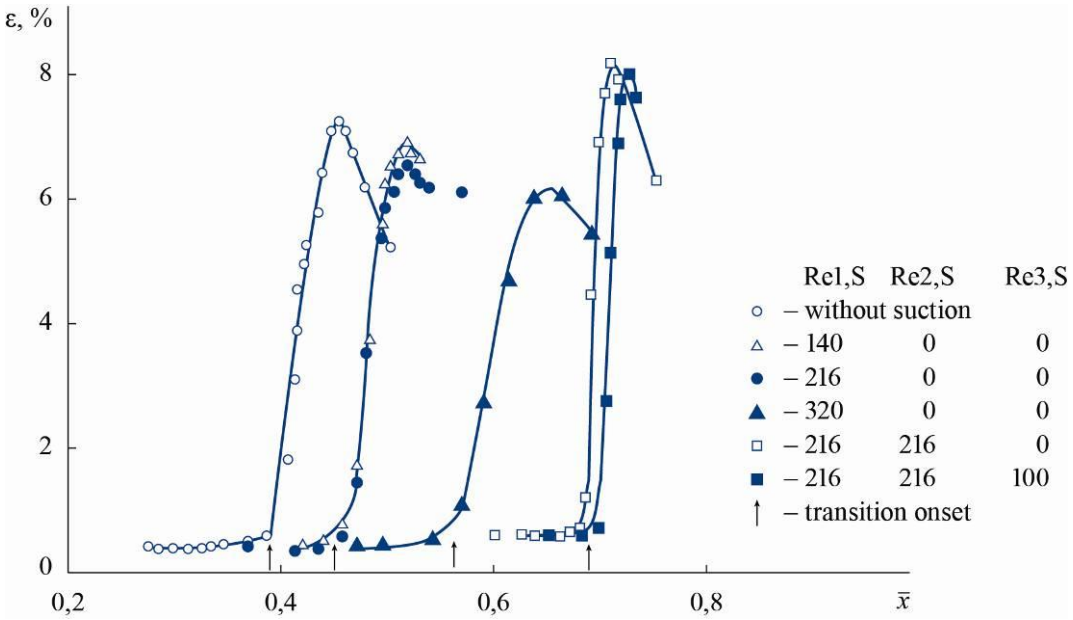


Figure 15: Measured variations of the intensity of the longitudinal velocity fluctuations ε near wall

According to this data, $\overline{X}_t = 0.392$ in the case when the suction is absent. The increase in the suction intensity through the first slot to $Re_{1,s}$ = 216, did not practically affect the transition onset position: the coordinate $\overline{X}_t$ turned out to be 0.45. But having increased $Re_{1,s}$ to 320, this point moved to $\overline{X}_t = 0.56$. The simultaneous suction from the first and second slots at $Re_{1,s} = Re_{2,s}$ = 216 resulted in displacing the transition onset position to $\overline{X}_t = 0.685$. The addition of suction through the third slot did not practically influence the transition onset position: in this case $\overline{X}_t$ was equal to 0.69. It should be born in mind that in the two latter cases transition was caused by the separation of the boundary layer which cannot be prevented with suction.

The influence of suction on velocity profiles and velocity fluctuations was investigated as well.

The air expense through each slot was chosen on the basis of the single principle: it must be sufficient to provide laminar flow behind the streamwise next slot. Thus, the optimum value of Q_1 through the first slot was considered to be that at which the chordwise distribution of ε reached a minimum just behind the second slot. The air expense through the second slot was determined from similar considerations. The suction scheme described has allowed the extent of the boundary layer's laminar part to be increased from 40 to 70 percent of the chord.

4.2 Numerical simulation of laminar - turbulent transition with considering boundary layer suction

Calculations were performed using unstructured computational mesh (Figure 16), consisting of 51.3 million elements and 18.8 million nodes. In the boundary layer zone a special finite-difference mesh consisting of 50 layers of prismatic elements was created. The thickness of the first prismatic layer was 0.005 mm (relative height of the first cell of the computational grid $y^+ \leq 1$). Total number of prismatic elements is equal to 30 million.

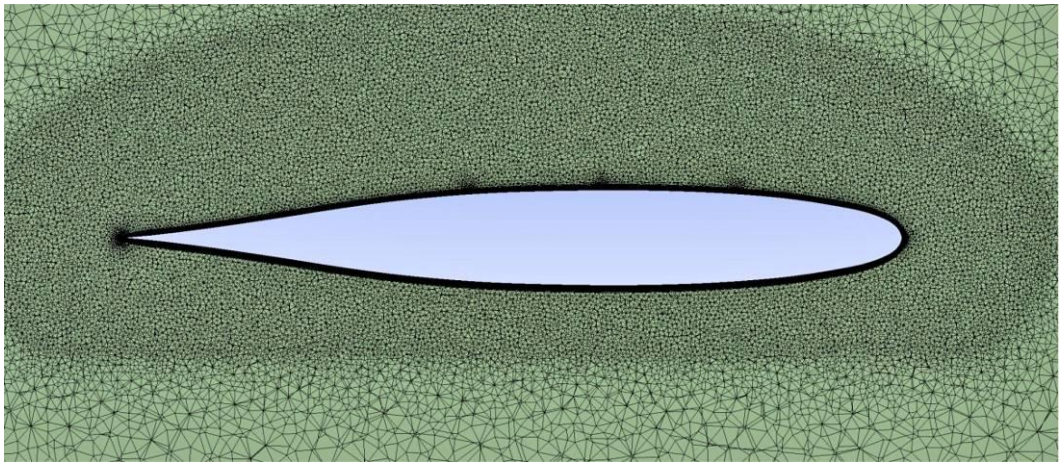


Figure 16: SR0012 wing CFD mesh

The following cases of flow over a wing were considered both in calculations and experiment:

- Case No.1 – flow without suction;
- Case No.2 – suction from the first slot at $\mathrm{Re}_{1,s} = 140$;
- Case No.3 – suction from the first slot at $\mathrm{Re}_{1,s} = 216$;
- Case No.4 – suction from the first slot at $\mathrm{Re}_{1,s} = 320$;
- Case t No.5 – suction from the first and second slots at $\mathrm{Re}_{1,s} = \mathrm{Re}_{2,s} = 216$;
- Case No.6 – suction from the through all three slots at $\mathrm{Re}_{1,s} = \mathrm{Re}_{2,s} = 216$; $\mathrm{Re}_{3,s} = 100$

Influence of boundary layer suction on the position of the laminar-turbulent transition was calculated using two models of turbulence – γ-Re_θ and γ- models. The calculated and experimental values of the coordinates of the transition beginning in the central section of the wing, depending on the case of the suction of the boundary layer, are shown in Figure 17.

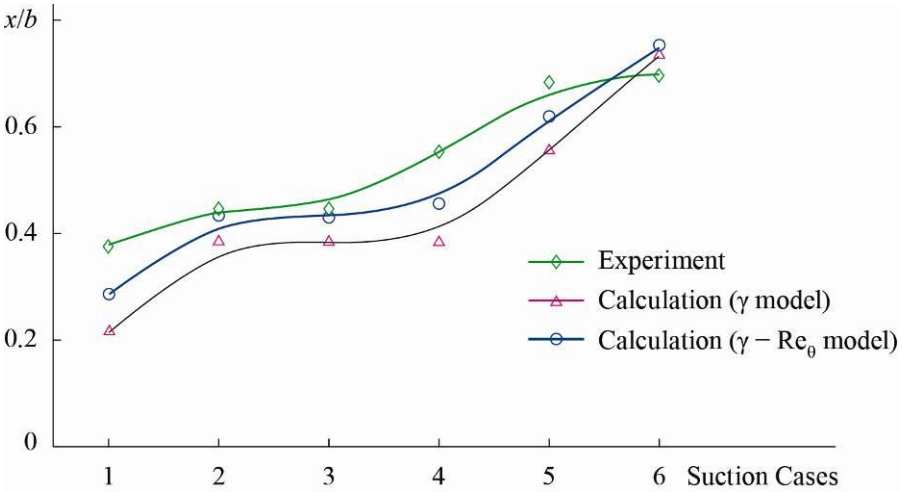


Figure 17: The position of transition in the central section of the wing, depending on the case of the suction

It can be noted that both models provide an acceptable agreement with the experiment. Langtry and Menter $\gamma - \mathrm{Re}_\theta$ model shows the results closer to the experimental ones. Compared with the results of measurements CFD-calculation predicts earlier transition. With both models calculations were performed using standard values of the model coefficients. Perhaps

the selection of model coefficients can adapt this model for better simulation of the experiment described above, but these studies have not been conducted in the framework of this study.

5 CONCLUSIONS

- The detailed experimental investigations of the transition zone on the straight and swept wings were carried out in the low-turbulence T-124 TsAGI wind tunnel. The influences of the incoming flow turbulence, acoustic disturbances and local boundary layer suction were studied. The validation of various transition prediction methods by comparison of experimental and calculated data obtained under the simulated influence of the above-mentioned factors was performed.
- The possibilities of application of different turbulence models for prediction of the laminar-turbulent transition and for investigation of the increased free stream disturbances and boundary layer suction effects on the transition characteristics in 2D and 3D flow are analyzed.
- The Walters-Cokljat and Langtry-Menter $\gamma - \mathrm{Re}_\theta$ turbulence models enable, in principle, to predict the laminar-turbulent transition in the cases when the cross-flow instability is insignificant. Compared with the transition model of Walters-Cokljat based on the transport equation for laminar and turbulent kinetic energy, $\gamma - \mathrm{Re}_\theta$ model is more versatile since it allows calculate the intermittency in the transition zone.
- The possibility of taking into account the cross-flow instability is realized in new Local-Correlation-Based Transition Model (γ - model). This model solves only one transport equation for the intermittency γ and allows the detection of cross-flow instability with the use of the Arnal C1-criterion.
- Both $\gamma - \mathrm{Re}_\theta$ and γ - model give acceptable results when modeling the influence of the increased level of free-stream turbulence and boundary-layer suction on the laminar-turbulent transition on the straight and swept wings. However, for application in specific conditions the fine tuning of the models is required. So additional researches are necessary to define the operational limits for each of the laminar-turbulence transition models to develop recommendations for their practical use.
- The empirical and semi-empirical models, allowing estimate the location of transition basing only on the averaged flow characteristics are still very robust and effective means for laminar-turbulent transition prediction at elevated free-stream turbulence level, at least in 2D boundary layers. Some of empirical criteria permit to calculate the location of transition on the swept wings, but they can't be applied at high level of free-stream turbulence.
- The e^N-method rather successfully correlates with laminar-turbulent transition for 2D boundary layer; however, the values of N-factor corresponding to transition are substantially different for various external conditions. Contrary to the 2D flows, for the swept wing boundary layer the semi-empirical e^N -method gives substantial scatter of the critical N-factor values and does not permit to correlate reliably the transition data. The reason of this phenomenon is dominating of the non-linear effects in the transition processes in the 3D boundary layer even at low level of the external turbulence.

REFERENCES

[1] S.L. Chernyshev, A.I. Ivanov, A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, S.V. Zhigulev, Investigations of influence of free stream turbulence and acoustic disturbances on the laminar-turbulent transition on LV6 airfoil and swept wing models. *AIAA Paper 2011-210*, 18 p., 2011.

[2] V.A. Vlasov, S.V. Zhigulev, A.I. Ivanov, A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, S.L. Chernyshev, Laminar-turbulent transition on the LV6 laminarized airfoil: natural transition. *TsAGI Science Journal*, **42**, No. 5, 565-591, 2011.

[3] V.A. Vlasov, S.V. Zhigulev, A.I. Ivanov, A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, S.L. Chernyshev, Laminar-turbulent transition on the LV6 laminarized airfoil. Part II: effect of free stream disturbances. *TsAGI Science Journal*, **42**, No. 6, 729-756, 2011.

[4] O.V. Babich, P.P. Vorotnikov, A.Ph. Kiselev, L.L. Chernyshev, S.L. Chernyshev, Laminar flow control on low-sweep wing. *Book abstracts of International Conference on Fundamental Research in Aerospace Science (1st FRAS)*, Section 1, 6–9, Zhukovsky, Russia, September 22-24, 1994.

[5] S.L. Chernyshev, A.I. Ivanov, A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, L.L. Teperin , V.V. Vozhdaev, Comparison of the laminar-turbulent transition prediction using different methods with the laminar wing test results. E. O˜nate, J. Oliver and A. Huerta eds. *Proceedings of 11th World Congress on Computational Mechanics (WCCM) / XI 5th European Conference on Computational Mechanics (ECCM V) / 6th European Conference on Computational Fluid Dynamics (ECFD VI)*, **5**, 6310-6321, Barcelona, Spain, July 20 -25, 2014.

[6] S.L. Chernyshev, A.I. Ivanov, A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, The free-stream turbulence effect on the laminar-turbulent transition in the swept wing boundary layer. E. O˜nate, J. Oliver and A. Huerta eds. *Proceedings of 11th World Congress on Computational Mechanics (WCCM XI) / 5th European Conference on Computational Mechanics (ECCM V) / 6th European Conference on Computational Fluid Dynamics (ECFD VI)*, **6**, 7511-7520, Barcelona, Spain, July 20 -25, 2014.

[7] R. Langtry, F.R. Menter, Correlation-based transition modeling for unstructured parallelized computational fluid dynamics codes, *AIAA Journal*, **47**, No. 12, 2894-2906, 2009.

[8] D.K. Walters, D. Cokljat, A three-equation eddy-viscosity model for Reynolds-averaged Navier-Stokes simulations of transitional flows, *Journal of Fluids Engineering*, **130**, No. 12, 121401–1 – 121401–14, 2008.

[9] F.R. Menter, P. Smirnov, Laminar-Turbulent Transition Modelling based on a New Intermittency Model Formulation. *11th World Congress on Computational Mechanics (WCCM XI) / 5th European Conference on Computational Mechanics (ECCM V) / 6th European Conference on Computational Fluid Dynamics (ECFD VI)*, Barcelona, Spain, July 20 -25, 2014.

[10] V.M. Filippov, Characteristics of fluctuations in flow through low-turdulence arerodynamic wind tunnel T-124 designed for small speeds. *Uchenye Zapiski TsAGI.* **39**, No.1-2, 68-80, 2008 (in Russian).

[11] R. Narasimha, The laminar-turbulent transition zone in the boundary layer. *Progress in Aerospace Sciences*, **22**, No. 1, 29-80, 1985.

[12] R.E. Mayle, The role of laminar-turbulent transition in gas turbine engines. *Journal of Turbomachinery*, **113**, No. 4, 509-537, 1991.

[13] D. Arnal, M. Habiballah, E. Coustols, Theorie de l'instabilite laminaire et criteres de transition en ecoulement bi et tridimensionnel. *La Recherche Aerospatiale*, No. 2, 125-143, 1984.

[14] A.Ph. Kiselev, V.A. Kuzminsky, D.S. Sboev, The laminar-turbulent transition zone in 2D and 3D boundary layers with emphasis on effect of free stream turbulence. *29th Congress of the International Council of the Aeronautical Sciences (ICAS 2014)*, St. Petersburg, Russia, 7-12 September, 10 p., 2014г.

[15] W.J. Solomon, G.J. Walker, J.P. Gostelow, Transition length prediction for flows with rapidly changing pressure gradients. *Journal of Turbomachinery*, **118**, No. 4, 744-751, 1996.

[16] S.K. Roberts, M.I. Yaras, Modeling transition in separated and attached boundary layers. *Journal of Turbomachinery*, **127**, No. 2, 402-411, 2005.

[17] W.B. Brown, A stability criterion for three-dimensional laminar boundary layers. in *Boundary layer and flow control.* **2**. Pergamon Press, 913-923, 1961.

[18] V.A. Barinov, V.M. Lutovinov, On the parameters of approximate relationship of critical Reynolds number in the three-dimensional boundary layer. *Uchenye Zapiski TsAGI*, **4**, No. 4, 27-32, 1973 (in Russian).

[19] M.V. Usinov, Statistical description of laminar-turbulent transition in a boundary layer at high freestream turbulence degree. *Fluid Dynamics*, **48**, No. 2, 192-200, 2013.

[20] V.A. Kuzminsky, Matrix numerical method of stability calculation of three-dimensional boundary layers. *Uchenye Zapiski TsAGI*. **38**, No. 3-4, 44-56, 2007 (*in Russian*).

[21] L.M. Mack, Transition prediction and linear stability theory. *AGARD CP*-224, 11-22, 1977.

[22] D. Arnal, Boundary layer transition: predictions based on linear theory. *Special course on progress in transition modelling*. AGARD Rep. No. 793, 2-1 - 2-63, 1994

[23] H. Bippes, Basic experiments on transition in three-dimensional boundary layers dominated by crossflow instability. *Progress of Aerospace Sciences*, **35**, No. 4, 363-412, 1999.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ASSESSMENT OF RANS AND DES METHODS FOR THE AHMED BODY

E. Guilmineau, G.B. Deng, A. Leroyer, P. Queutey, M. Visonneau and J. Wackers

LHEEA, CNRS UMR 6598, Ecole Centrale de Nantes
1 rue de la No, BP 92101, 44321 Nantes Cedex 3, France
e-mail: Emmanuel.Guilmineau@ec-nantes.fr

Keywords: DES models, IDDES, RANS, Ahmed body, Turbulence models.

Abstract. *This paper presents numerical simulations for the prediction of the flow around the Ahmed body, with the 25° and 35° slant angles, obtained with the flow solver ISIS-CFD. A RANS (Reynolds Averaged Navier-Stokes) turbulence model, as EARSM (Explicit Algebraic Stress Model) and two hybrid RANS-LES models, as DES (Detached Eddy Simulation) and IDDES (Improved Delay Detached Eddy Simulation) models, are used. All these turbulence models are based on the k-ω model.The flow for the slant angle 35° is well predicted by all turbulence models with a slight advantage for the IDDES model. However, for the flow with the slant angle 25° , only the IDDES hybrid RANS-LES model predict the recirculation bubble on the slant.*

1 INTRODUCTION

The external aerodynamics of a car determines many relevant aspects of an automobile such as stability, comfort and fuel consumption at high cruising speed [1]. The flow around vehicles is characterized by highly turbulent and three-dimensional separations. There is a growing need for more insight into the physical features of these dynamical flows, on one hand, and powerful numerical tools to analyze them on the other hand. Computations based on Reynolds-Averaged Navier-Stokes (RANS) equations are common in industry today. Although they are very successful in predicting many parts of the flow around a vehicle, they are unable to predict unsteadiness in the wake regions. The failure in predicting the base pressure is the major reason for large discrepancy in drag prediction between experiments and numerical simulations.

In an attempt to improve the predictive capabilities of turbulence models in highly separated regions, Spalart et al. [2] proposed a hybrid approach which combines features of classical RANS formulations with elements of Large Eddy Simulations (LES) method. This concept has been termed Detached Eddy Simulation (DES) and is based on the idea of covering the boundary layer by a RANS model and of switching the model to a LES mode in detached regions. Compared to classical LES methods, DES save orders of magnitude of computing power for high Reynolds number flows, due to the moderated costs of a RANS model in the boundary layer region, but still offers some of the advantage of a LES method in separated regions. A variant of the DES model, like Improved Delayed DES, IDDES, seems to be attractive [3].

The purpose of this paper is to conduct a validation of the flow around the Ahmed body [4] to compare RANS model and hybrid RANS-LES methodologies. The Ahmed body is a generic car geometry comprising a flat front with rounded corners and a sharp slanted rear surface. A detailed experimental study by Lienhart and Becker [5] has been performed. This was undertaken for the 25° and 35° slant angles. While representing a much simplified car, the Ahmed body provides many of the flow features of a realistic car, such as the large 3D separation region behind the car body and the roll up of the vortices at the rear corners. The wake behind the body is a complex interaction between the counter-rotating vortices and the highly turbulent recirculating flow. The structure of the wake and the reattachment point of the recirculation bubble depend on the angle of the rear slant. At 35°, the counter-rotating vortices are weaker, which results in the flow being completely separated over the entire slant. As the angle is reduced, the strength of the counter-rotating vortices is increased relative to the recirculating turbulent flow. At 25°, the flow is able to reattach over the slant.

Both 25° and 35° slant back angles are investigated in this study to assess the capability of each turbulence model to capture the important changes in flow physics. These turbulence models are used with the ISIS-CFD unstructured finite-volume solver.

2 TEST CASE

The Ahmed model used for the numerical simulations is shown in Figure 1. The slant angle is adjustable and is the main variable model-parameter in the experimental investigations [4]. In this paper, the 25° and 35° slant angles are investigated. The length of the model is L = 1044 mm, the width is W = 369 mm, the height is H = 288 mm. The length of the slant is l = 222 mm. The ground clearance is G = 50 mm, and the diameter of the four feet, which are used to secure the model to the floor of the wind tunnel, is ϕ = 30 mm. The reference axis (X,Y,Z) is linked to the model. The origin of these axes is related to the point O located on the floor of the wind tunnel at the base of the model ans in the symmetry plane of the model, see Figure 1.

The upstream velocity is U_∞ = 40 m/s. The Reynolds number, based on the length of the

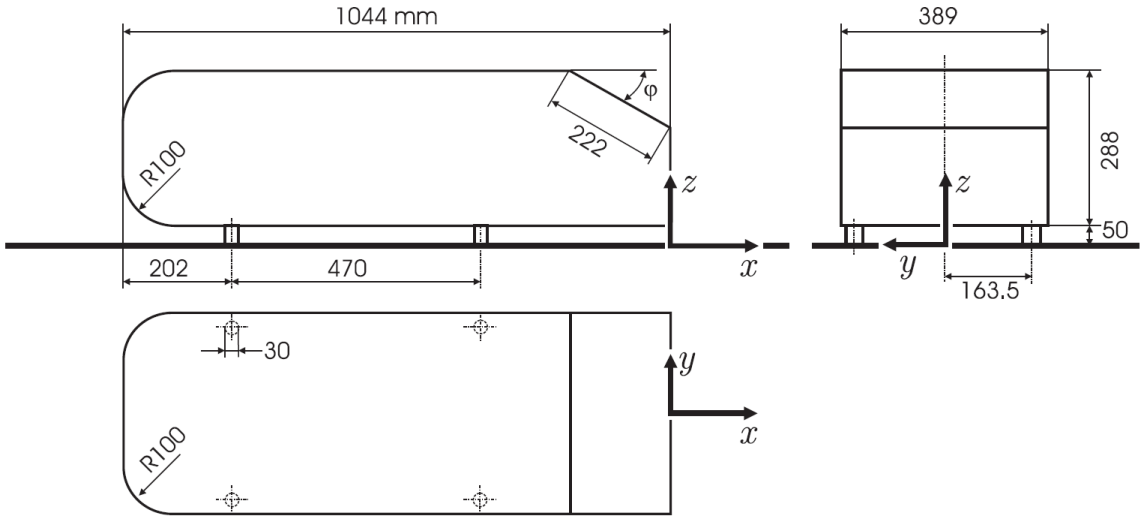


Figure 1: Ahmed body model with the coordinates system used. Distances are in mm.

model, is Re = 2.78×10^6. Measurements were performed by Lienhart and Becker [5].

3 FLOW SOLVER

ISIS-CFD, developed by the Ecole Centrale de Nantes and CNRS and available as a part of the FINE/Marine computing suite, is an incompressible unsteady Reynolds-averaged Navier-Stokes (URANS) method. The solver is based on the finite volume method to build the spatial discretization of the transport equations. The unstructured discretization is face-based, which means that cells with an arbitrary number of arbitrarily shaped faces are accepted. A second order backward difference scheme is used to discretize time. The solver can simulate both steady and unsteady flows. The velocity field is obtained from the momentum conservation equations and the pressure field is extracted from the mass equation constraint, or continuity equation, transformed into a pressure equation. In the case of turbulent flows, transport equations for the variables in the turbulence model are added to the discretization. A detailed description of the solver is given by Queutey and Visonneau [6].

The solver features sophisticated turbulence models: apart from the classical two-equation k-ε and k-ω models, the anisotropic two-equation Explicit Algebraic Reynolds Stress Model (EARSM), as well as Reynolds Stress Transport Models, are available, see Duvigneau et al. [7] and Deng and Visonneau [8]. All these are RANS models. A Detached Eddy Simulation (DES) approach has been introduced, see Guilmineau et al. [9]. Recently, some modifications of this formulation proposed by Griskevich et al. [10] includes recalibrated empirical constants in the shielding function and a simplification of the original Spalart-Allmaras-based formulation. This new model is called Improved Delayed Detached Eddy Simulation (IDDES).

4 NUMERICAL SIMULATION SET-UP

The computational domain starts 2L in front of the model and extends to 5L behind the model. The width of the domain is 1.87 m and its height is 1.4 m. These dimensions are recommended for the ERCOFTAC workshop on Refined Turbulence Modelling [11]. The mesh is generated using Hexpress, an automatic unstructured mesh generator. This software generates

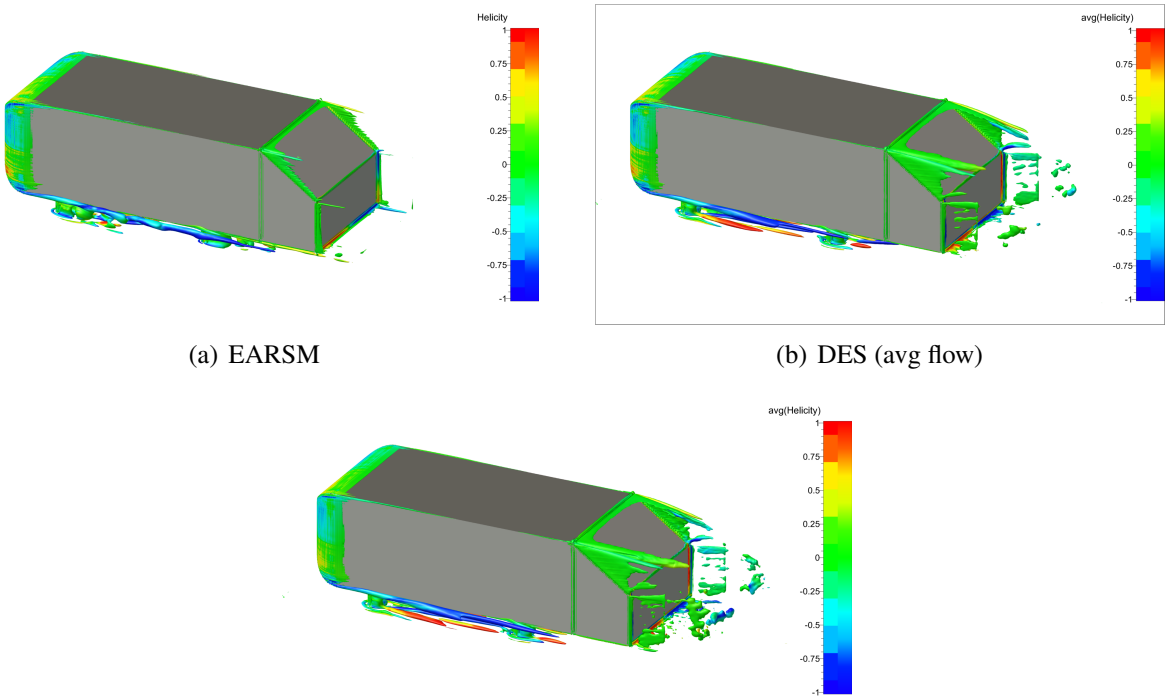
meshes containing only hexahedrals. For the surface of the car model and the floor, a no-slip boundary condition is used and the wall normal resolution is set 0.0007 mm, i.e; $y^+ \leq 0.7$. For the 25° slant angle, the mesh consists of 23.1×10^6 cells and the model is described by 384,090 faces. For the 35° slant angle, the mesh consists of 22.2×10^6 cells and the model is described by 379,358 faces.

To capture the unsteadiness of the flow, an unsteady simulation is carried out with the RANS turbulence models. In this case the time step is $\Delta t = 0.001$ s. The numerical simulation converge to a steady flow. With the hybrid RANS-LES models, the flow is by nature unsteady and the time step is $\Delta t = 2.5 \times 10^{-4}$ s. The averaging time, $t \times U_\infty / L$, in the simulations is 40.

5 RESULTS

5.1 35° slant angle

Figure 2 shows the vortex structures by means of the dimensionless Q-criterion for the Ahmed body with the 35° slant angle. For the hybrid RANS-LES models, the visualizations presents the mean flow. The vortex are more pronounced with the hybrid RANS-LES approaches than the RANS model.



(a) EARSM

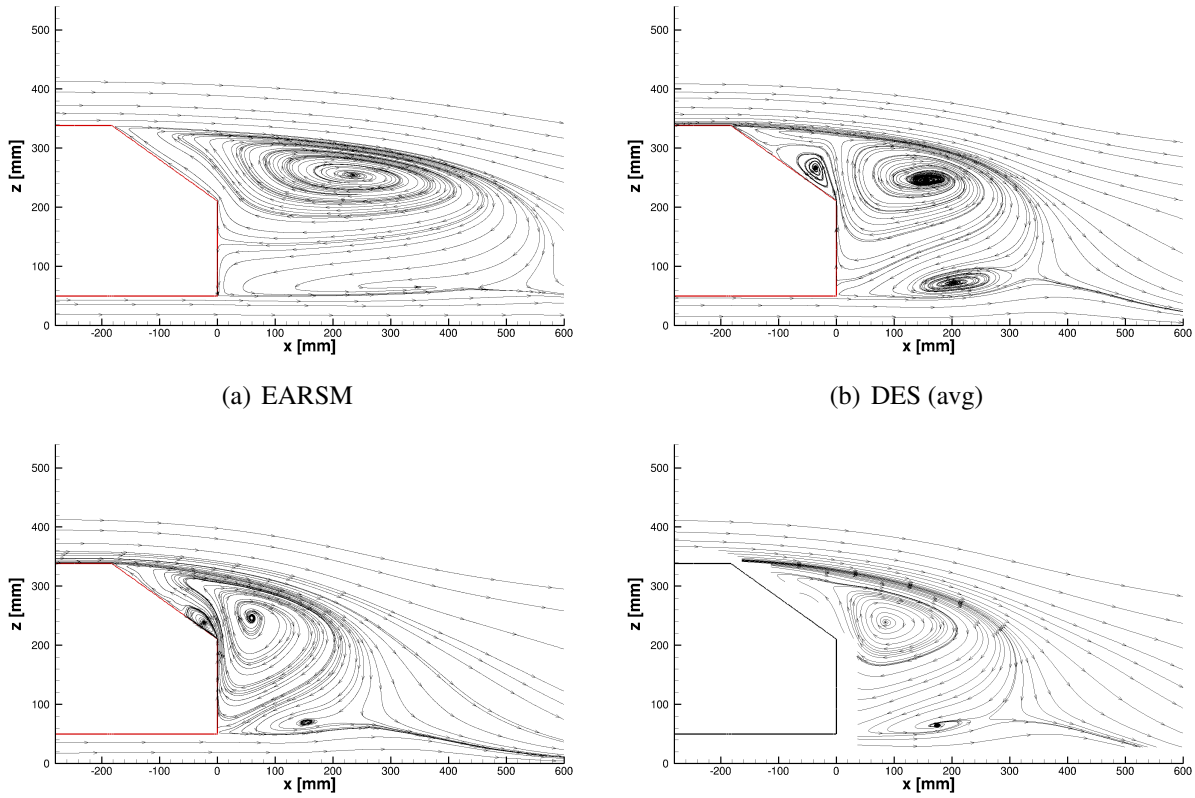
(b) DES (avg flow)

(c) IDDES (avg flow)

Figure 2: 35° slant angle - Vortex structures around the Ahmed body visualized by iso-surfaces of the non-dimensional Q-criterion (Q=50).

The flow in the symmetry plane obtained with all turbulence models is reported in Figure 3 together with the experimental results [5]. With the EARSM model, a massive separation in the wake is predicted while with the hybrid RANS-LES approaches, this recirculation is smaller and in better agreement with the measurements. With this turbulence modelization, a small separation is observed at the end of the slant. In the experiments, it is difficult to say if this

recirculation is present. The shape of the lowest recirculation in the wake obtained with the IDDES model is in better agreement with the experiments than this obtained with the DES approach.



(a) EARSM (b) DES (avg)

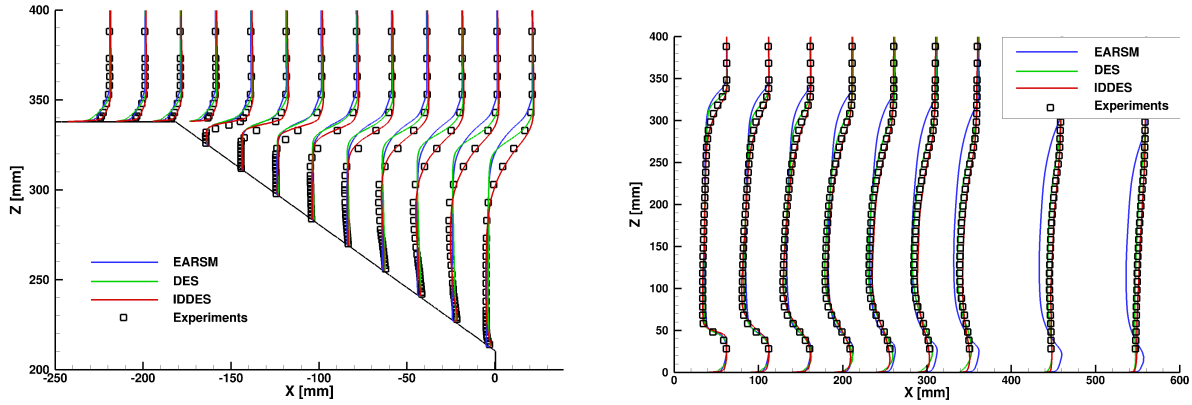
(c) IDDES (avg) (d) Experiments [5]

Figure 3: 35° slant angle - Streamlines in the symmetry plane Y = 0.

Figure 4 shows a comparison of the streamwise velocity in the symmetry plane for the Ahmed body with the 35° slant angle. The experimental profiles are those obtained by Lienhart et al. [12]. All turbulence models predict the correct flow, a fully separated flow, in the symmetry pane. The IDDES model gives a better agreement at the shear layer region.

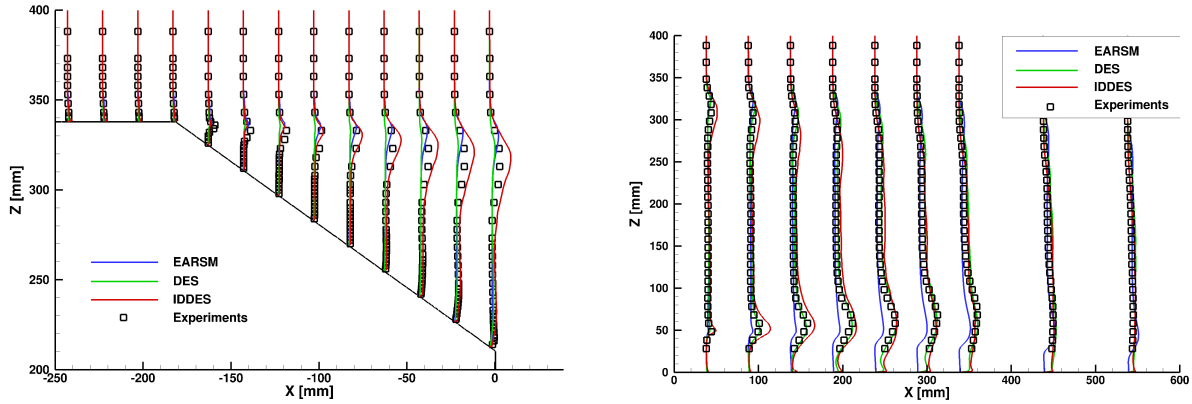
Figure 5 presents the turbulent kinetic energy (TKE) in the symmetry plane. Each model gives a good agreement with the experimental data. Even for a fully separated flow, a RANS model gives a good agreement with the experimental data. This is in agreement with many previous studies, including the ERCOFTAC workshops [13, 11]. At the shear layer, the IDDES model over-predicts TKE while the other models under-predict TKE. In the wake, a better agreement with the experimental data is obtained with the hybrid RANS-LES models.

A comparison of the drag coefficient versus the turbulence model used, with experimental measurements [4], is given in Table 1. The hybrid RANS-LES models overestimated the experimentally measured coefficient of drag. In experiments, it seems that the drag of the feet is not taken into account. For the RANS model, the drag of feet represents 10% of the drag while for the hybrid RANS-LES approaches, the percentage is 11%.



(a) Over the slant (b) In the wake

Figure 4: 35° slant angle - Streamwise velocity on the rear slant and in the wake.



(a) Over the slant (b) In the wake

Figure 5: 35° slant angle - Turbulent kinetic energy on the rear slant and in the wake.

5.2 25° slant angle

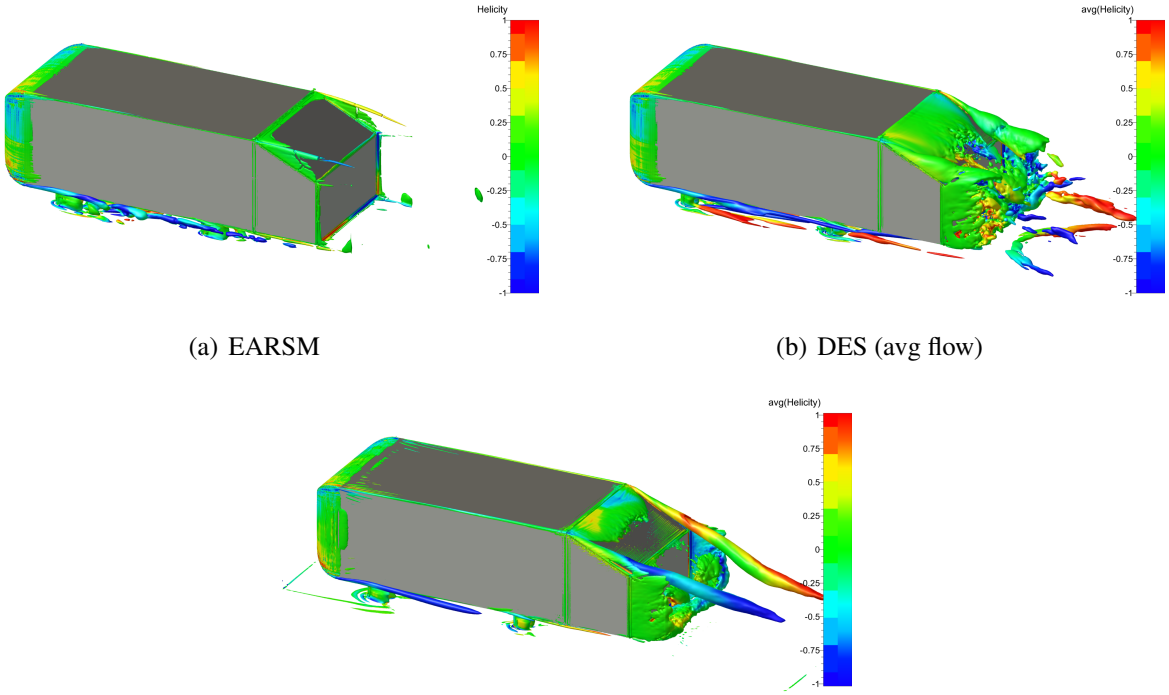
Figure 6 presents the vortex structure of the dimensionless Q-criterion for the Ahmed body with the 25° slant angle. With the RANS model, the C-pillar vortices are not predicted while with the IDDES approach, these vortices are very pronounced. With the hybrid RANS-LES models, a separation bubble is present on the slant.

Figure 7 shows the streamlines in the symmetry plane for all turbulence models compared to the experimental data. With the EARSM model, the results are similar to those of the previous slant angle, i.e. a massive separation. Thus, a RANS model does not see the modification of the geometry. With the hybrid RANS-LES models, the wake is different. The massive separation does not exist, only small separations are present. On the slant, a recirculation bubble is predicted with the IDDES model. The reattachment point is located at X = 168 mm, from the upstream edge of the slant, which is good agreement with the experimental measurement of Thacker [14] for which the value is X = 160 mm. With the DES model, the separation covers the whole slant.

Comparisons of the streamwise velocity profiles with measurements, in the symmetry plane, are presented in Figure 8. As expected, the EARSM turbulence models fails to describe cor-

EARSM	DES	IDDES	Experiments [4]
0.260	0.316	0.355	0.256

Table 1: 35° slant angle - Drag coefficient.



(a) EARSM

(b) DES (avg flow)

(c) IDDES (avg flow)

Figure 6: 25° slant angle - Vortex structures around the Ahmed body visualized by iso-surfaces of the non-dimensional Q-criterion (Q=50).

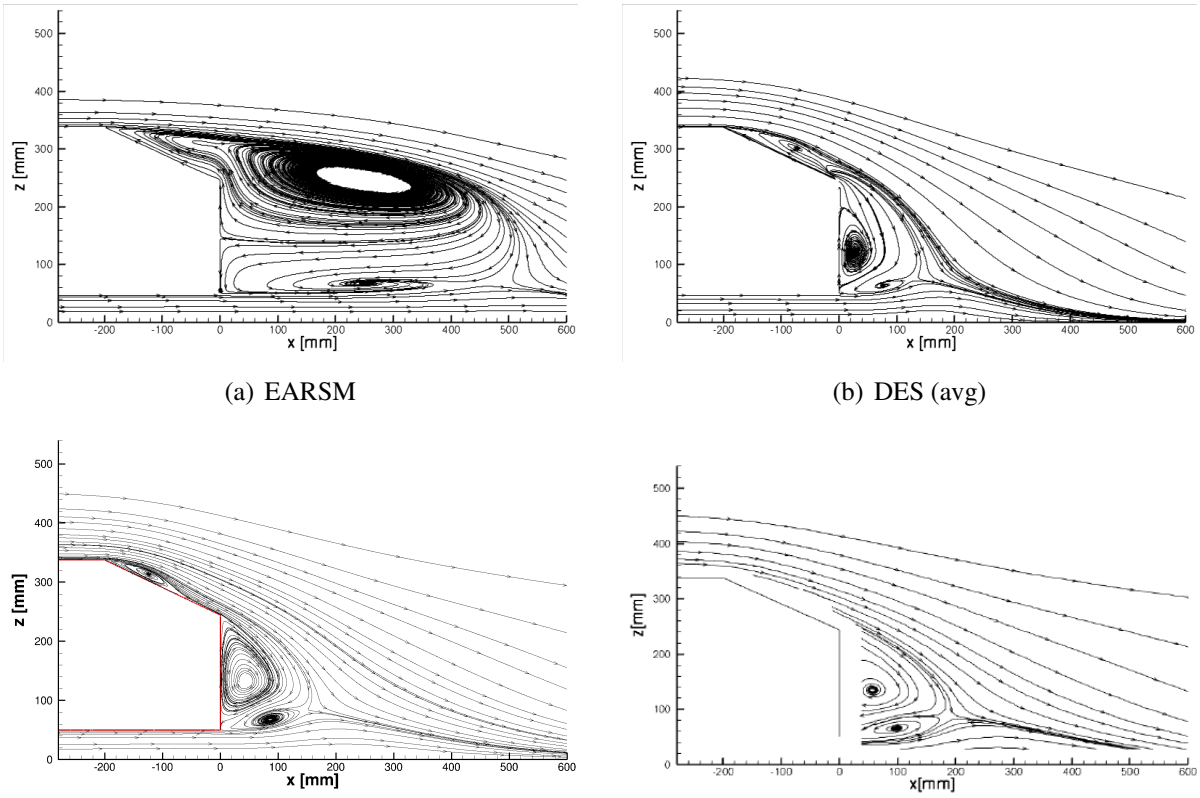
rectly the velocity field and overestimates the size of the recirculation in the wake. The hybrid RANS-LES models provided an improved prediction. In the wake, the results are very similar. On the slant, the IDDES model gives a better agreement due to the recirculation bubble which is correctly predicted.

A comparison of TKE is presented in Figure 9. The TKE obtained with the EARSM turbulence model is underestimated on the slant which means less turbulent mixing and thus a greater recirculation region. With the hybrid RANS-LES models, TKE is larger than the values obtained with the RANS model. In the initial separated region, TKE predicted with the IDDES model is in good agreement with the experimental value.

Table 2 presents a comparison of the drag coefficient. Compared to the previous slant angle, the drag coefficient have increased. The value of the drag is dependent of the turbulence model used. If we compare the drag coefficient measured by Thacker [14], the IDDES model gives the better agreement.

6 CONCLUSIONS

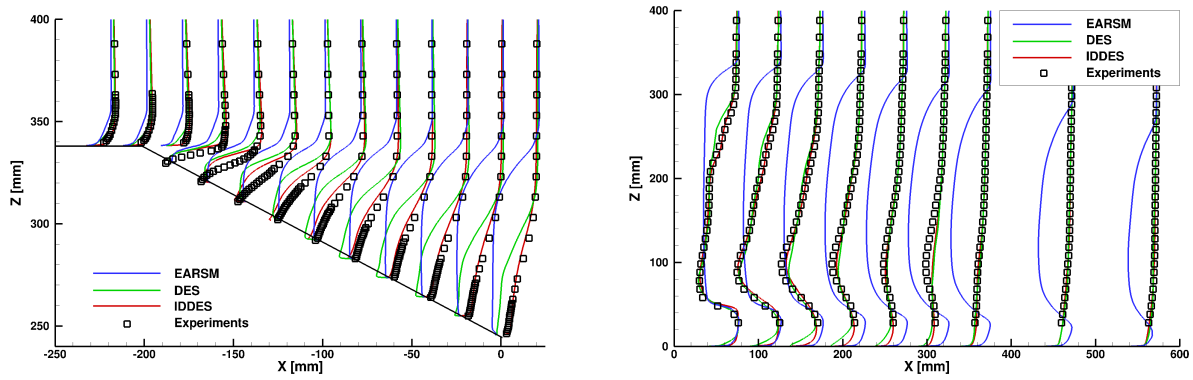
In this paper, an investigation of RANS and hybrid RANS-LES models for the Ahmed car body at 25° and 35° slant angle cases has been conducted. The RANS model used is the EARSM



(a) EARSM (b) DES (avg)

(c) IDDES (avg) (d) Experiments [5]

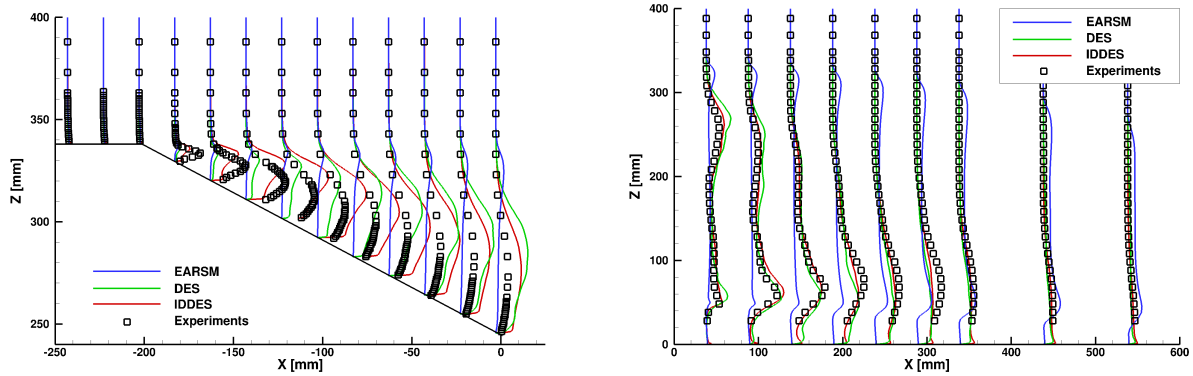
Figure 7: 25° slant angle - Streamlines in the symmetry plane Y = 0.



(a) Over the slant (b) In the wake

Figure 8: 25° slant angle - Streamwise velocity on the rear slant and in the wake.

turbulence model and the hybrid RANS-LES models are the DES model and the IDDES model. These both models are based on the k-ω SST model. It has been shown that the use of a hybrid RANS-LES model, in particular the IDDES model, offers an advantage over RANS model in terms of the force coefficients and general flow field, even for the 25° test case where the reattachment of the flow is correctly predicted. Only, the IDDES model predicts the flow correctly for both slant angles. With the RANS model, a massive separation is present for both



(a) Over the slant (b) In the wake

Figure 9: 25° slant angle - Turbulent kinetic energy on the rear slant and in the wake.

EARSM	DES	IDDES	Experiments [4]	Experiments [14]
0.280	0.437	0.382	0.285	0.384

Table 2: 25° slant angle - Drag coefficient.

slant angles.

ACKNOWLEDGMENT

This work was granted access to the HPC resources of CINES/IDRIS under the allocation 2015-2a0129 made by GENCI.

REFERENCES

[1] W.H. Hucho, *Aerodynamics of road vehicles*. SAE International, 1998.

[2] P. R. Spalart, W. Jou, M. Strelets, S. Allmaras, Comments on the feasibility of LES for wings and on a hybrid RANS/LES approach, *1st AFOSR International Conference on DNS/LES* (C. . Z. Liu, ed.), Advances in DNS/LES, Greyden Press, 1997.

[3] N. Ashton, A. West, S. Lardeau, A. Revell, Assessment of RANS and DES methods for realistic automotive models, *Computers and Fluids*, **128**, 1–15, 2016.

[4] S.R. Ahmed, G. Ramm, G. Faltin, Some salient features of the time-averaged ground vehicle wake. *SAE World Congress*, Paper 840300, 1984.

[5] H. Lienhart, S.Becker, Flow and turbulence in the wake of a simplified car model. *SAE World Congress*, Paper 2003-01-0656, 2003.

[6] P. Queutey, M. Visonneau, An interface capturing method for free-surface hydrodynamic flows, *Computers and Fluids*, **36**, 1481–1510, 2007.

[7] R. Duvigneau, M. Visonneau, G.B. Deng, On the role played by turbulence closures in hull ship optimization at model and full scale, *Journal of Marine Science and Technology*, **8**, 11–25, 2003.

[8] G.D. Deng, M. Visonneau, Comparison of explicit algebraic stress models and second-order turbulence closures for steady flow around ships, *7th Symposium on Numerical Ship Hydrodynamics*, Nantes, France, July 1999.

[9] E. Guilmineau, G.B. Deng, J. Wackers, Numerical simulation with a DES approach for automotive flows, *Journal of Fluids and Structures*, **27**, 807–816, 2011.

[10] M.S. Gritskevich, A.V. Garbaruk, J.Schütze, F. R. Menter, Development of DDES and IDDES formulations for the k-ω shear stress transport model, *Flow, Turbulence and Combustion*, **88**, 431–449, 2012.

[11] R. Manceau, J.Bonnet, eds., *10th joint ERCOFTAC/IAHR/QNET-CFD Workshop on Refined Turbulence Modelling*, Poitiers, France 2002.

[12] H. Lienhart, C. Stoots, S. Becker, Flow and turbulence structures in the wake of a simplified car model Ahmed model), *DGLR Fach Symposium der AG STAB*, Stuggart University, November 2000.

[13] S. Jakirlić, R. Jester-Zrker, C. Tropea, eds., *9th joint ERCOFTAC/IAHR/QNET-CFD Workshop on Refined Turbulence Modelling*, Darmstadt, Germany 2001.

[14] A. Thacker, *Contribution expérimentale à l'analyse stationnaire et instationnaire de l'écoulement à l'arrière d'un corps de faible allongement*. PhD thesis, Université d'Orléans, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A WAY TO IMPROVE JET MODELING WITHIN RANS EQUATION SYSTEM

Alexey Troshin[1,2]

[1]Central Aerohydrodynamic Institute (TsAGI)
1 Zhukovsky Street, Zhukovsky, Moscow Region, 140180, Russia
[2]Moscow Institute of Physics and Technology (MIPT)
9 Institutskiy per., Dolgoprudny, Moscow Region, 141700, Russia
e-mail: ai-troshin@yandex.ru

Keywords: Turbulence Model, Mixing Layer, Jet.

Abstract. *Performance of several standard turbulence models in predicting the flow field of a plane jet is analyzed. Jet potential core length is shown to be overestimated by all the models considered. To solve the problem, an additional source term in the turbulence characteristic frequency ω equation is used. It accounts the longitudinal flow inhomogeneity and entrainment which influence turbulence in jet mixing layers. In comparison to the earlier publications on this source term, a coefficient in its formulation has been slightly altered to better predict round jets. The modified SSG/LRR-ω differential Reynolds stress model is used to compute several test cases including free subsonic plane and round jets, a supersonic underexpanded free round jet, and a coaxial jet. In all the cases, improvements over the standard SSG/LRR-ω and SST models are shown. Further steps involving RANS turbulence models verification with high order of accuracy LES computations are discussed.*

1 INTRODUCTION

Today it is widely recognized that most turbulence models overestimate jet potential core length L_{ini}. In [1] L_{ini} overprediction is reported for $k-\varepsilon$ and SST models, in [2] the same is shown for an EARSM, and in [3] SSG/LRR-ω differential Reynolds stress model (DRSM) is found to give similar results. This issue can negatively influence the modeling of many problems involving civil aircraft propulsive jets, their interaction with nozzle walls and downstream structural elements of a plane. Several solutions of this problem have been proposed. In [1], L_{ini} is reduced by taking into account the increase in turbulent diffusion intensity near the jet axis due to acoustic interaction between different parts of the mixing layer; in [4], another modification of turbulent diffusion is proposed compatible with the concept of mixing layer self-similarity. In [3], the problem is further investigated and another approach to improve the jet potential core modeling is developed which is not connected to the amplification of the modeled turbulence diffusion. Instead, the influence of longitudinal mean flow inghomogeneity on turbulence statistics is analyzed and taken into account. This approach is followed in the present paper. Firstly, more computational results confirming the inadequacies of standard models in predicting L_{ini} are presented. After that, a minor update to one of the coefficients in the model proposed in [3] is made. Finally, new test cases computations using both modified and original models are reported.

The structure of the paper is as follows. In Section 2, the results of free plane jet computations using several standard turbulence models are compared, and common shortcomings of the solutions are discussed. In Section 3, a modification to the SSG/LRR-ω turbulence model is formulated and briefly commented. In Section 4, the computations of free subsonic plane and round jets, a supersonic underexpanded free round jet, and a coaxial jet are reported, and performance of the modified model is compared to the standard one and to the eddy viscosity SST model. Further steps in RANS turbulence models verification involving high order of accuracy LES computations are discussed in Section 5. The conclusions are made in Section 6.

2 STANDARD TURBULENCE MODELS

2.1 Solver and turbulence models

All the computations presented here are conducted using the EWT-TsAGI in-house code [5]. The solver uses structured multiblock hexahedral meshes. Hanging nodes at the block boundaries are allowed.

The following complete unsteady equation systems for the compressible air flow can be solved by the code:

- Euler equations;
- Navier–Stokes equations;
- Favre averaged Reynolds equations with one of the turbulence models listed below:
  - — Menter SST model [6];
  - — Coakley $q-\omega$ model [7];
  - — Spalart–Allmaras model [8];
  - — Wilcox Stress-ω DRSM [9];
  - — SSG/LRR-ω DRSM [10] and its modified version [3].

Second order of accuracy finite volume Godunov-Kolgan-Rodionov scheme is implemented in the solver. Explicit second order two step time marching (global, fractional, and local time steppings), first order backward Euler unconditionally stable implicit time marching (global and local time steppings), and second order dual time stepping are available.

The following notes concerning the numerical method need to be made. First, turbulence variables at a face of a cell are reconstructed with the same method (using van Leer limiter by default) as main variables thus maintaining the same accuracy order for all the equations. Second, exact iterative Godunov Riemann solver is used (turbulence variables are treated as passive scalars). To our experience, these features, aimed at increasing the accuracy of the method, generally do not degrade convergence.

Since the test cases reported here are steady, implicit scheme has been used in the computations.

2.2 Test case specification

To demonstrate the performance of the standard turbulence models, computations of a free subsonic plane jet of a cold air have been performed. Nozzle width h based Reynolds number $\mathrm{Re}_h = u_0 h/\nu$ is 6.7×10^6, Mach number is 0.30. At the nozzle exit, approximately top hat velocity profile forms with thin turbulent boundary layers (each boundary layer has relative $0.99u_0$-velocity width δ_{99}/h of 0.014). The primary role of these boundary layers is to supply the initial mixing layers with sufficient turbulence level for smooth self-similar development. On the other hand, both turbulence intensity and turbulent viscosity ratio in the potential core flow has been set negligibly small ($\langle u' \rangle / u_0 \sim \sqrt{k}/u_0 \approx 10^{-5}$, $\omega h/u_0 \sim 10^{-3}$, $\nu_t/\nu \sim 0.6$).

Computational domain scheme and boundary conditions (BCs) are shown in Figure 1 (scaling is not proportional). Soft Riemann invariants based BCs are specified on the left boundary (above the jet flow region) and on the upper half of the right boundary. On the lower part of the right boundary, extrapolation BCs are used to better predict the outflow of the jet. Symmetry plane is set at the top (to avoid incorrect entrainment flow patterns) and the botoom (which is jet centerplane). External boundaries are placed $200h$ away from the jet centerplane. No-slip BCs are set on the nozzle wall which length is equal to h. Ambient outer flow with relative velocity $u_\infty/u_0 = 0.01$ is specified in order to avoid instabilities due to interaction of soft BCs with stagnant air.

There are several published sets of experimental data on free subsonic plane jets. In this paper, data on velocity distributions along the centerline are taken from [11, 12, 13, 14]. According to them, jet potential core length L_{ini} defined as the distance from the nozzle exit where centerplane velocity is $0.99u_0$, lies in the range $5h \le L_{\text{ini}} \le 6h$.

2.3 Computational meshes and mesh convergence study

A set of three nested meshes has been generated containing approximately 60000, 15000, and 4000 cells. An overview and nozzle region of the coarse mesh are presented in Figure 2. Hanging nodes outside the jet regions are clearly visible. They allow to significantly reduce the number of cells without affecting the accuracy of the results.

The fine mesh is designed to reproduce in details the mean flow in mixing layers (about 120 cells across the turbulent zone) and shear region downstream the initial region of the jet (about 140 cells per jet half-width). Nozzle boundary layer is resolved with 60 cells across it, the first cell height in wall units being of the order of 1.

To study the mesh convergence, computations with SST turbulence model have been per-

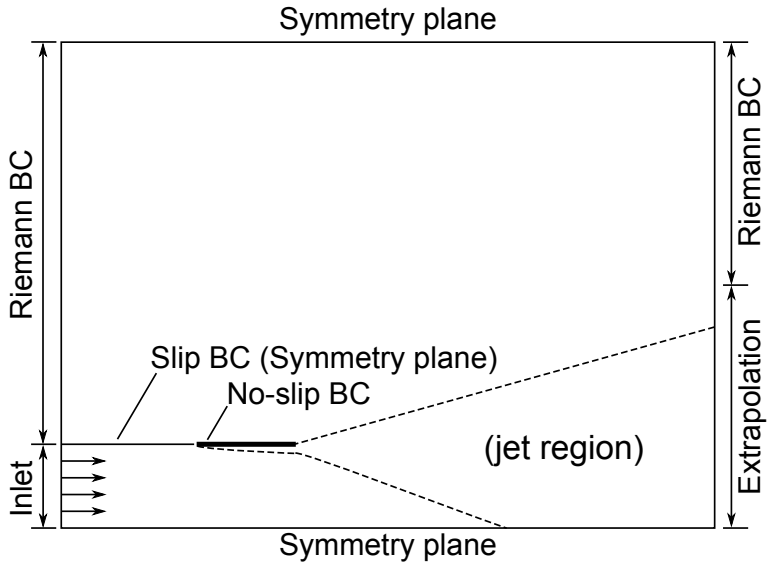


Figure 1: Computational domain scheme and boundary conditions of the plane jet test.

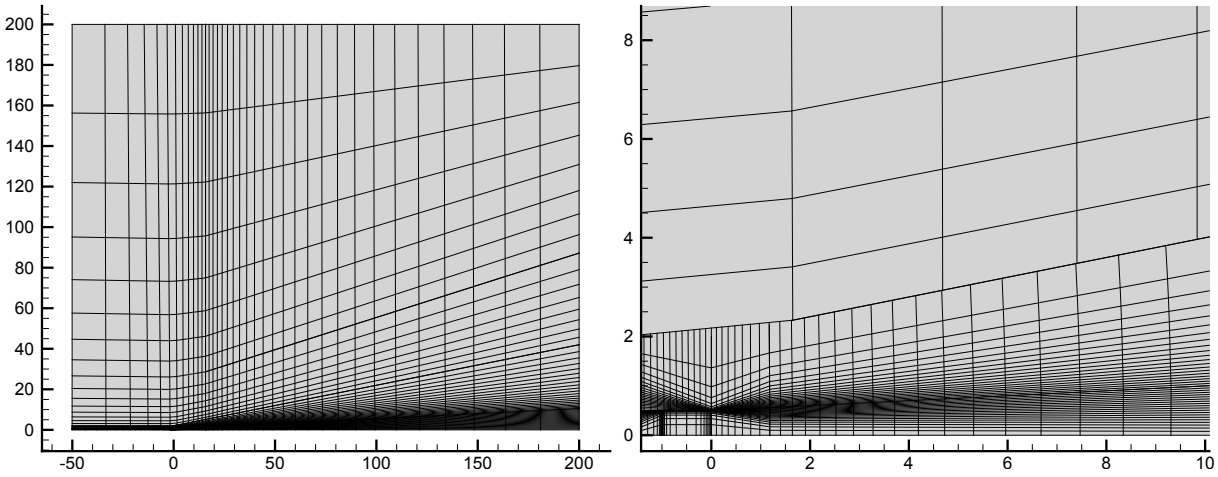


Figure 2: An overview (left) and nozzle region (right) of the coarse mesh for the plane jet test.

formed on each mesh. Resulting centerplane velocity and turbulent kinetic energy k distributions in the region $0 \leq x/h \leq 30$ are shown in Figure 3.

The difference in the velocity distributions is no more than 1.3% between the fine and medium meshes and no more than 2.6% between the medium and coarse meshes. Peak turbulent kinetic energies almost coincide, positions of k maxima differing within 0.1% and 5%, respectively. It is concluded that medium mesh is sufficient for further analysis in this paper. Another argument against the fine mesh is difficulties in obtaining the steady solution. Kelvin–Helmholtz type instability easily develops in the mixing layer, and to avoid it, it is necessary to start the computation with the first order numerical scheme, and after getting the steady field, proceed with second order scheme but with significantly limited time step. Even with these tricks, some travelling pressure waves persist is the mixing layers which slightly influence the centerplane velocity in the potential core. On the other hand, computations on the medium mesh converge to steady solutions smoothly.

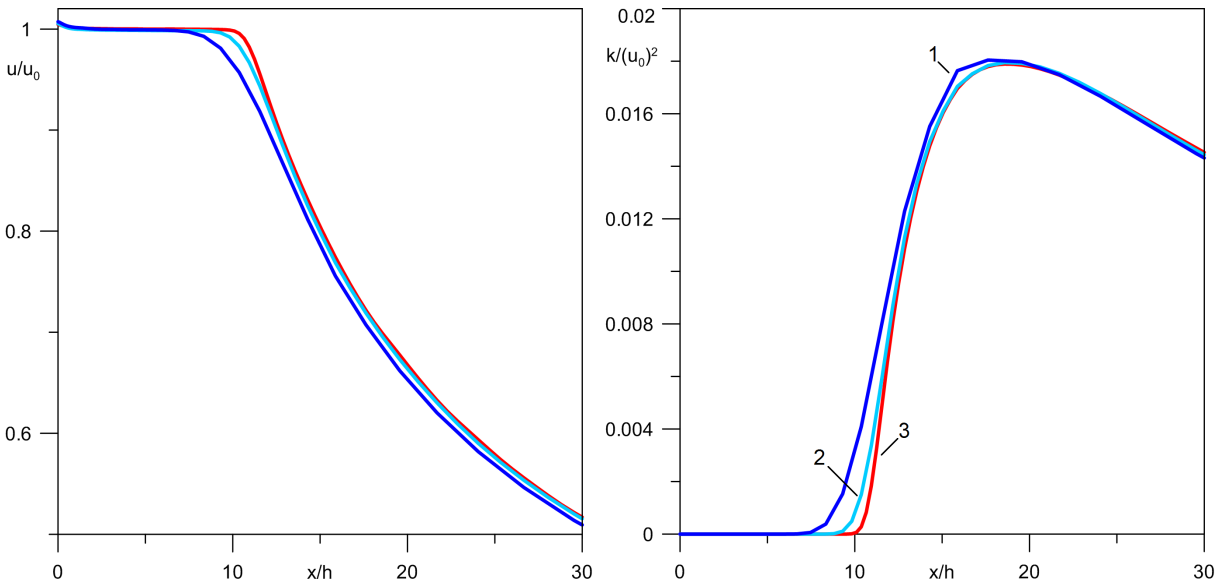


Figure 3: Centerplane velocity (left) and kinetic energy (right) distributions obtained on a set of nested meshes. 1 — coarse mesh, 2 — medium mesh, 3 — fine mesh.

2.4 Results comparison

After the mesh convergence study with SST model, computations with Spalart–Allmaras, Wilcox Stress-ω, and SSG/LRR-ω models have been performed. In computations with DRSMs, isotropic turbulence has been specified at inflow boundaries. Centerplane velocity distributions obtained in these computations are compared with experimental data in Figure 4.

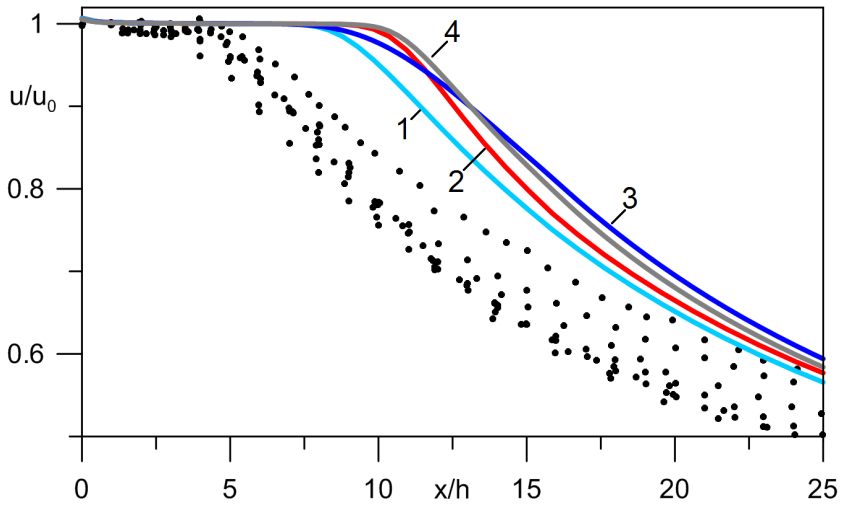


Figure 4: Centerplane velocity distributions obtained in plane jet test with different turbulence models. 1 — Spalart–Allmaras model, 2 — SST, 3 — Stress-ω, 4 — SSG/LRR-ω.

It is clearly seen that all the models overestimate jet potential core length L_{ini}. The least error in L_{ini} compared to the experimental data give Spalart–Allmaras model, the simplest among the models considered, and Stress-ω DRSM. They overpredict L_{ini} by 60%. The higher error is produced by SST and SSG/LRR-ω models (80%). At the same time, mixing layer width growth rate $dD_{0.1}/dx$ ($D_{0.1}$ is the distance between the points where $u = 0.1u_0$ and $u = 0.9u_0$) is

predicted much better and lies within $\pm 15\%$ around experimentally observed value for all the turbulence models considered. Obtained values are collected in Table 1.

Data	L_{ini}/b	$dD_{0.1}/dx$
experiments	$5-6$	$0.14-0.18$
Spalart–Allmaras	9	0.17
SST	10	0.15
Stress-ω	9	0.15
SSG/LRR-ω	10	0.14

Table 1: Plane jet potential core lengths and mixing layer width growth rates predicted by different turbulence models.

It turns out that different classes of turbulence models, from one equation eddy viscosity models to DRSMs, are affected by the same issue of overestimating L_{ini} while correctly predicting $dD_{0.1}/dx$. This suggests an idea that mixing layer velocity profile is distorted such that the whole turbulent zone is "rotated" outwards the jet centerplane. In the following Section, a modification to the turbulence characteristic frequency ω equation is described which fixes the mixing layer velocity profile and recovers the correct L_{ini} in computations.

3 MODIFIED TURBULENCE MODEL

The ideas of the modification are given here in brief and consecutive derivation of the formulas is omitted because these questions were the subjects of another paper [3] where they are described in details. The main goal of the current paper is to extend the number of test cases of the modified model and to more thoroughly assess its performance.

3.1 The idea of the modification

The first step of the SSG/LRR-ω model modification is recalibration of its coefficients in order to obtain as accurate description of mixing layers as possible. During the coefficients tuning it appears that temporal mixing layer velocity profile [15] is easy to reproduce, but single stream spatial mixing layer is a challenge. The process of tuning is described in [3]. The new "free stream" coefficient values using the designations of [10] are:

$$\alpha_\omega = 0.48, \quad \beta_\omega = 0.0774, \quad D = 0.20, \quad \sigma_\omega = 0.659, \quad \sigma_d = 0.373. \tag{1}$$

Velocity profiles in these two flows obtained after optimization of the coefficients are shown in Figure 5.

With the optimized coefficient set, temporal mixing layer fits the available data, but spatial mixing layer velocity profile is too sharp at the high velocity boundary of the mixing layer and too wide at the low velocity boundary. Indeed, this behavior is typical for most of the turbulence models. It leads to the idea that turbulence models do not take into account the influence of longitudinal flow inhomogeneity and entrainment effects in spatial mixing layer on turbulence. To describe these effects, a mechanism should be reproduced which increases the turbulent kinetic energy dissipation rate in the low velocity region of the spatial mixing layer and decreases it in the high velocity region by taking into account longitudinal gradients of different variables (velocity components u_i and turbulence variables ω, $\overline{u_i' u_j'}$, and its invariants).

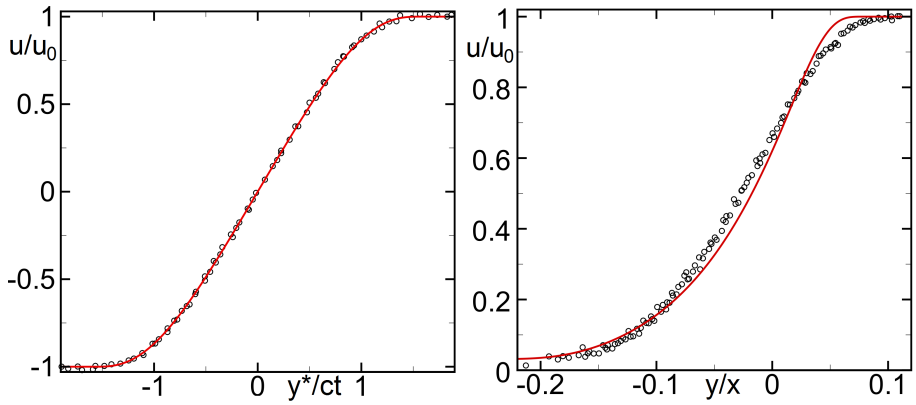


Figure 5: Velocity profiles in temporal (left) and single stream spatial (right) mixing layers after optimization of the SSG/LRR-ω model coefficients.

3.2 Modified ω equation

One possible form of taking into account the effects mentioned above is an additional source term I_ω in ω equation [3]. It can be written as

$$\frac{\partial \bar{\rho}\omega}{\partial t} + \frac{\partial}{\partial x_k}\left(\bar{\rho}\omega\tilde{u}_k\right) = S_{\bar{\rho}\omega} + \bar{\rho}I_\omega,$$

$$I_\omega = -C_{\omega 3}\ 0.03\,\mathrm{th}\left(\frac{2\tilde{\Omega}_{ij}}{0.03\,\omega^4}\frac{\partial\omega}{\partial x_i}\frac{\partial k}{\partial x_j}\right)\ \omega^2,$$

$$\tilde{\Omega}_{ij} = \frac{1}{2}\left(\frac{\partial\tilde{u}_i}{\partial x_j} - \frac{\partial\tilde{u}_j}{\partial x_i}\right), \quad C_{\omega 3} = 20,$$

where $S_{\bar{\rho}\omega}$ are the standard diffusion and source terms in ω equation.

I_ω is a Galilean invariant, local source term which takes different signs at the opposite edges of spatial mixing layers. In temporal mixing layer, it is zero. Recommended "free stream" $C_{\omega 3}$ value is 20 instead of previously reported value 22 to better capture the initial region of round supersonic jets. Near walls, I_ω is not used ("near-wall" value of $C_{\omega 3}$ is 0). The effect of I_ω inclusion in ω equation on single stream spatial mixing layer and on far field of an incompressible free plane jet is shown in Figure 6. Optimized coefficient set (1) was used in all these computations. It is seen that with I_ω term, the two flows considered fit the experimental data.

4 TEST CASES

The modified SSG/LRR-ω model as well as its original version and eddy viscosity SST model have been used to compute four jet test cases described below.

4.1 Free subsonic plane jet

The first test is a free subsonic plane jet described in Section 2. In addition to the results obtained above, modified SSG/LRR-ω model has been used. The same boundary conditions have been specified on the medium mesh. Resulting centerline velocity profiles are shown in Figure 7.

With the modified model, L_{ini} is predicted to be $6h$ which lies at the boundary of experimental range $5h - 6h$ and is approximately two times lower than with two other models considered. Mixing layer width growth rate $dD_{0.1}/dx$ is 0.17 which is within experimental range.

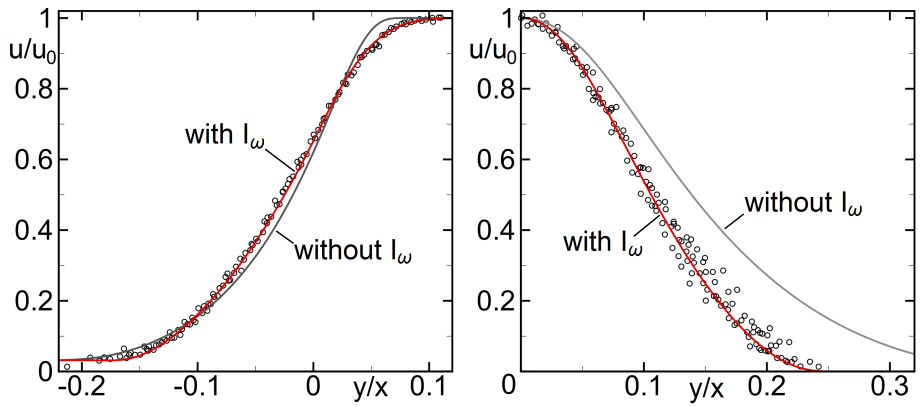


Figure 6: Velocity profiles in single stream spatial mixing layer (left) and plane jet (right) after optimization of the SSG/LRR-ω model coefficients with and without inclusion of I_ω term.

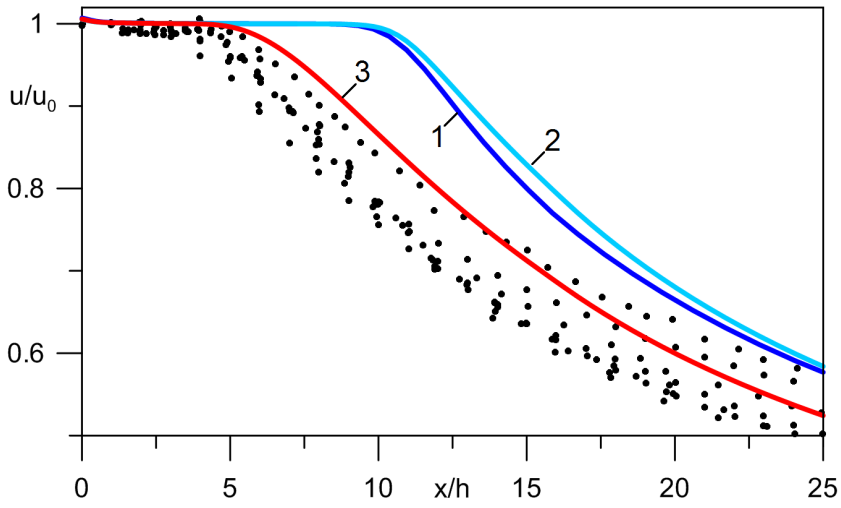


Figure 7: Centerplane velocity distributions obtained in plane jet test with: 1 — SST model, 2 — original SSG/LRR-ω model, 3 — its modified version.

4.2 Free subsonic round jet

This test case corresponds to the free subsonic plane jet described above except that computational mesh is a 5° sector in azimuthal direction. In (x, r) plane, the mesh coincides with the medium mesh used in the plane jet test. The resulting axial velocity distributions obtained with the three turbulence models are presented in Figure 8. Experimental values are taken from [16, 17, 18]. In round jets, experimental L_{ini} values fall in the range $5d - 6d$, where d is nozzle diameter.

Again, the modified model gives $L_{\text{ini}} = 5d$ which lies in the experimental range. The original SSG/LRR-ω model gives $L_{\text{ini}} = 8.5d$ and SST model gives $L_{\text{ini}} = 8d$. As in plane jet, mixing layer width growth rate fits the experimental data for all three turbulence models. Downstream the initial region, all three models predict too fast axial velocity decay. This is due to a well-known "round jet/plane jet anomaly" which can be eliminated by using either axial Pope correction or nonlinear pressure-strain models [19].

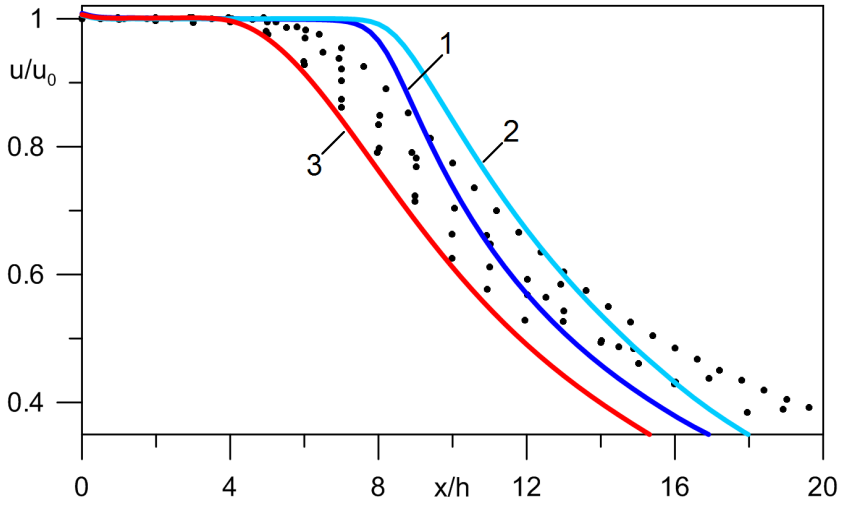


Figure 8: Axial velocity distributions obtained in round jet test with: 1 — SST model, 2 — original SSG/LRR-ω model, 3 — its modified version.

4.3 Underexpanded free round jet

In this test, a cold free supersonic round air jet has been modeled. It was studied experimentally at ITAM SB RAS [20]. The jet issues from a converging nozzle with Mach number at the exit equal to 1. Nozzle pressure ratio is 2.8. Nozzle diameter d based Reynolds number number $\mathrm{Re}_d = u_e d/\nu_e$ is 8.5×10^5, where parameters with subscript "e" are taken at the nozzle exit. Turbulence level at the nozzle exit Tu_e is taken to be 0.9% to reproduce the experimental conditions.

An overview of the computational domain, mesh, and boundary conditions is shown in Figure 9. When generating the mesh, its density has been taken to be approximately the same as in medium mesh in free subsonic plane jet, which had been shown to produce mesh converged solutions.

In Figure 10, axial distributions of relative Pitot pressure p_t/p_0, where p_0 is the total pressure in the core of the nozzle, are shown. SST and original SSG/LRR-ω models capture the first 4 cells of the jet, after which they do not predict experimental p_t fall. Moreover, SST model underpredicts the amplitude of p_t oscillations at $x/d > 5$. The modified SSG/LRR-ω model captures 5 jet cells and p_t fall at $x/d > 5$, which fits the experimental data better.

In Figure 11, three cross-sections of p_t/p_0 obtained in computations are compared with the experimental data. SST and original SSG/LRR-ω models give similar results predicting too sharp high velocity boundary of the mixing layer. The modified SSG/LRR-ω model make mixing layer smoother in this region and captures the experimental behavior of Pitot pressure.

4.4 Coaxial jet

This flow was also studied experimentally at ITAM SB RAS. Now it is one of the tests in First TILDA Workshop on Industrial LES & DNS (2016). A cold free air jet issues from a dual-stream round nozzle. The inner contour is subsonic with nozzle pressure ratio $\mathrm{NPR}_1 = 1.72$ and Mach number at the exit $M_1 = 0.8$. The outer contour is supersonic with $\mathrm{NPR}_2 = 2.25$ (underexpanded flow) and $M_2 = 1.0$. Outer contour diameter D based Reynolds number $\mathrm{Re}_D = u_2 D/\nu_2$ is equal to 2.9×10^6.

Computational mesh near the nozzle is shown in Figure 12. As in previous test, mesh density

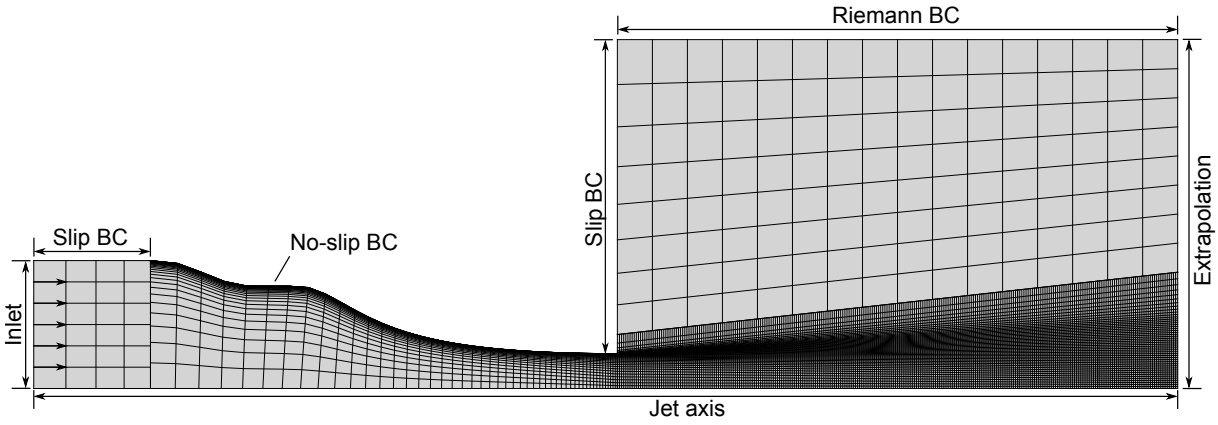


Figure 9: Computational domain, mesh, and boundary conditions in unedrexpanded free round jet test.

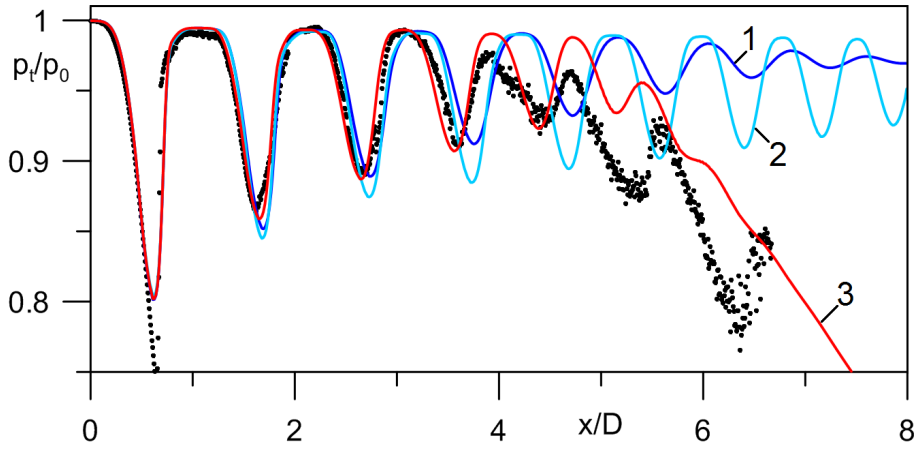


Figure 10: Axial distributions of p_t/p_0 in unedrexpanded free round jet test. 1 — SST model, 2 — original SSG/LRR-ω, 3 — its modified version.

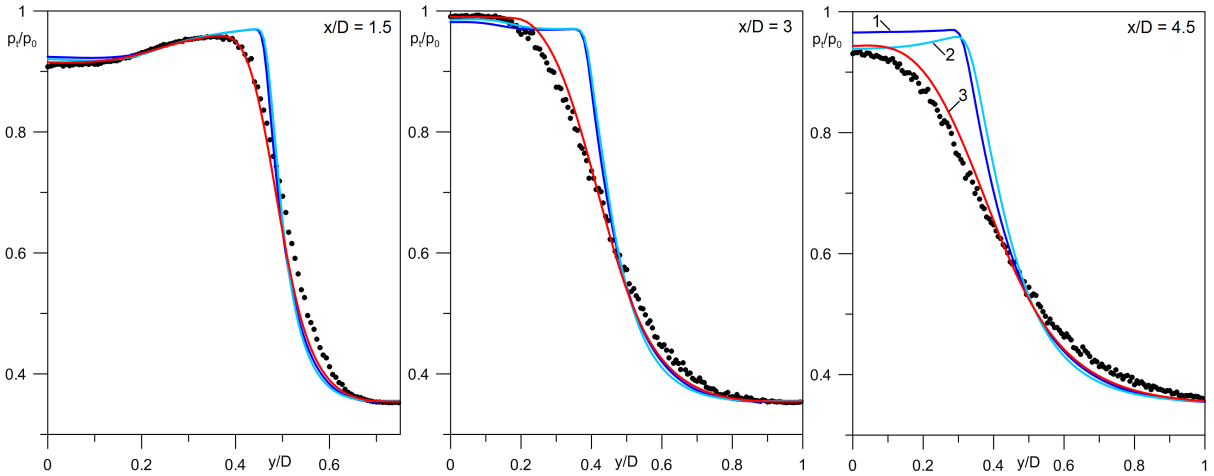


Figure 11: Cross-sections of p_t/p_0 in unedrexpanded free round jet test. 1 — SST model, 2 — original SSG/LRR-ω, 3 — its modified version.

has been taken to be approximately the same as in medium mesh in free subsonic plane jet to get mesh converged solutions. In this test, it contains approximately 150000 cells.

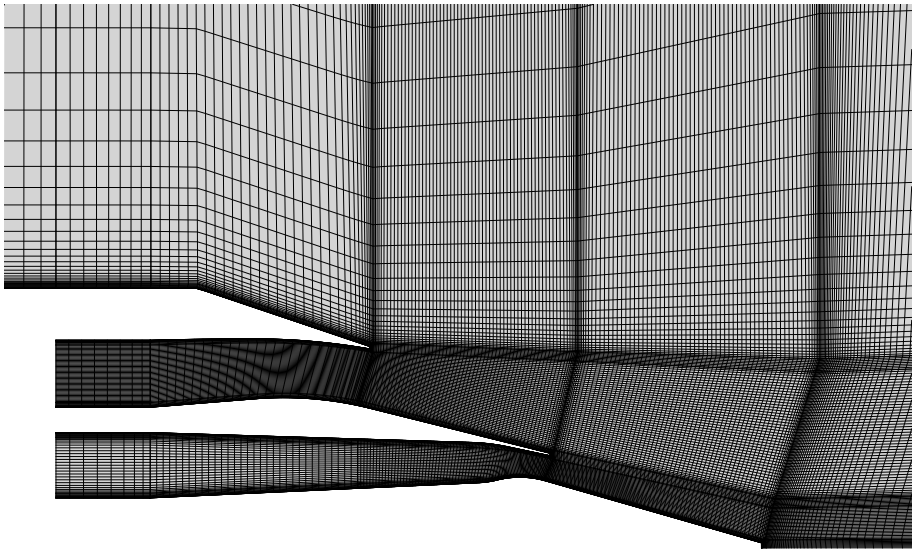


Figure 12: Mesh in coaxial jet test. Every second mesh line is shown.

In Figure 13, three computed cross-sections of p_t/p_0, where p_0 is the total pressure in the air supply of the outer contour, are compared with the experimental data. Longitudinal coordinate axis x origin is at the tip of the central body. As clearly seen from the experimental data, boundary layers at the walls of the outer contour are much thicker than in computations. In fact, their traces collapse already at $x/D = 0.08$. This leads to lower peak p_t values at $y/D \approx 0.3$ than in computations. SST model predicts a small region of p_t decrease at $y/D \approx 0.2$ downstream the position $x/D = 0.80$ which is absent both in experiments and in computations using DRSMs. On the other hand, wake intensity behind the central body is overpredicted by all the models considered. This is an interesting issue to investigate in the future. Apart from this discrepancies, the overall agreement between the computations and the experiment is satisfactory. The modified SSG/LRR-ω model predicts a different velocity profile in the outer mixing layer which seems to be realistic since similar behavior allowed to better fit the experimental data in the previous tests.

5 FURTHER STEPS: LES BASED VALIDATION

As has been shown in paper, the modified model improves modeling accuracy of the jet flows considered in the paper. However, several questions are still open, some of them are:

- Are there other, maybe simpler, forms of additional source term in ω equations which play the same role as I_ω used in the paper? If there are, which one to prefer?
- How to further modify turbulence model equations to better capture the decay of turbulent wakes, e.g. central body wake in the coaxial jet test?
- How do mixing layers develop when they originate from thick boundary layers? Are there any specific features? Does I_ω term needs a correction in this case? This can be studied in a series of coaxial jet test computations with correct boundary layers in the outer contour.

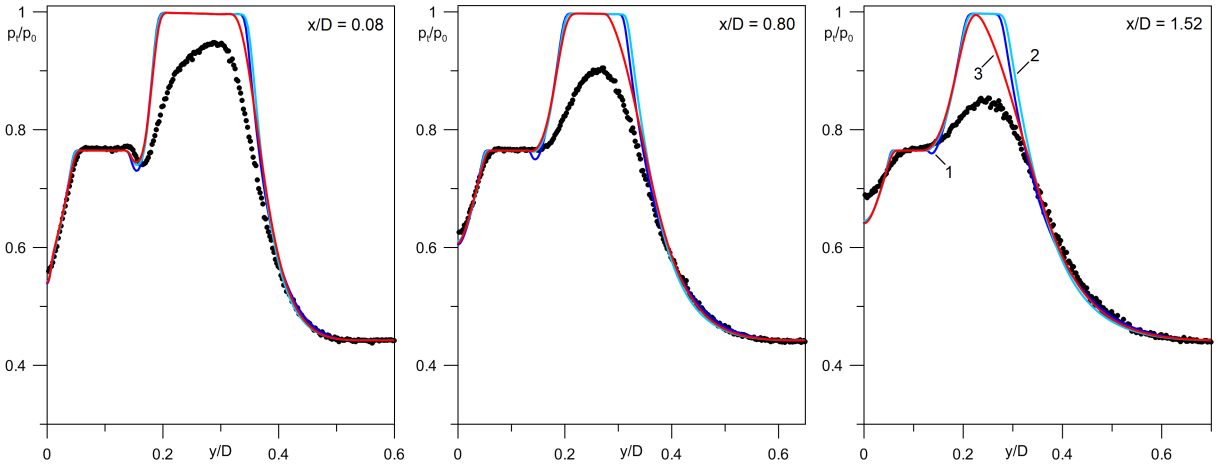


Figure 13: Cross-sections of p_t/p_0 in coaxial jet test. 1 — SST model, 2 — original SSG/LRR-ω, 3 — its modified version.

Nowadays, there is strong progress in developing and validating the high order LES techniques, among which is Discontinuous Galerkin based LES [21]. These techniques are capable of accurately capturing turbulence phenomena in the range of scales, and it looks natural to use these capabilities for term-by-term validation and tuning of RANS turbulence models.

As the next step for the research presented in the paper, we plan to conduct LES based computations of the coaxial jet to get answers to at least some of the questions listed above. We have developed a new in-house code with the following features:

- High spatial order Discontinuous Galerkin method with up to piecewice cubic polynomials.
- Explicit time discretization (multistage Runge-Kutta schemes of up to 4^{th} order) with fractional time stepping technique.
- Complete compressible LES equation system with Smagorinsky subgrid scale model.
- MPI parallelism with efficient scalability on up to $\sim 10^4$ cores.

The data obtained in LES computations will allow to extract distributions of individual terms in Reynolds stress and ω equations. They will be compared to the data in RANS computations to improve the engineering models for different terms in RANS turbulence models.

6 CONCLUSIONS

It is shown that different standard turbulence models, from linear eddy viscosity models to DRSMs, incorrectly predict jet potential core length. SSG/LRR-ω DRSM is recalibrated and supplemented by an additional source term in the turbulence characteristic frequency ω equation which accounts the longitudinal flow inhomogeneity and entrainment. These effects influence turbulence in jet mixing layers and change its velocity profile. Four jet test cases are considered, in all of which the modified model improves the accuracy of flow field predicion over the standard models examined in the paper.

Theoretical part of the study was supported by RFBR, research project No. 16-38-00760.

Computations presented in the paper were supported by the Ministry of Education and Science, project No. 14.628.21.0005.

REFERENCES

[1] W.A. Engblom, N.J. Georgiadis, A. Khavaran, Investigation of variable-diffusion turbulence model correction for round jets. *11th AIAA/CEAS Aeroacoustics Conference*, Monterey, California, May 23-25, 2005.

[2] N.J. Georgiadis, D. Papamoschou, Computational investigations of high-speed dual-stream jets. *9th AIAA/CEAS Aeroacoustics Conference and Exhibit*, Hilton Head, South Carolina, May 12-14, 2003.

[3] A.I. Troshin, A turbulence model taking into account the longitudinal flow inhomogeneity in mixing layers and jets. *6th European Conference for Aerospace Sciences*, Krakow, Poland , June 29 - July 3, 2015.

[4] A.I. Troshin, A turbulence model with variable coefficients for calculating mixing layers and jets. *Fluid Dynamics*, **47**, No. 3, 320–328, 2012.

[5] V. Neyland, S. Bosniakov, S. Glazkov, A. Ivanov, S. Matyash, S. Mikhailov et al., Conception of electronic wind tunnel and first results of its implementation. *Progress in Aerospace Sciences*, **37**, No. 2, 121–145, 2001.

[6] F.R. Menter, M. Kuntz, R. Langtry, Ten years of industrial experience with the SST turbulence model. *Turbulence, Heat and Mass Transfer*, **4**, 625–632, 2003.

[7] T.J. Coakley, Turbulence modeling methods for the compressible Navier-Stokes equations. *AlAA 16th Fluid and Plasma Dynamics Conference*, Danvers, Massachusetts, July 12-14, 1983.

[8] P.R. Spalart, S.R. Allmaras, A one-equation turbulence model for aerodynamic flows. *30th Aerospace Sciences Meeting & Exhibit*, Reno, Nevada, January 6-9, 1992.

[9] D.C. Wilcox, *Turbulence modeling for CFD. 3rd Edition*. DCW Industries, 2006.

[10] R.-D. Cécora, R. Radespiel, B. Eisfeld, A. Probst, Differential Reynolds-stress modeling for aeronautics. *AIAA Journal*, **53**, 739–755, 2015.

[11] E. Gutmark, I.J. Wygnanski, The planar turbulent jet. *Journal of Fluid Mechanics*, **73**, 465–495, 1976.

[12] B.R. Ramaprian, M.S. Chandrasekhara, LDA measurements in plane turbulent jets. *Journal of Fluids Engineering*, **107**, 264–271, 1985.

[13] M. Alnahhal, Th. Panidis, The effect of sidewalls on rectangular jets. *Experimental Thermal and Fluid Science*, **33**, 838–851, 2009.

[14] R.C. Deo, J. Mi, G.J. Nathan, The influence of Reynolds number on a plane jet. *Physics of Fluids*, **20**, 075108, 2008.

[15] J.H. Bell, R.D. Mehta, Development of a two-stream mixing layer from tripped and untripped boundary layers. *AIAA Journal*, **28**, 2034–2042, 1990.

[16] H.J. Hussein, S.P. Capp, W.K. George, Velocity measurements in a high-Reynolds-number, momentum-conserving, axisymmetric, turbulent jet. *Journal of Fluid Mechanics*, **258**, 31–75, 1994.

[17] J.C. Lau, P.J. Morris, M.J. Fischer, Measurements in subsonic and supersonic free jets using a laser velocimeter. *Journal of Fluid Mechanics*, **93**, 1–27, 1979.

[18] J. Bridges, M. Wernet, Establishing consensus turbulence statistics for hot subsonic jets. *16th AIAA/CEAS Aeroacoustics Conference*, Stockholm, Sweden, June 7-9, 2010.

[19] K. Hanjalić, B. Launder, *Modelling turbulence in engineering and the environment: second-moment routes to closure*. Cambridge University Press, 2011.

[20] V.I. Zapryagaev, N.P. Kiselev, A.A. Pivovarov, Gasdynamic structure of an axisymmetric supersonic underexpanded jet. *Fluid Dynamics*, **50**, No. 1, 87–97, 2015.

[21] F. Bassi, L. Botti, A. Colombo, A. Crivellini, A. Ghidoni, A. Nigro, S. Rebay, Time integration in the Discontinuous Galerkin code MIGALE - unsteady problems. *Numerical Fluid Mechanics and Multidisciplinary Design*, **128**, 205–230, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MODIFIED SSG/LRR-ω REYNOLDS STRESS MODEL FOR PREDICTING BLUFF BODY AERODYNAMICS

Csaba Klajbár[1], László Könözsy[2], and Karl W. Jenkins[2]

[1]Centre for Fluid Mechanics and Computational Science
Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom
e-mail: c.klajbar@cranfield.ac.uk

[2] Centre for Fluid Mechanics and Computational Science
Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom
e-mail: laszlo.konozsy@cranfield.ac.uk, k.w.jenkins@cranfield.ac.uk

Keywords: SSG/LRR-ω Reynolds Stress Models, turbulent flow separation, complex vortical structures, computational fluid dynamics (CFD).

Abstract. *This work focuses on a complex turbulent flow around a blunt body by predicting vortical structures, separation/re-attachment locations and velocity/turbulent kinetic energy profiles relying on a modified SSG/LRR-ω Reynolds Stress Model (RSM). The investigated physical problem is in the centre of research interest, because the Reynolds-Averaged Navier-Stokes (RANS) turbulence models and Large Eddy Simulation (LES) approaches usually fail to reproduce the physically correct flow field. Due to the fact that there is lack of knowledge on the SSG/LRR-ω hybrid RSM closure model proposed by Cecora et al. [1], therefore this work has been devoted to further investigate the overall numerical behaviour of a modified hybrid RSM approach. The advantage of these RSM closure models is to take into account the anisotropy of Reynolds stresses caused by the streamline curvature, which has importance in maintaining the quasi non-diffusive nature of turbulent vortices. The numerical model implementation has been verified through the classical test case of a turbulent flow over a flat plate at zero pressure-gradient on a sequence of nested grids. A preliminary analysis performed on this benchmark problem showed that the hybrid model is capable of predicting the near-wall turbulence in the fully-developed boundary layer with a reasonable accuracy. It has also been observed that the SSG/LRR-ω adopts a free-stream independence feature of the specific dissipation rate from the Menters [2] k-ω SST model, thus the turbulent quantities in the near-wall exhibit almost negligible sensitivity respect to the specific dissipation rate outside of the shear layer. In this work, a modified hybrid SSG/LRR-ω RSM closure in conjunction with a simplified diffusion model has been proposed, which has been investigated through the classical cubic obstacle problem in a three-dimensional channel flow of Martinuzzi and Tropea [3]. An increased accuracy and demanding computational time have been observed by employing the modified SSG/LRR-ω formulation, which could still be advantageous, because it unifies the favourable features of two distinct differential Reynolds stress models producing accurate results in the near-wall and the shear layer regions.*

1 INTRODUCTION

Most engineering flows used in aerodynamic design and optimisations for a wide range of applications are almost exclusively turbulent [1, 2, 3]. Therefore several advanced turbulence modelling approaches are proposed over the last twenty years, including advanced scale-resolving techniques as Direct Numerical Simulation (DNS) or Large-Eddy Simulation (LES). The numerical treatment of turbulent flows in general is complicated task and the reliability of numerical tools is always in the centre of interest. The practical applicability of these methods is computationally very demanding, because when the Reynolds number increases, the size of the smallest turbulent structures decreases, thus the computational cost drastically increases. In contrast with that, Reynolds-Averaged Navier-Stokes (RANS) approaches treat turbulent flows by considering the average influence of turbulence on the mean flow by employing classical turbulence models. These engineering turbulence modelling tools are widely used for industrial applications due to their efficiency, robustness and acceptable accuracy, however there are many limitations related to these models caused by the modelling assumptions.

The differential Reynolds stress models (RSMs) are considered as the most advanced RANS tools. This concept originates from the work of Launder et al. [4] and these models completely abandon the Boussinesq-hypothesis, which usually causes the poor performance of the popular eddy-viscosity models. Six transport equations are solved for the Reynolds stresses tensor components and one equation for the turbulent length scale. In these models, the contribution of Reynolds stresses are added directly to the momentum equation instead of employing eddy-viscosity models. The SSG/LRR-ω Reynolds stress model was first proposed by Eisfeild and Brodersen [5]. In addition to the original version of the aforementioned Reynolds stress model, a hybrid SSG/LRR-ω model was developed by Cecora et al. [1] in 2015.

In the present work, the recently developed hybrid SSG/LRR-ω model [1] has been further improved, thus a modified hybrid SSG/LRR-ω approach with a simplified diffusion model has been proposed. The hybrid RSM of Cecora et al. [1] was developed by adopting Menter's blending concept [2], the SSG model of Speziale et al [6] is employed in the near-wall region, and the LRR-ω model of Launder et al. [4] is applied on the rest of the domain. The hybrid RSM closure [1] was primarily developed for aerospace applications, thus its performance for bluff body flows is not yet well-known and it is an interest of the present paper. Therefore the physical problem of an incompressible flow behind a cubic obstacle placed in a three-dimensional channel, which is a classical bluff body test case investigated by Martinuzzi and Tropea [3], has been used to assess the proposed modified hybrid SSG/LRR-ω model.

2 GOVERNING EQUATIONS AND NUMERICAL IMPLEMENTATION

2.1 Reynolds-Averaged Navier-Stokes (RANS) equations

For modelling unsteady, incompressible turbulent flows with Reynolds-Averaged Navier-Stokes (RANS) approach, the governing equations consist of the continuity and momentum conservation laws which can be written [9] with index notation as

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0, \tag{1}$$

$$\frac{\partial}{\partial t}(\rho \bar{u}_i) + \bar{u}_j \frac{\partial}{\partial x_j}(\rho \bar{u}_i) = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}\left(2\mu S_{ij} - \rho R_{ij}\right), \tag{2}$$

where $\bar{u}_i$ is the Reynolds-averaged (time-averaged) velocity vector, $\bar{p}$ is the hydrodynamic pressure, ρ is the fluid density, μ is the dynamic viscosity which are considered to be constants. The mean rate of strain tensor S_{ij} represents the averaged viscous stresses of a Newtonian fluid as

$$S_{ij} = \frac{1}{2}\left(\frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \bar{u}_j}{\partial x_i}\right). \tag{3}$$

The term ρR_{ij} is the Reynolds stress tensor on the right-hand side of the momentum equation 2, which represents the averaged contribution of turbulence to the mean flow. The elements of this tensor are unknown, thus six additional terms need to be modelled by a suitable closure.

2.2 Differential Reynolds Stress Models

The highest level approximation of the Reynolds stress tensor can be achieved by employing differential Reynolds stress models. In these cases, the eddy-viscosity hypothesis is completely removed from the six transport equations for the components of the Reynolds stress tensor and an additional equation has to be solved for the turbulent length scale which is required for the closure. These RSM models are computationally more expensive, because the solution of the coupled system of partial differential transport equations could lead to numerical instabilities. In addition to this, the unknown terms require modelling assumptions similar to the RANS modelling approaches, which are inevitable sources of uncertainties. Due to these reasons, Reynolds stress models are rarely used for industrial applications, because the general performance of a well-calibrated RANS eddy-viscosity model could also be better. The Reynolds stress transport equations by neglecting the effects of buoyancy and system rotation can be written as

$$\frac{\partial}{\partial t}\left(\rho R_{ij}\right) + \frac{\partial}{\partial x_k}\left(\rho R_{ij}\bar{u}_k\right) = \rho P_{ij} + \rho \Pi_{ij} - \rho \epsilon_{ij} + \rho D_{ij} + \rho M_{ij}, \tag{4}$$

where ρP_{ij} represents the Reynolds stress production, $\rho \Pi_{ij}$ is the pressure-strain correlation tensor, $\rho \epsilon_{ij}$ is the kinetic energy dissipation tensor, ρD_{ij} is the diffusion tensor. It is important to note that the last term is the fluctuating mass flux tensor ρM_{ij} which is neglected for modelling incompressible flows. The elements of the Reynolds stress production tensor are functions of the Reynolds stresses and the derivatives of the mean velocity components as

$$\rho P_{ij} = -\rho R_{ik}\frac{\partial \bar{u}_j}{\partial x_k} - \rho R_{jk}\frac{\partial \bar{u}_i}{\partial x_k}. \tag{5}$$

The reason for neglecting the fluctuating mass flux is that it would represent the contributions of density fluctuations to the Reynolds stress transport equation, which are not present for incompressble flows, thus this term is equal to zero ($\rho M_{ij} = 0$).

2.3 Transport equations of the hybrid SSG/LRR-ω Reynolds Stress model

The SSG/LRR-ω full Reynolds stress model proposed by Cecora et al. [1] is a hybrid combination of Speziale-Sarkar-Gatski (SSG) [6] and Launder-Reece-Rodi (LRR) [4] models by employing Menter's [2] blending concept. The primary motivation for constructing this model was to unify the advantageous features of the aforementioned RSMs. The SSG model formulation is a widely used pressure-strain correlation model, however it employs the turbulent dissipation rate ϵ as a length scale variable. For external aerodynamics, the specific dissipation rate ω was proven to be more advantageous than the application of the dissipation rate ϵ,

therefore the SSG model is used in conjunction with a modified form of Menter's equation for the specific dissipation rate ω in the farfield [2]. The LRR-ω re-distribution model is employed without wall-reflection terms near to the viscous wall. For the hybrid SSG/LRR-ω model proposed by Cecora et al. [1], the pressure-strain correlation tensor is modelled by neglecting the effects of pressure dilation as

$$\begin{aligned}\rho\Pi_{ij} = &-\left(C_1\rho\epsilon + \frac{1}{2}C_1^*\rho P_{kk}\right)a_{ij} + C_2\rho\epsilon\left(a_{ik}a_{kj} - \frac{\delta_{ij}}{3}a_{kl}a_{kl}\right) + \\ &+\left(C_3 - C_3^*\sqrt{a_{kl}a_{kl}}\right)\rho k S_{ij}^* + C_4\rho k\left(a_{ik}S_{jk} + a_{jk}S_{ik} - \frac{2}{3}a_{kl}S_{kl}\delta_{ij}\right) + \\ &+ C_5\rho k\left(a_{ik}W_{jk} + a_{jk}W_{ik}\right),\end{aligned} \tag{6}$$

and the Reynolds stress anisotropy tensor can be computed with the turbulent kinetic energy by

$$a_{ij} = \frac{R_{ij}}{k} - \frac{2}{3}\delta_{ij}, \tag{7}$$

and the mean rate of rotation tensor can be defined as

$$W_{ij} = \frac{1}{2}\left(\frac{\partial\bar{u}_i}{\partial x_j} - \frac{\partial\bar{u}_j}{\partial x_i}\right). \tag{8}$$

The turbulent dissipation rate in the pressure-strain correlation tensor (6) can be expressed by using the specific turbulent dissipation rate ω with a closure constant $C_\mu = 0.09$ as

$$\epsilon = C_\mu k\omega, \tag{9}$$

and the model coefficients in Eq. (6) can be obtained with a blended manner as

$$C_i = F_1 C_i^{(\omega)} + (1 - F_1)C_i^{(\epsilon)}, \quad i = 1, ..., 5; \tag{10}$$

$$C_i^* = (1 - F_1)C_i^{(\epsilon)}, \quad i = 1, 3, \tag{11}$$

where F_1 denotes the blending function, which is similar to that one used in the k-ω SST model proposed by Menter [2]. The model constants used in the present work have been shown in Table 1, which are divided into inner ω and outer ϵ coefficients. The closure parameters were derived from the $C_i^{(\omega)}$ and $C_i^{(\epsilon)}$ constants of the LRR-ω and SSG models, respectively.

The blending function F_1 in Eqs. (10) and (11) ensures that the LRR-ω model acts close to the viscous wall and the SSG formulation is employed on the rest of the computational domain. The function F_1 was developed in a way to be equal to 1 starting from the wall up to approximately 50% of the boundary layer and after that it tends gradually to 0. The blending function depends on global fluid properties, such as dynamic viscosity and fluid density, and on turbulence-related quantities, such as turbulent kinetic energy and specific dissipation rate, and it also takes into account the distance d of the cell to the nearest viscous wall as

$$\begin{aligned}&F_1 = tanh(\zeta^4),\\ &\zeta = min\left[max\left(\frac{\sqrt{k}}{C_\mu\omega d}, \frac{500\mu}{\rho\omega d^2}\right), \frac{4\sigma_\omega^{(\epsilon)}\rho k}{CD_\omega d^2}\right],\\ &CD_\omega = \frac{\sigma_d^\epsilon\rho}{\omega}max\left(\frac{\partial k}{\partial x_k}\frac{\partial\omega}{\partial x_k}, 0\right).\end{aligned} \tag{12}$$

Inner constants: LRR (ω)		Outer constants: SSG (ϵ)	
$\alpha_\omega^{(\omega)} = 0.5556$	$\beta_\omega^{(\omega)} = 0.075$	$\alpha_\omega^{(\epsilon)} = 0.44$	$\omega^{(\epsilon)} = 0.08282$
$\sigma_\omega^{(\omega)} = 0.5$	$\sigma_d^{(\omega)} = 0$	$\sigma_\omega^{(\epsilon)} = 0.856$	$\sigma_d^{(\epsilon)} = 1.712$
$C_1^{(\omega)} = 1.8$	$C_1^{*(\omega)} = 0$	$C_1^{(\epsilon)} = 1.7$	$C_1^{*(\epsilon)} = 0.9$
$C_2^{(\omega)} = 0$	$C_3^{(\omega)} = 0.8$	$C_2^{(\epsilon)} = 1.05$	$C_3^{(\epsilon)} = 0.8$
$C_3^{*(\omega)} = 0$	$C_4^{(\omega)} = \big(9C_2^{(LRR)} + 6\big)/11$	$C_3^{*(\epsilon)} = 0.65$	$C_4^{*(\epsilon)} = 0.625$
$D^{(\omega)} = 0.75C_\mu$	$C_5^{(\omega)} = \big(-7C_2^{(LRR)} + 10\big)/11$	$D^{(\epsilon)} = 0.22$	$C_5^{(\epsilon)} = 0.2$
$C_2^{(LRR)} = 0.52$			

Table 1: Closure coefficients of the SSG/LRR-ω full Reynolds stress model proposed by Cecora et al. [1].

In order to ensure computational robustness and efficiency, the diffusion term is modelled by employing a so-called *simplified diffusion model*, which takes into account the effects of molecular and turbulent diffusion in a simplified way as

$$\rho D_{ij} = \frac{\partial}{\partial x_k}\left[\left(\mu + D_S \frac{\rho k}{C_\mu \omega}\right)\frac{\partial R_{ij}}{\partial x_k}\right], \quad D_S = 0.5C_\mu F_1 + \frac{2}{3}0.22(1 - F_1), \tag{13}$$

and the dissipation term is modelled through an isotropic tensor as

$$\rho\epsilon_{ij} = \frac{2}{3}\rho\epsilon\delta_{ij}, \tag{14}$$

where the turbulent dissipation rate ϵ is predicted by using the specific dissipation rate ω based on Eq. (9). An altered form of Menter's ω-equation is used to supply the length scale where only minor differences were made in the production (ρG_ω), destruction (ρY_ω) and cross-diffusion terms (ρCD_ω) compared to the original ω-equation [2] to adapt to the hybrid model as

$$\begin{aligned}
&\frac{\partial}{\partial t}(\rho\omega) + \frac{\partial}{\partial x_i}(\rho U_i \omega) = \rho G_\omega - \rho Y_\omega + \rho D_\omega + \rho CD_\omega, \\
&\rho G_\omega = \alpha_\omega \frac{\omega}{k}\frac{\rho P_{kk}}{2}, \quad \rho Y_\omega = \beta_\omega \rho \omega^2, \quad \rho D_\omega = \frac{\partial}{\partial x_k}\left[\left(\mu + \sigma_\omega \frac{\rho k}{\omega}\right)\frac{\partial \omega}{\partial x_k}\right], \\
&\rho CD_\omega = \frac{\sigma_d^\omega \rho}{\omega} max\left(\frac{\partial k}{\partial x_k}\frac{\partial \omega}{\partial x_k}, 0\right)
\end{aligned} \tag{15}$$

It is important to note that the cross-diffusion term ρCD_ω makes the link between the k-ϵ and k-ω model behaviours, and Reynolds stress models inherits this term. The closure coefficients are obtained in a blended manner in Eq. (10) with $C_i = \alpha_\omega,\ \beta_\omega,\ \sigma_\omega$ and σ_d, and the model constants of the ω-equation can be found in Table 1.

2.4 Implementation of the modified SSG/LRR-ω model

The modified hybrid SSG/LRR-ω approach with a simplified diffusion model is not available by default in the ANSYS-FLUENT commercial software package [10], therefore according to the best of our knowledge, the present work is one of the first attemps to implement the modified RSM in the ANSYS-FLUENT environment. The modified hybrid SSG/LRR-ω model has been implemented by using User-Defined Functions (UDFs) in C programming language. For the system of governing equations, 6+1 User-Defined Scalar (UDS) transport equations have

been solved by adding the corresponding turbulent source terms to the scalar momentum equations. The incompressible continuity and Reynolds momentum equations were solved by using the pressure-based ANSYS-FLUENT solver [10] through the implemented modified hybrid SSG/LRR-ω model by solving the additional RSM scalar transport equations. The convective flux terms of the momentum equations and the turbulent RSM-UDS transport equations were discretised by using a 2^{nd}-order upwind scheme. In order to ensure convergence to the numerical solutions, all residuals were required to be dropped below a threshold value of 10^{-6}. It was shown by monitoring all shear-stress related quantities that the prescribed criterion was a sufficient condition for ensuring numerical convergence of the solution.

3 RESULTS AND DISCUSSION

The implementation of the modified hybrid SSG/LRR-ω approach with a simplified diffusion model has been compared to the SSG and LRR-ω models on two fundamentally different test cases. The first benchmark problem is a turbulent flow over a flat plate at zero pressure gradient which is an extensively used validation test case for investigating and verifying RANS turbulence model code implementations. The second one is a very challenging flow problem as a turbulent flow past a cube placed in a three-dimensional channel flow which was investigated by Martinuzzi and Tropea [3]. For modelling this physical problem, RANS approaches may fundamentally fail and even LES techniques may also not able to capture accurately the vortical structures of an aerodynamic turbulent flow behind a bluff body at high Reynolds numbers.

3.1 Turbulent flow over a flat plate at zero pressure gradient

The sketch of a fairly simple two-dimensional turbulent flow problem over a flat plate at zero pressure gradient has been shown in Figure 1. The lower boundary starts with an *inviscid* wall section which is followed by a *viscous* wall extending from $x/L = 0.0$ to $x/L = 2.0$, where $L = 1\,m$ is the reference length. The subsonic air enters into the domain at the left velocity inlet boundary at a low Mach number ($Ma = 0.2$). Due to the low magnitude of the farfield Mach number, the fluid flow can be considered as incompressible. The Reynolds number Re_L is defined by the reference length ($Re_L = 5 \cdot 10^6$). By considering these physical circumstances, the maximum boundary layer thickness is estimated as $\delta_{max} \approx 0.03 \cdot L$, thus the domain height of $y/L = 1.0$ can be taken to be high enough to avoid the effect of the upper *inviscid* wall on the near-wall boundary layer flow. The fluid flow leaves the domain at the right *outflow* boundary where the normal derivatives of the velocity components are vanished.

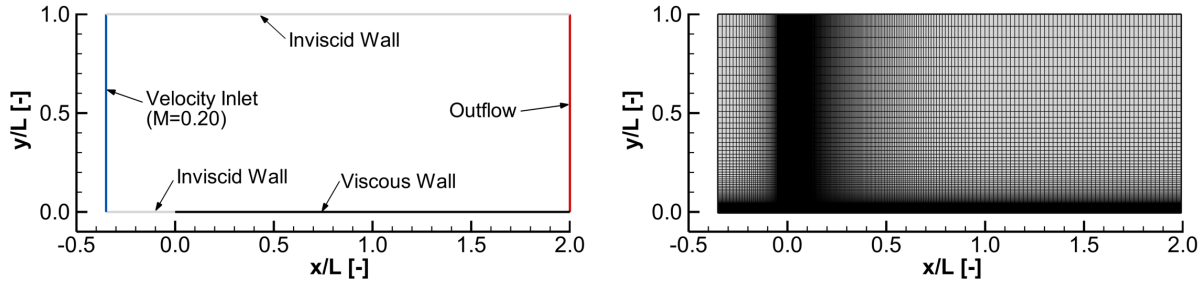


Figure 1: Computational domain with boundary conditions (left) and the computational grid (right) of the flow over a flat plate problem.

A structured computational grid consists of 273×193 node points was used for the com-

putations on the solid plate (see Figure 1). The dimensionless distance y^+_{wall} was kept below 1 in the near wall cells, because the boundary layer was fully-resolved without using any type of wall-function. Furthermore, grid points have been clustered in the direction y and stretched near the leading edge of the plate in order to capture accurately the high gradients of flow variables.

3.1.1 Performance of the modified SSG/LRR-ω model for the flat plate problem

The dimensionless velocity profile u^+ as a function of the dimensionless wall-distance y^+ and the skin friction coefficient c_f distribution along the plate surface have been used to assess the model performance as

$$u^+ = \frac{u}{u_\tau}, \quad y^+ = \frac{u_\tau \cdot y}{\nu}, \quad \text{where} \quad u_\tau = \sqrt{\frac{\tau_w}{\rho}}, \tag{16}$$

and the skin friction coefficient can be written as

$$c_f = \frac{\tau_w}{0.5\rho U^2_\infty A}. \tag{17}$$

The numerical results have been compared to the theoretical curves of Coles [11] and von Kármán & Schroerr [12] (see Figure 2). The worst agreement was produced by the ϵ-based SSG model, because significant discrepancy has been observed in the u^+-y^+ profile outside of the buffer layer while the skin friction c_f distribution was over-predicted. The observed relatively poor performance for predicting near-wall turbulence can be explained with the fact that the SSG model was developed primarily for shear flows. The LRR-ω RSM model performed fairly well for both metrics and an almost perfect match was achieved for the law-of-the-wall and the skin friction coefficient distributions. This property is related to the excellent ability of ω-based models for predicting wall-bounded turbulent flows. The modified SSG/LRR-ω model performed in between the previous two formulations. The tendency of the dimensionless velocity profile was predicted correctly, and above a certain portion of the log-law region, it produced significant discrepancy. The skin friction curve matched well with the theoretical prediction.

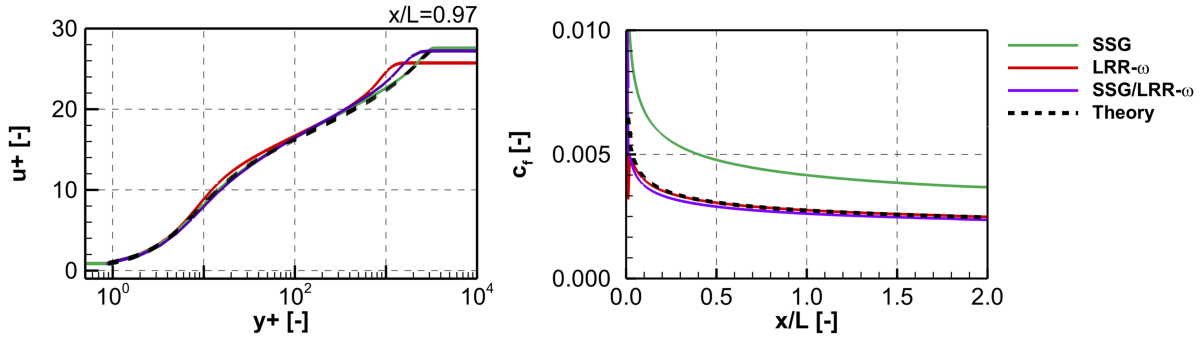


Figure 2: Computed law-of-the-wall compared with the theory of Coles [11] at the $x/L = 0.97$ location (left) and skin-friction coefficient distribution along the wall compared with the von Kármán-Schoenherr [12] theory (right).

3.1.2 Freestream dependency of the specific turbulent dissipation rate ω

There is another important advantage of the SSG/LRR-ω model related to its 7^{th} transport equation. The LRR-ω model employs an altered form of Wilcox's ω-equation [8] for determining the turbulent length scale while the SSG/LRR-ω model was developed with a slightly

modified version of Menter's ω-equation [2]. The k-ω formulation of Wilcox is known to produce freestream dependence of ω which means that the results near the wall strongly depend on the magnitude of the turbulent specific dissipation rate specified outside of the shear layer. This can lead to significant increase in the turbulent viscosity, thus excessive μ_t may prevent flow separation in certain cases. In the development of the k-ω SST model, Menter's [2] one of key motivations was to propose an ω-based turbulence model which has an excellent freestream independence of the k-ϵ formulation. This property is accomplished by adding the so-called cross-diffusion term to the ω-equation, and as a result of that, the predicted near-wall turbulence of the k-ω SST model is far less sensitive to the prescribed ω magnitude at the inlet boundary. This is a very important feature of the this model, because the ω-related boundary conditions are determined relying on differernt assumptions, which are usually sources of uncertainty.

It was observed by relying on numerical experience that although the contribution of ω is obviously different for Reynolds stress models compared to the linear eddy-viscosity models, however the previous properties are inherited in some extent. Numerical investigations have been carried out by using three different ω^{inlet} values of $0.1 \cdot \omega_\infty$, $1 \cdot \omega_\infty$, $10 \cdot \omega_\infty$, where ω_∞ is defined based on the recommendations of Rumsey [13] as

$$\omega_\infty = \frac{3/2(Tu_\infty)^2 \rho U_\infty^2}{\mu_{t,\infty}}, \tag{18}$$

where the farfield turbulence intensity Tu_∞ is taken to be equal to 0.001 and the farfield turbulent viscosity $\mu_{t,\infty}$ is assumed to be equal to 0.1μ. The normalised eddy-viscosity (μ_t/μ) and normalised Reynolds stress (Re^+_{xx}, Re^+_{yy}) distributions can be expressed by

$$\frac{\mu_t}{\mu} = \frac{\rho k}{\omega \mu}, \qquad Re^+_{xx} = \frac{Re_{xx}}{u_\tau} \quad \text{and} \quad Re^+_{yy} = \frac{Re_{yy}}{u_\tau}, \tag{19}$$

which are compared at the $x/L = 0.97$ coordinate with the LRR-ω and the modified SSG/LRR-ω models (see Figure 3). The LRR-ω formulation predicted a clearly higher magniture of μ_t/μ peak over an elongated extent for the lower freestream turbulent dissipation case ($0.1 \cdot \omega_\infty$) compared to the other inlet values. The hybrid RSM exhibited considerably less numerical sensitivity, because very minor deviations can be observed between the corresponding μ_t/μ curves. This is also valid for the normalised Reynolds stress distributions: a) the LRR-ω model predicted approximately 20% increase in the peak mean stress magnitudes for the $\omega^{inlet} = 0.1 \cdot \omega_\infty$, and b) the SSG/LRR-$\omega$ closure produced almost identical Reynolds stress distributions.

3.2 Flow past a cube placed in a three-dimensional channel

The benchmark problem of Martinuzzi and Tropea [3] is a flow past a cubic obstacle placed in a fully-developed turbulent channel flow (see Figure 4). Despite of the fact that the geometry is relatively simple, fairly complex flow structures can be observed around the bluff object. Due to the blockage effect of the body, a strong adverse pressure gradient develops resulting to the formation of the system of horseshoe- and hairpin vortices. Capturing these flow features poses challenges for classical RANS and even advanced scale-resolving turbulence models, thus the investigation of this physical problem can be considered as a long-lasting research topic. In this case, the subsonic inflow is characterized by a bulk Reynolds number ($Re_b = 40\,000$) which is defined by the obstacle height H and bulk average flow velocity U_b. Unsteady simulations have been carried out with a CFL number of 0.95 until the flow time of $50 \cdot H/U_{mean}$ is reached. After that $20 \cdot H/U_{mean}$ time period was used to collect statistical averages of the flow field.

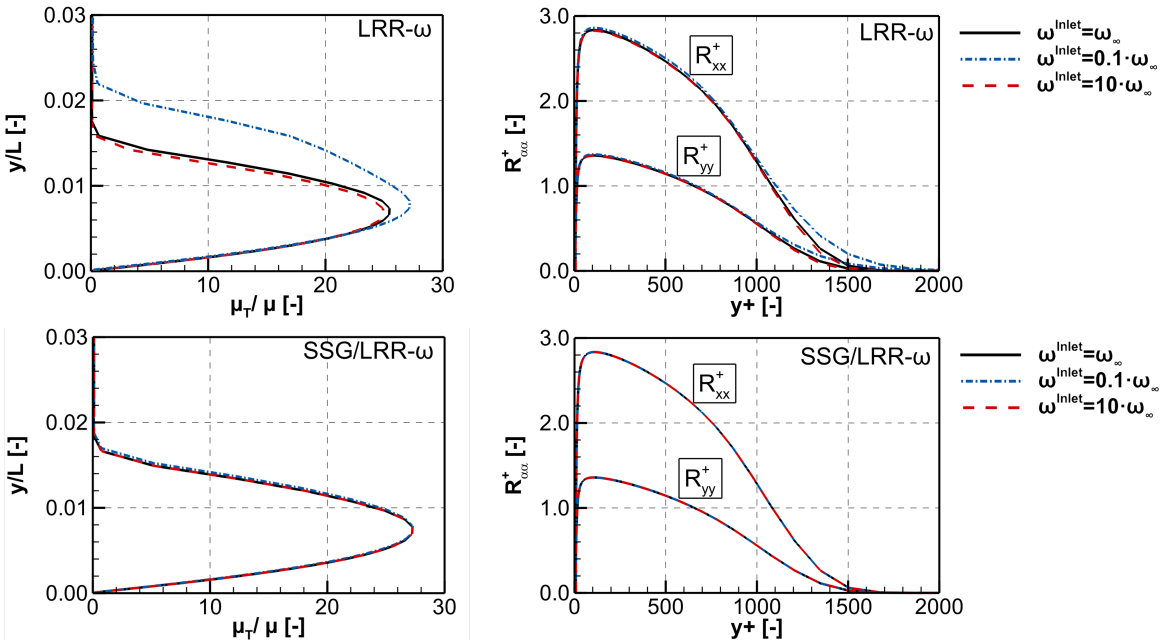


Figure 3: Normalised turbulent viscosity distribution along the vertical direction (left) and normalised Reynolds stress distribution as the function of dimensionless wall distance (right) at the $x/L = 0.97$ location.

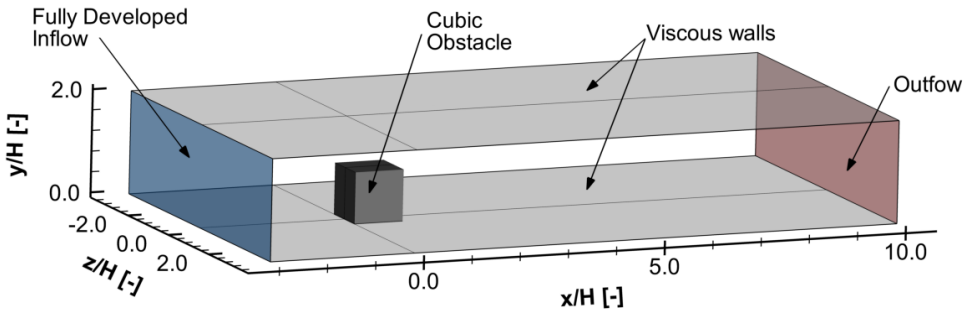


Figure 4: Sketch and boundary conditions of the flow past a cubic obstacle problem.

The boundary conditions of this benchmark problem have been shown in Figure 4. A fully-developed inflow is specified at the inlet plane and the velocity distribution has been obtained through preliminary streamwise periodic channel flow simulations. The cubic obstacle and the upper/lower channel surfaces are treated as viscous walls with no-slip boundary conditions. The two side walls are treated as symmetry planes and the normal derivatives of the velocity compontnes vanish at the outlet section of the channel. A structured computational grid has been employed for the analysis of this benchmark problem (see Figure 5). After performing a grid sensitivity study, the final configuration employs $190 \times 100 \times 120$ cells in directions of x, y and z with the total number of $2\,280\,000$ hexahedral volumes. Near the viscous walls, the grid spacing has been chosen in a way to ensure y^+_{wall} approximately equal to 1.

3.2.1 Performance of the modified SSG/LRR-ω model for a flow past a cube

The numerical results obtained with the aforementioned three turbulence models have been presented in this subsection. The iso-surfaces of Q-criterion indicate that similar upstream

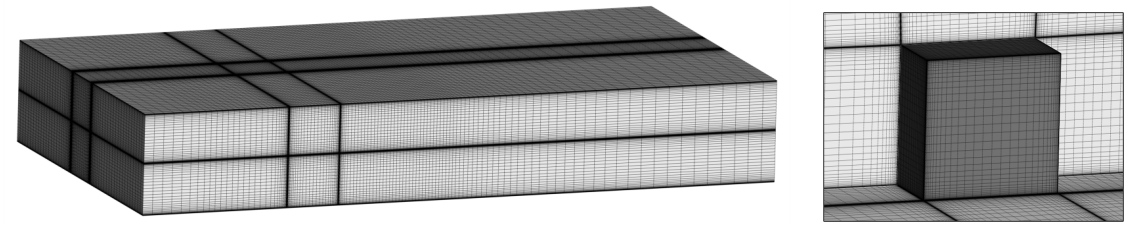


Figure 5: Details of the structured grid prepared for the flow past cube problem.

horseshoe vortex system can be predicted by all models (see Figure 6). However, some remarkable qualitative differences can be observed downstream. In contrast with the SSG closure model, the LRR-ω model was not able to reproduce properly the slanted horseshoe vortex structures behind the cube. The modified SSG/LRR-ω model resulted qualitatively better agreement with the expected vortical flow structures, however the hairpin vortex system on the spanwise faces of the cube were missing and not captured by the SSG model.

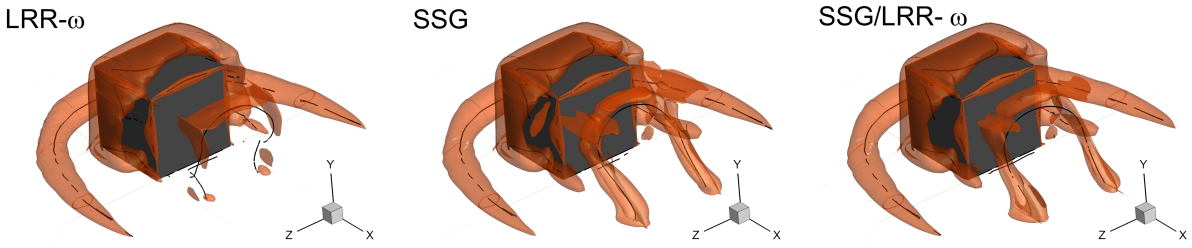


Figure 6: Computed iso-surfaces of Q-criterion for the flow past cubic obstacle problem.

Two mean flow quantities have been compared to the experimental data of Martinuzzi and Tropea [3] (see Table 2). These quantities are the flow separation point upstream and the reattachment location behind the cube in the mid-plane ($z/H = 0$), respectively. In general, all considered models produced remarkable differences for these measures. The separation location was predicted with at least approximately 13% relative error, and the downstream flow quantity was computed with even worse fidelity. These observations confirm again how difficult is to capture such a complex flow around a bluff body.

Configuration	$x_S/H[-]$	$x_R/H[-]$	$\lvert r_S\rvert[\%]$	$\lvert r_R\rvert[\%]$
Experiment	-1.040	2.612	$-$	$-$
SSG	-1.183	3.148	13.75	20.52
LRR-ω	-0.844	3.426	18.85	31.16
SSG/LRR-ω	-1.174	3.207	12.89	22.78

Table 2: Computed flow separation and reattachment locations along the X coordinate direction in the $z/H = 0$ plane, compared with the experimental data of Martinuzzi and Tropea [3].

The normalised streamwise velocity distributions have been compared at the $z/H = 0$ plane at three distinct x coordinates with the experimental data of Martinuzzi and Tropea [3] and the LES solution of Sedighi and Farhadi [14] (see Figure 7). It can be seen that none of the models was particularly successful again. At the $x/H = 0.5$ location, the SSG model slightly

under-predicted the strength of the mean flow above the cube and produced stronger reversed-flow near the cube surface. In contrast with that, the LRR-ω model resulted a stronger mean flow and a weaker recirculatory region. The modified SSG/LRR-ω performed again in between these two approaches indicating the inhereted mixed model properties. At the trailing edge of the cube ($x/H = 1.0$), larger recirculatory zone has been obtained with all RSMs, and it can be seen that even the reference LES simulation produced serious discrepancies. One edge distance downstream of the cube ($x/H = 2.0$), the flow is still massively separated posing serious challenges for all models. Note that the LRR-ω model resulted oscillatory velocity distribution which has not been observed with the SSG and SSG/LRR-ω models. The SSG RSM performed better in terms of agreement with experimental data compared to the hybrid model.

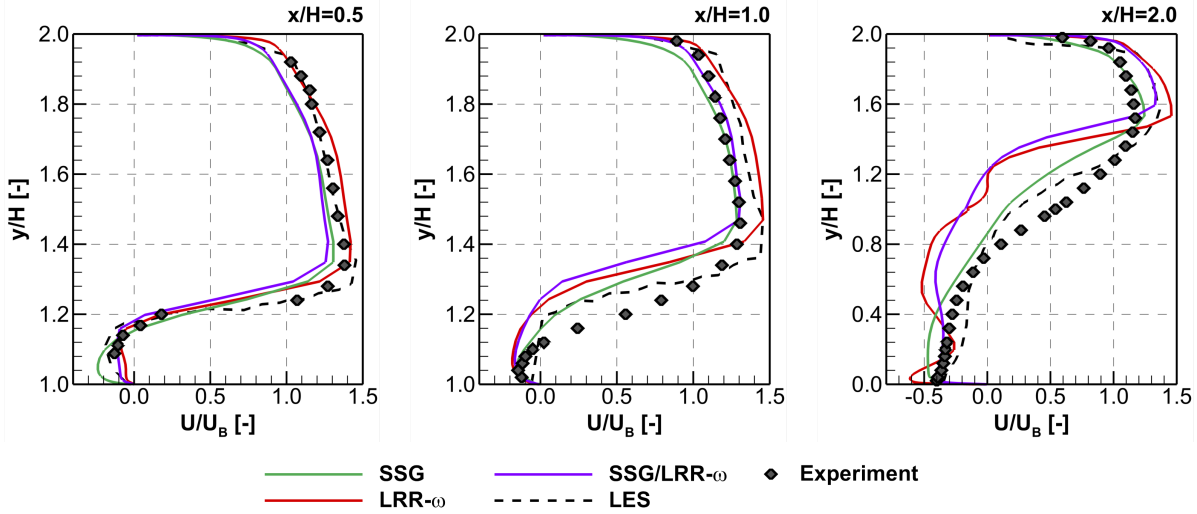


Figure 7: Comparison of normalized streamvise velocity component in the $z/H = 0$ plane with the measured data of Martinuzzi and Tropea [3] and LES solution of Sedighi and Farhadi [14].

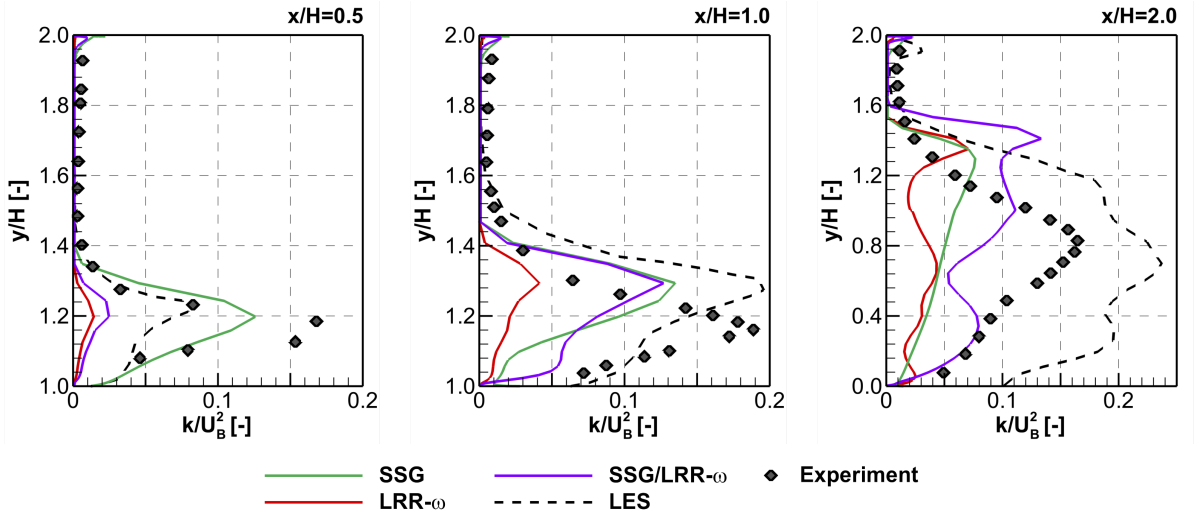


Figure 8: Comparison of normalized turbulent kinetic energy in the $z/H = 0$ plane with the measured data of Martinuzzi and Tropea [3] and LES solution of Sedighi and Farhadi [14].

The normalised turbulent kinetic energy distribution have been investigated along the previously shown locations (see Figure 8). In general, all Reynolds stress models under-predicted the peak magnitudes which indicates that the investigated RSMs produced excessive dissipation of the turbulent flow structures. This dissipative behaviour was the most striking for the LRR-ω model and the less drastic for the SSG model. The hybrid closure RSM performed in between these aforementioned two RSMs. It is important to note again that large errors have been obtained against the LES results of Sedighi and Farhadi [14].

4 CONCLUSIONS

In this work, a modified hybrid SSG/LRR-ω RSM approach with a simplified diffusion model has been investigated through the classical cubic obstacle problem in a three-dimensional channel flow of Martinuzzi and Tropea [3]. This model is blending the advantageous features of two distinct RSMs: a) the LRR-ω model, which is used to compute simple wall bounded flows, and b) the SSG model, which was developed for shear flows. Preliminary performance analysis on the problem of a flow over a flat plate at zero pressure gradient showed that the proposed RSM was able to reproduce law-of-the-wall accurately. It was also observed that the hybrid closure provides the freestream independence of the specific turbulent dissipation rate ω, which means that the results at near the wall was not sensitive to the ω quantity specified outside of the shear layer. In addition to this, the modified hybrid SSG/LRR-ω model has been assessed for the classical bluff body problem of a flow past cube in a three-dimensional channel flow. It was observed that the accuracy of the investigated RSM model was in between the obtained results of the LRR-ω and SSG approaches to predict a complex vortical flow around a bluff body. Since the original version of the hybrid SSG/LRR-ω model proposed by Cecora et al. [1] was primarily developed for aeronautical applications, therefore further improvements for bluff body aerodynamics are still required. The numerical results indicate that existing RANS models should also be further improved for capturing the aforementioned complex flow structures, furthermore it can be seen that LES techniques, which requires orders of magnitude higher computational effort, are challenged to predict these flows accurately. Overall, it can be concluded that the modified hybrid SSG/LRR-ω RSM approach with a simplified diffusion model exhibited stable numerical behaviour for computing a complex flow around a bluff body.

REFERENCES

[1] Cecora, R.-D., Radespiel, R., Eisfeld, B., Probst, A., Differential Reynolds Stress Modeling for Aeronautics. *AIAA Journal*, **53**(3), 739–755, 2015.

[2] Menter, F. R., Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications. *AIAA Journal*, **32**(8), 1598–1605, 1994.

[3] Martinuzzi, R., Tropea, C., The flow around surface-mounted prismatic obstacles placed in a fully developed channel flow. *Journal of Fluids Engineering*, **115**, 85–91, 1993.

[4] Launder, B. E., Reece, G. J., Rodi, W., Progress in the development of a Reynolds-stress turbulence closure. *Journal of Fluid Mechanics*, **68**, Part 3, 537–566, 1975.

[5] Eisfeild, B., Brodersen, O., Advanced Turbulence Modelling and Stress Analysis for the DLR-F6 Configuration. *AIAA Paper* 2005–4727, 2005.

[6] Speziale, C. G., Sarkar, S., Gatski, T. B., Modelling Modelling the Pressure-Strain Correlation of Turbulence - An Invariant Dynamical Systems Approach. *NASA Contractor Report 181979*, *ICASE Report*, **90-5**, 1990.

[7] Gibson, M. M., Launder, B. E., Ground Effects on Pressure Fluctuations in the Atmospheric Boundary Layer. *Journal of Fluid Mechanics*, **86**, 491–511, 1978.

[8] Wilcox, D. C., Re-assessment of the scale-determining equation for advanced turbulence models. *AIAA Journal*, **26**(11), 1299–1310, 1988.

[9] Wilcox, D. C., *Turbulence Modeling for CFD, 2nd Edition*. DCW Industries, La Canada, 1998.

[10] ANSYS Inc., ANSYS FLUENT Theory Guide. *ANSYS*, Release 14.0, November, 2011.

[11] Coles, D. J., Fluid Mechanics. *RAND Corp. Report*, **R-403-PR**, 191–226, 1962.

[12] von Kármán, Th., Turbulence and Skin Friction. *Journal of the Aeronautical Sciences*, **1**(1), 1–20, 1934.

[13] Rumsey, C. L., Turbulence Modelling Resource. *NASA, Langley Research Center*, Available online: http://turbmodels.larc.nasa.gov/, cited on 20^{th} of February 2016.

[14] Sedighi, K., Farhadi, M., Three-Dimensional Study of Vortical Structure Around a Cubic Bluff Body in a Channel. *Facta Universitas, Mechanical Engineering*, **4**(1), 1–16, 2006.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL ANALYSIS OF RAREFIED GAS FLOWS USING THE ACADEMIC CFD CODE GALATEA

Angelos G. Klothakis[1], Georgios N. Lygidakis[2] and Ioannis K. Nikolos[3]

[123]Technical University of Crete
School of Production Engineering and Management
University Campus, Chania, GR-73100, Greece
e-mail: [1]anklothakis@isc.tuc.gr,[2]glygidakis@isc.tuc.gr,[3]jnikolo@dpem.tuc.gr

Keywords: Rarefied Gas Flows, Navier-Stokes Equations, Velocity Slip Conditions, Temperature Jump Conditions, Direct Simulation Monte Carlo (DSMC).

Abstract. *During the last decades considerable efforts have been exerted for the development of micro air vehicles as well as microelectromechanical systems in general, for a wide range of applications. However, such systems involve microscale rarefied gas flows, which appear to be significantly different comparing to flows at macroscale and continuum regime; it is this the reason the Navier-Stokes equations fail to simulate such phenomena without further modification. To this end, the enhancement of the in-house academic Computational Fluid Dynamics solver Galatea to encounter such simulations is reported in this study. In case of rarefied gas flows and particularly for fluids in slip flow regime (Knudsen number greater than 0.01) the no-slip condition on solid wall surfaces is no longer valid; hence, velocity slip conditions as well as temperature jump ones have to be included instead. Furthermore, to increase accuracy at the same region the second-order accurate spatial slip model of Beskok and Karniadakis has been incorporated, which avoids the numerical difficulties, entailed by the evaluation of the second derivative of slip velocity when complex geometries along with unstructured hybrid grids are encountered. Due to oscillations that might appear, especially during the initial steps of the iterative procedure, a normalization scheme is additionally employed, to allow for the gradual increase of the corresponding slip/jump values. Galatea has been validated against a benchmark test case concerning rarefied laminar flow (inside the slip flow regime) over a wing with a NACA0012 airfoil in different angles of attack. The obtained results were compared with those of a reference solver, and with those obtained with the paralleld open-source kernel SPARTA, based on the Direct Simulation Monte-Carlo method. According to this last approach, the flow domain is divided into a finite number of computational cells, while the required sample macroscopic flow properties are retrieved assuming intermolecular collisions of the simulated particles inside such cells. An excellent agreement was achieved between the results obtained by Galatea and SPARTA as well.*

1 INTRODUCTION

During the past decade a significant effort has been exerted by various researchers for the development of micro air vehicles as well as microelectromechanical systems in general [1, 2, 3]. However, such systems may involve rarefied gas flows, which appear to be considerably different, compared to flows at the continuum regime; thus, the Navier-Stokes PDEs (Partial Differential Equations), used at macroscale CFD (Computational Fluid Dynamics) solvers, appear to fail simulating such phenomena without further adaptions and modifications [1]. In practice the rarefied gas flows are categorized depending on the computed Knudsen number, a classification originally proposed by Schaaf and Chambre [4]: For Knudsen numbers less than *1.0E-2* (continuum regime) the Navier-Stokes PDEs are valid without any further modification, allowing ordinary CFD (Computational Fluid Dynamics) solvers to be employed. Nevertheless, for values between *1.0E-2* and *1.0E-1* (slip flow regime) special treatment of wall boundary conditions is required; velocity slip conditions as well as temperature jump ones have to be applied (as in this work) [1]. If greater than *1.0E-1* Knudsen numbers are encountered (transition regime and free molecular regime) the rarefaction effects become the sovereign ones, necessitating for alternative methodologies, depending on the DSMC (Direct Simulation Monte Carlo) [5] approach or the solution of the Boltzmann Equation [6]. Figure 1 includes a schematic representation of the previously described classification. Although the latter method (based on the Boltzmann Equation) can actually be employed throughout all the regimes [1], the CFD approaches are usually preferred at continuum and slip flow ones; the implementation of Boltzmann Equation at those regimes calls for very large numbers of particles and consequently for excessive computational load and storage.

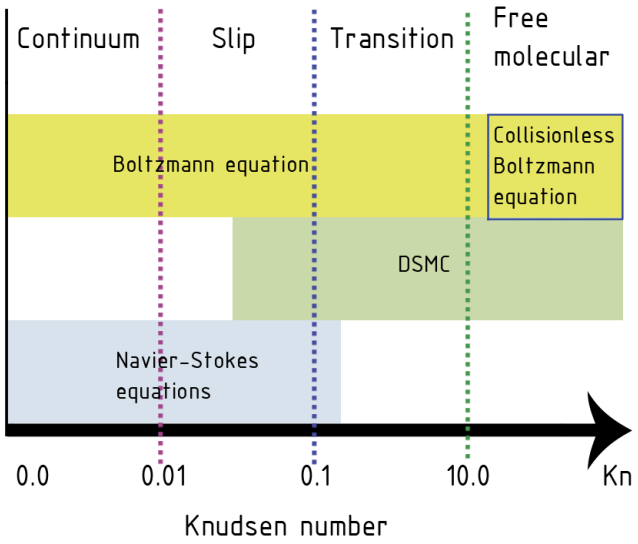


Figure 1: Classification of flow regimes

As mentioned above, in the slip flow regime (Knudsen number between *1.0E-2* and *1.0E-1*) the Navier-Stokes PDEs remain valid, although special treatment of solid wall boundaries is required. The no-slip condition cannot longer be applied on such surfaces; velocity slip conditions as well as temperature jump ones have to be imposed instead [1, 2, 3]. Despite the main concept of the aforementioned conditions was initially proposed by Navier [1], the first corresponding mathematical model was introduced by Maxwell in 1879 [7]; according to this approach the

values of slip velocity and temperature are defined with the corresponding normal to the boundary surface gradients [3]. Since then, various studies have been conducted on the prediction of the attitude of rarefied gases,focusing on the utilization of two- or three-dimensional computational fields, structured or unstructured grids, various geometries, etc., [8, 9, 10, 11, 12, 13]. Furthermore, significant efforts have been exerted for the development of higher-order spatial schemes to increase the accuracy of the final steady-state solution at the region of the solid wall boundaries [14, 15, 16, 17]. Among them the model of Beskok and Karniadakis appears to be especially attractive, as it avoids the numerical difficulties, entailed by the evaluation of the second derivative of slip velocity when complex geometries along with unstructured hybrid grids are encountered [14, 15]. As the aforementioned boundary conditions are imposed in a wall-function mode, they are susceptible to produce residual oscillations, especially during the initial steps of an explicit iterative solution; a remedy to this drawback was proposed by Ferras et al. [18], where a normalization scheme is considered, allowing for the gradual increase of the slip/jump values. Finally, few quite many studies include comparisons of the DSMC and the modified by the slip/jump boundary conditions Navier-Stokes PDEs approaches (as in this work), demonstrating the potential of such methodologies to effectively simulate rarefied gas flows at slip flow regime [1, 14, 19].

Despite the development of slip velocity and temperature jump boundary conditions allowed CFD solvers to extend their applications at slip flow regime, the latter appear to be inadequate to predict flows with larger values of Knudsen number (greater than *1.0E-1*) [1]. It is this observation along with the excessive computational requirements of methodologies depending on the Boltzmann equation that created the need for the development of the DSMC approach, originally introduced by Bird [5]. Mainly due to the statistical-stochastic nature of this modelling type, it wasn't initially accepted by the scientific community, confining its implementation in relatively simple simulations [5]. Nevertheless, since then it made its way, extending its applications in various engineering problems and demonstrating its capability for accurate flow predictions. The main drawback of the DSMC approach is the relatively excessive computational requirements in test cases involving high pressure and large-scale computational fields. This drawback was alleviated by the advance of the available computer systems and subsequently the capability of parallel processing. The first reported parallel DSMC approach (a NASA production code nowadays) was the DSMC Analysis Code (DAC), developed by LeBeau [20, 21], while since then quite many such algorithms have been developed, such as the SMILE [22], the MONACO [23], the ICARUS [24], the MGDS [25], the dsmcFoam [26] and the most recent one the open-source kernel SPARTA [27] (used also in this work).

In this study the enhancement of the in-house academic CFD solver *Galatea* to simulate rarefied gas flows inside the slip flow regime (Knudsen number between *1.0E-2* and *1.0E-1*) is reported [28]. *Galatea* employs the Navier-Stokes PDEs in dimensionless formulation along with a node-centered finite-volume scheme to predict inviscid, viscous laminar, and viscous turbulent flows (with RANS-Reynolds Averaged Navier-Stokes approach) of compressible fluids on tetrahedral or hybrid unstructured grids [28]. For rarefied gases the capability of implementing the aforementioned slip velocity and temperature jump boundary conditions [1] on solid wall surfaces has been incorporated. In order to increase the accuracy at the same regions, a second-order accurate slip scheme has been included. Particularly, the slip model of Beskok and Karniadakis [14, 15] was adopted due to its relatively easy implementation on unstructured, tetrahedral or hybrid grids; it can overcome the numerical difficulties, entailed by the evaluation of the second derivative of slip velocity [14]. Furthermore, the normalization scheme, reported in [18], was included to mitigate the excessive oscillations, caused by the Dirichlet-type of the

slip/jump boundary conditions, especially during the initial steps of the iterative solution procedure. The proposed solver was validated against a benchmark test case concerning rarefied laminar flow over a wing with a NACA0012 airfoil in different angles of attack [13, 19, 29]. The obtained results have been compared with those of another reference solver, confirming its potential for such simulations. In addition, further evaluation of the *Galatea* algorithm was succeeded with the parallel open-source kernel SPARTA [27], a DSMC code as mentioned above. An excellent agreement was achieved between the results obtained by *Galatea* and SPARTA as well.

The structure of this paper is as follows: In Section 2 the *Galatea* solver is described in brief, while the main attention is directed towards the analysis of the incorporated slip velocity and temperature jump conditions. Section 3 contains an outline of the DSMC approach, as well as of the employed open-source kernel SPARTA. In Section 4 the extracted by *Galatea* and SPARTA results against a test case considering rarefied laminar flow over a rectangular wing in different angles of attack are presented and compared, while Section 5 includes the corresponding conclusions and some information on ongoing work.

2 THE CFD APPROACH

2.1 The *Galatea* code

In this study the in-house academic CFD code, named *Galatea* [28], was used, enhanced though with appropriate modifications to account for rarefied gas flows. The proposed solver employs a node-centered finite-volume method on three-dimensional unstructured grids, composed of tetrahedral, prismatic and pyramidical elements, to simulate inviscid, viscous laminar and viscous turbulent compressible fluid flows. For turbulence prediction the RANS approach is implemented along with appropriate statistical two-equation models, namely the $k-\varepsilon$ [30], the $k-\omega$ [31] and the SST (Shear Stress Transport) [32] models. An upwind method is implemented for the computation of the inviscid fluxes, applying the Roe's [33] or the HLLC (Harten-Lax-van Leer-Contact) [34] approximate Riemann solvers, coupled with a higher-order accurate spatial scheme, based on the MUSCL (Monotone Upwind Scheme for Conservation Laws) approach along with appropriate slope limiters, namely the Van Albada-Van-Leer [35], the Min-mod [36], the Barth-Jespersen [37] and the MLP-Venkatakrishnan (Multi-dimensional Limiting Process-Venkatakrishnan) [38, 39] limiter. For the computation of the viscous fluxes the required velocity and temperature gradients are evaluated with an element-based (edge-dual volume) or a nodal-averaging method [28], either selected by the user. Time integration and iterative approximation of the final steady-state solution is succeeded with either an explicit scheme, applying a second-order temporal accurate four-stage Runge-Kutta (RK(4)) method, or an implicit one, implementing the Jacobi or the Gauss-Seidel algorithm [28]. In order to increase the efficiency of the proposed solver, appropriate acceleration techniques have been incorporated, such as an edge-based data structure [40], a local time-stepping technique [40], parallel processing based on the domain decomposition approach and the MPI (Message Passing Interface) library functions [28], and an agglomeration multigrid scheme [28, 41]. Finally, an h-refinement technique has been included to increase accuracy at pre-selected regions of the examined grid, by enriching them with more nodes during the solution procedure, while simultaneously avoiding the generation of dense meshes from the very beginning [28]. Further details for the *Galatea* solver can be found in [28] and [41].

2.2 Velocity slip and temperature jump conditions

As mentioned in Introduction, the no-slip conditions on solid wall surfaces appear to fail in the slip flow regime; velocity slip and temperature jump wall functions have to substitute them [1, 19]. Although their main idea was initially captured by Navier [1], the first such model was reported by Maxwell in 1879 [7]. According to this approach, the values of slip velocity and temperature on wall boundaries are defined with the corresponding normal to the boundary surface gradients. A schematic representation of the previous state for slip velocity values is illustrated in Figure 2; moreover it highlights the difference between the no-slip and the slip boundary conditions. Mathematically, the Maxwell model is described for the dimensionless slip velocity U_s as follows [1, 7, 12, 14]

$$U_s - U_w = \frac{2-\sigma_u}{\sigma_u} Kn \frac{\partial U_s}{\partial n} + \frac{3}{2\pi} \frac{(\gamma-1)}{\gamma} \frac{Kn^2 Re}{E_c} \frac{\partial T}{\partial s} \tag{1}$$

where U_w denotes the velocity on the wall surface, K_n is the Knudsen number and $\frac{\partial U_s}{\partial n}$ the transverse velocity gradient, i.e., the derivative of the tangential slip velocity normal to the wall surface (denoted by vector n) [12]. Further, σ_u is the tangential momentum accommodation coefficient (equal to unity in this study) [13, 19], which actually models the momentum exchange of the gas molecules impinging on the solid boundaries. It depends on the surface quality; small values of this constant can considerably increase the slip velocity even for low Knudsen numbers [14]. The second term includes the values of the dimensionless Reynolds Re and Eckert E_c numbers, along with the ideal gas constant γ and the tangential to the wall surface derivative of the temperature. It actually represents the slip velocity contribution induced by the thermal creep; it appears usually to be relatively small, due to the second-order in Knudsen number as well as due to the fact that in common engineering problems there is a small temperature change across the wall surfaces; therefore, in this study it is neglected [12].

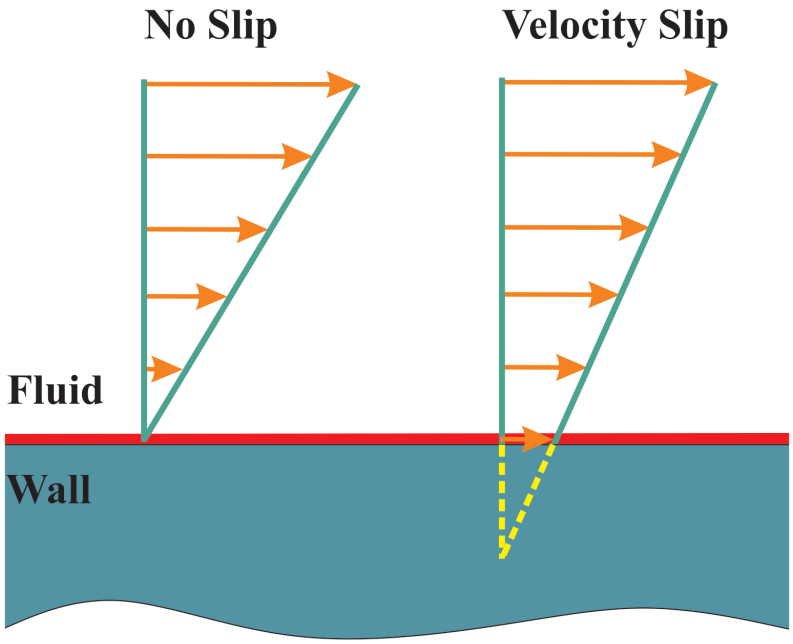


Figure 2: Velocity slip boundary conditions.

The temperature jump boundary conditions are defined similarly to the slip velocity ones as [12]

$$T_s - T_w = \frac{2-\sigma_T}{\sigma_T}\frac{2\gamma}{\gamma+1}\frac{Kn}{Pr}\frac{\partial T}{\partial n} \tag{2}$$

where T_s represents the slip temperature, T_w the wall temperature and $\frac{\partial T}{\partial n}$ the derivative of the temperature normal to the wall surface. Pr is the dimensionless Prandtl number, while σ_T is the thermal or energy accommodation coefficient; similarly to the momentum coefficient the thermal one depends on the surface quality [14], while in this work it was set equal to unity [13, 19].

In order to increase the accuracy of the final steady-state solution at the solid walls' region, several different second-order accurate spatial schemes have been reported in the open literature [14, 15, 16, 17]. Their main concept is based on the reconstruction of the first term of Equation (1) with the Taylor series [16, 17]. Nevertheless, such a reconstruction is not a straightforward procedure when complex three-dimensional geometries, along with unstructured hybrid grids, are encountered; the calculation of the second derivative of the tangential velocity in the normal direction to the surface may introduce numerical errors, despite the *a priori* computational difficulty it entails in such test cases [14]. Therefore, in this study the approach, introduced by Beskok and Karniadakis [14] was adopted, according to which the second-order accurate slip velocity can be computed as

$$U_s - U_w = \frac{2-\sigma_u}{\sigma_u}\frac{Kn}{(1-bKn)}\frac{\partial U_s}{\partial n} \tag{3}$$

where b is the slip coefficient, which can be defined either experimentally or from other methodologies employing the Boltzmann equation or the DSMC approach [14]. It actually represents the vorticity flux into the surface divided by the vorticity of the flow field on the surface, obtained by the no-slip approximation; in this study it was set equal to *-1.0* [14].

The velocity slip and temperature jump boundary conditions are implemented in a Dirichlet way, hence they are susceptible to produce oscillations during the iterative procedure (especially during the initial steps of an explicit iterative scheme), or even lead simulation to fail [18]. Therefore, a normalization scheme was incorporated in *Galatea* solver, allowing for the gradual increase of the slip velocity and temperature values; mathematically, it is described as follows [18]

$$U_s^i = \alpha U_s^{i-1} + (1-\alpha)U_s^i \tag{4}$$

where α is the normalization coefficient, taking values smaller than unity; in this work it was set equal to *0.95*. In case a multigrid accelerated simulation is performed, the aforementioned slip velocity and temperature values are computed only at the finest resolution, while at next they are restricted to the coarser multigrid levels similarly to the rest nodal values [41].

3 THE DSMC APPROACH

3.1 The particle method

The DSMC method, originally introduced by Bird [5], is a stochastic-type scheme using a large number of particles to simulate gas flows, especially rarefied ones. Each of the aforementioned particles, called simulator particle, actually represent actually a large number of real particles [5]. According to the main idea of this methodology the computational domain is divided in Cartesian cells, which provide the geometric boundaries and volumes required to

sample the macroscopic properties of the flow. The latter are acquired taking into account the movements and collisions of the simulator particles; therefore, the time step has to be defined in a way allowing the movement and collisions of each particle to be decoupled. In conclusion, the DSMC algorithm can be divided in four main steps: a) Given a time step the particles move to their new positions. b) They are indexed into the computational cells. c) The collision pairs are selected and the corresponding intermolecular collisions are performed on a probabilistic manner. d) Since the temporary steady-state solution has been achieved, the required macroscopic flow properties, such as velocity and temperature, are obtained by averaging temporally (for several time steps) the microscopic properties of the particles in the cells [5]. The pre-described procedure is presented in the flow chart of Figure 3.

3.2 The SPARTA DSMC kernel

As mentioned in Introduction, the parallel open-source kernel SPARTA is used in this study, where SPARTA stands for Stochastic Parallel Rarefied-gas Time-Accurate Analyzer. It was developed in Sandia Laboratories by Gallis and Plimpton [27] and distributed as an open source code under the terms of the GPL license. It is capable of both serial and parallel processing, the latter based on the domain decomposition approach and the MPI communication protocol (similarly to the *Galatea* solver). Despite its robustness and flexibility it is not addressed to non-experienced users, as it calls for a thorough understanding of the DSMC approach; it executes by reading the commands in the script line by line [27].

3.2.1 Problem definition

In order to run a simulation with SPARTA the corresponding parameters have to be defined first, i.e., the simulation box (computational domain), the grid, the internal boundaries (inside the simulation box), the particle species (e.g., oxygen, nitrogen, ions) and the initial population of the particles. For a two-dimensional run, the simulation box is defined by the coordinates of the boundaries; minimum and maximum coordinates in x- and y- axis. For a three-dimensional simulation the coordinates of two more boundaries have to be defined (including coordinates in z-axis). As far as the computational discretization is concerned, SPARTA employs a hierarchical Cartesian grid strategy. The simulation domain is considered initially as a aingle grid cell (level 0). The aforementioned cell is divided in N_x by N_y by N_z cells at level 1, while at next each of these cells can be further divided into smaller ones at different levels. The properties of the particle species, such as their diameter and molecular weight, are acquired from an additional appropriately formatted file. Finally, the initial population of the particles is defined from the *Fnum* number, denoting the number of real particles represented by a simulator particle. As far as initialization of the procedure is concerned, the corresponding velocities are retrieved from a Maxwellian speed distribution function.

3.2.2 Settings

Since the problem definition stage is accomplished the settings of the input script have to be defined. In this section various parameters are determined, i.e., the collision model, the load balancing (static at the beginning and dynamic during the iterative procedure), the boundary conditions, the surface reflection models, the sample averaging-type and the desired simulation outputs [27]. This stage is of extreme importance, as it influences the accuracy of the final results as well as the required computation time. For example, the time step, the used flow

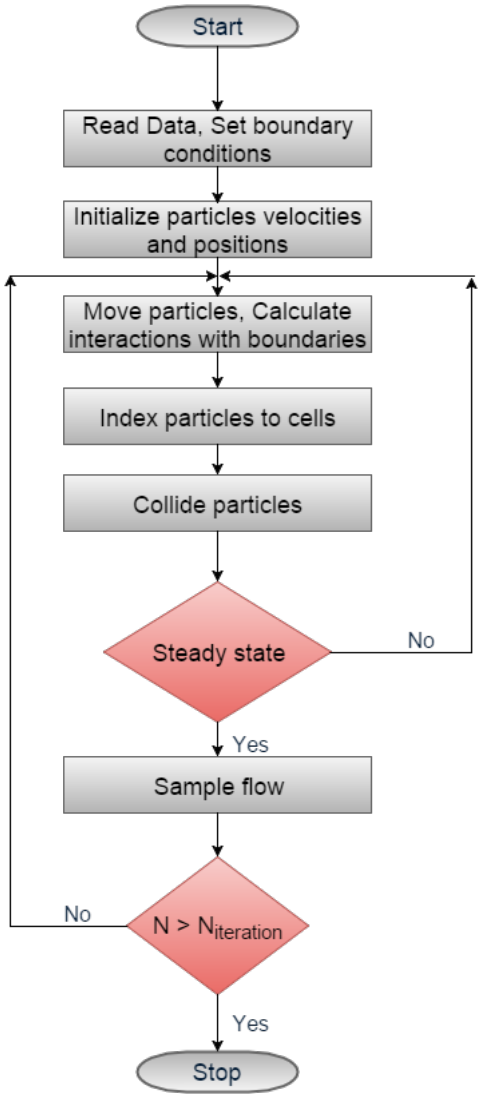


Figure 3: The DSMC Method

field variables, the characteristics of the boundary surfaces, and the sampling intervals included in this script have a serious impact on the extracted solution. Especially the latter parameter, defining the intervals at which the macroscopic flow properties will be extracted, appears to be very important, as an inappropriate value can lead to incorrect results or to an efficiency drop.

Furthermore, at this stage dynamic load-balancing type is defined. SPARTA can adjust the assignments of grid cells and their corresponding particles to each processor in order to spread more efficiently the computational load. Since the dynamic load-balancing type is determined, the cells assigned to each processor are either clumped or dispersed. Actually, there are three ways the aforementioned balancing can be succeeded; it can be performed depending on clumped assignments of child cells to each processor, dispersed assignments of child cells to each processor, or on a recursive coordinate bi-sectioning approach [27]. According to the first method each grid cell is assigned randomly to one of the processors; as a result, the processors do not get necessarily the same amount of grid cells. If the second strategy is followed, each processor selects randomly another processor to assign its first grid cell to; at next, it loops over its grid cells and assigns each of them to consecutive processors. Unfortunately, this method-

Case	$Mach_\infty$	Re_∞	Kn_∞	V_∞, m/s	α, deg	ρ_∞, kg/m^3	T_∞, K	T_w, K
A	2.0	106	0.026	506	0	6.026×10^-5	161	290
B	2.0	106	0.026	506	10	6.026×10^-5	161	290

Table 1: Flow conditions for Cases A and B.

Number density, m^-3	F_{num}	Time-step, sec	Transient period	Sample period
1.296×10^{21}	1.2×10^{15}	3.0×10^{-7}	140000	40000

Table 2: Parameters used for the SPARTA code runs.

ology suffers from the same to the first approach shortcoming. Finally, according to the third scheme, the recursive coordinate bisectioning method (RCB), the processors are assigned compact clumps of grid cells [27]. For every cell a weight coefficient is defined, hence, dynamic load balancing aims to assign equal total weights to all the available processors. This weight coefficient can be defined in various ways, e.g., if it is set equal to unity, each processor is assigned the same number of cells, while if it is set equal to the number of the included particles, the same number of particles is assigned to all the processors. Dynamic load-balancing is strongly influenced by two more parameters, namely the imbalance factor and the frequency interval. The first one defines the maximum number of particles run by a processor divided by the average number of particles per processor, while the second the frequency in which the imbalance of the simulation is checked for change. Both the aforementioned coefficients have to be wisely selected in order communication overheads to be avoided during the simulation. For the test cases encountered in this study, an imbalance factor of *20%-30%* (defined as *1.2* or *1.3* in the input script) was found to be sufficient enough, without causing additional communication load, while the imbalance was checked for possible change two times during the simulation.

4 NUMERICAL RESULTS

The *Galatea* solver was validated against a benchmark test case considering rarefied laminar compressible fluid flow over a rectangular wing with a NACA0012 airfoil [13, 19, 29]. The Mach number was assumed equal to *2.0*, while the Knudsen and Reynolds numbers equal to *0.026* and *106*, respectively [19]; the latter was based on the mean aerodynamic chord, which was redimensionalized to become equal to unity due to the dimensionless formulation of the proposed code. Two different angles of attack were examined, namely *0*° (Case A) and *10*° (Case B) [13, 19, 29]. Freestream temperature was assumed equal to $161K$, while the corresponding value on the wing surface equal to $290K$. The aforementioned flow conditions are summarized in Table 1; for *Galatea* solver simulations they were used though in dimensionless formulation [28]. For the representation of the computational field a hybrid unstructured grid was constructed, composed of *305,978* nodes, *566,245* tetrahedra and *394,760* prisms, the latter located on the solid wall region (wing) to allow for the effective prediction of the boundary layer region. Figure 4 illustrates the employed computational grid, while Figure 5 depicts its density on the symmetry surface.

For the computation of the inviscid fluxes the Roe's approximate Riemann solver along with a second-order spatial accurate scheme, coupled with the Min-mod limiter, was employed; the corresponding viscous fluxes were computed based on the nodal-averaging scheme. Time inte-

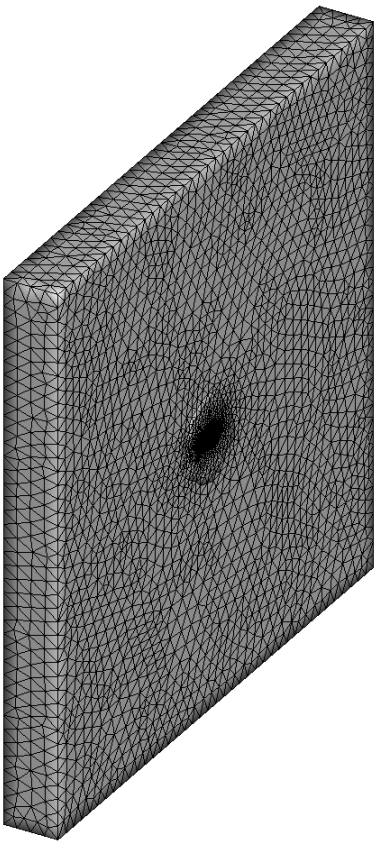


Figure 4: Computational grid used with the *Galatea* code.

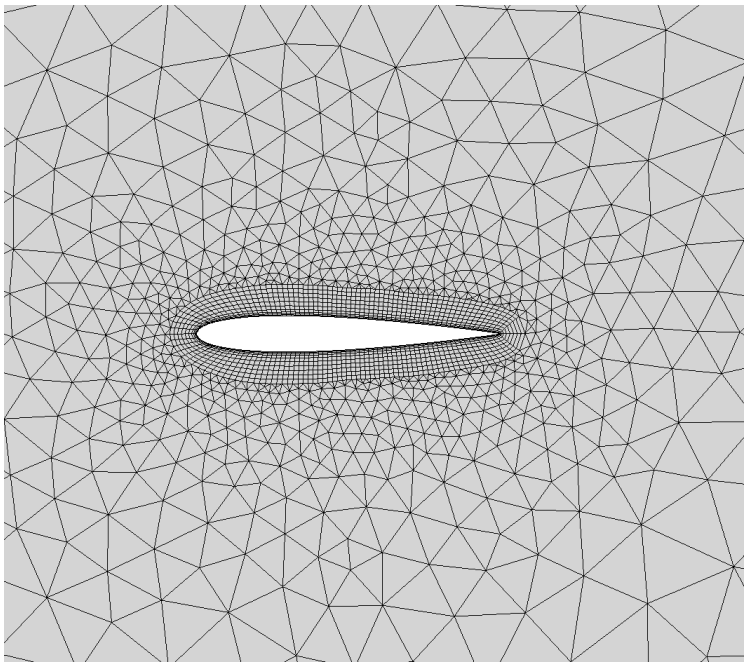


Figure 5: Density on the symmetry surface of the computational grid used with the *Galatea* code.

gration was succeeded via the incorporated explicit scheme (Runge-Kutta method) with a CFL number equal to *0.5*. Both the incorporated first- and second-order spatial accurate slip models were implemented [14]. However, no difference was identified between their results; this lack of difference is attributed to the relatively fine grid utilized. To accelerate the solution procedure on a (DELL^{TM} R815 Poweredge Server with four AMD Opteron^{TM} 6380 sixteen-core processors) the initial grid was decomposed in four sub-domains to be processed in parallel; two coarser levels were constructed for each of them applying the full-coarsening directional agglomeration strategy, in order the incorporated multigrid scheme to be applied [41].

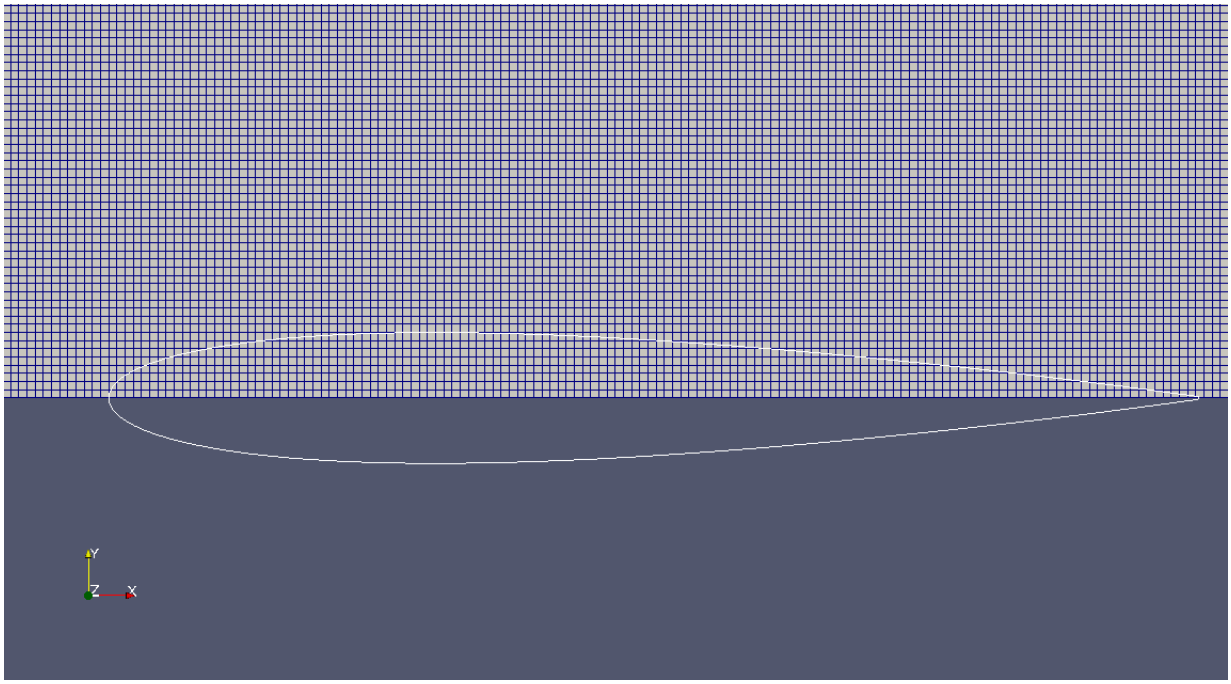
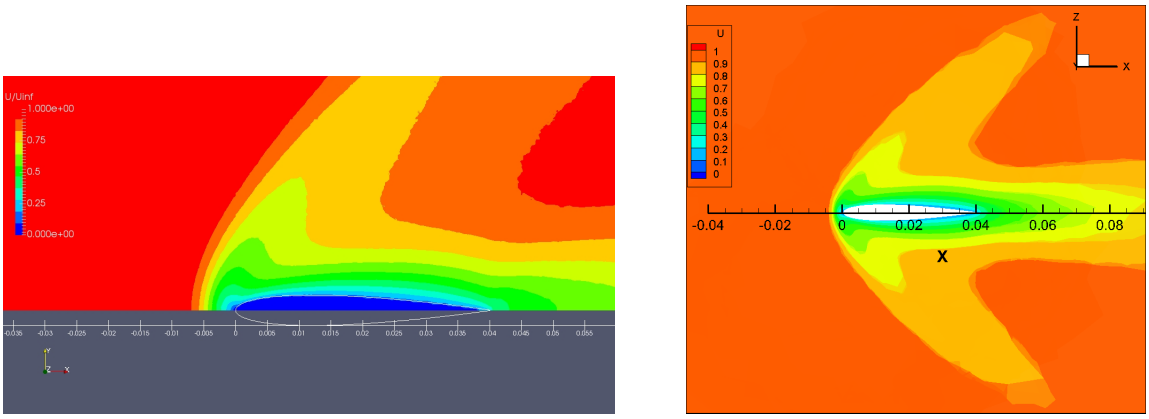


Figure 6: Computational grid used with the parallel open-source kernel SPARTA.

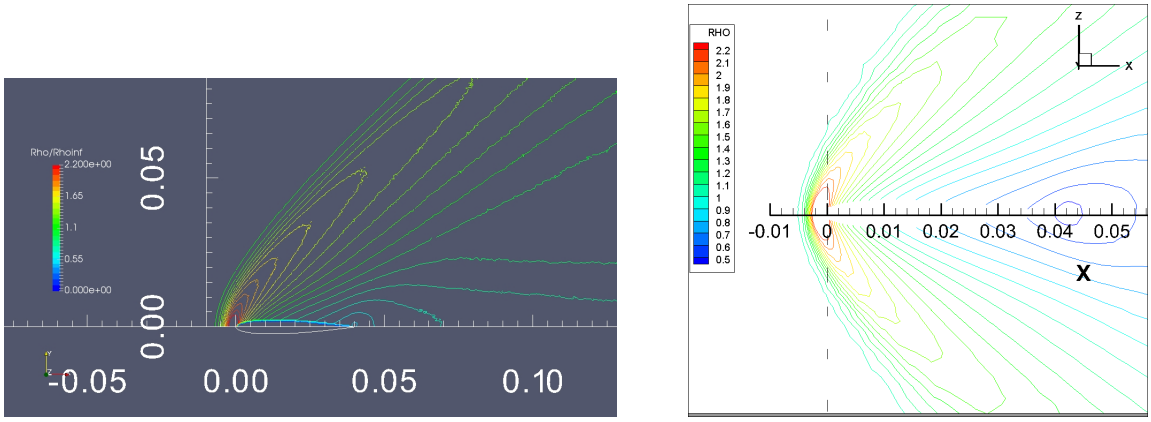
As mentioned in Introduction, the evaluation of the proposed solver was succeeded with a DSMC approach code, the parallel open-source kernel SPARTA [27]. The encountered computational domain was designed with dimensions *0.38m* and *0.24m* along the x- and y-axis, respectively; its discretization was succeeded with a Cartesian two-dimensional mesh, composed of *1200* cells along x-axis and *800* cells along y-axis. Considering that the mean free path in such test cases is approximately *0.000923m*, each grid cell was constructed to be more or less *1/3* the size of it, a common practice in such simulations [5]. Furthermore, the grid around the airfoil was generated such as no more than four points, used for airfoil representation (a total number of *1600* was used), to be included at each computational cell; as a result, a smoother discretization was achieved. Figure 6 illustrates a close-up view of the utilized computational grid around the NACA0012 airfoil.

The SPARTA code was run on the same to *Galatea* solver computer system with the flow conditions summarized in Table 1. As far as the simulation parameters are concerned, the number density was set equal to $1.296 \times 10^{21}\ m^{-3}$, while the time step equal to $3.0 \times 10^{-7} s$; the latter value along with the employed grid density ensures that each particle needs approximately three time steps to cross entirely a grid cell [5]. The simulation began with a transient period of *140,000* steps, producing the initial steady-state solution, while at next samples were taken for



(a) SPARTA. (b) *Galatea*.

Figure 7: Velocity contours (Case A).



(a) SPARTA. (b) *Galatea*.

Figure 8: Density contours (Case A).

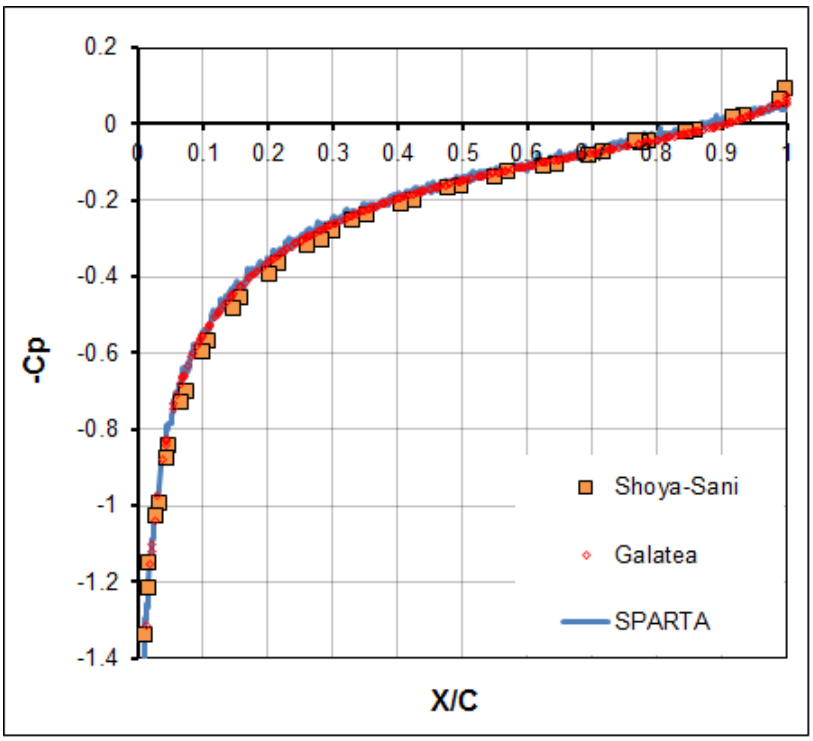


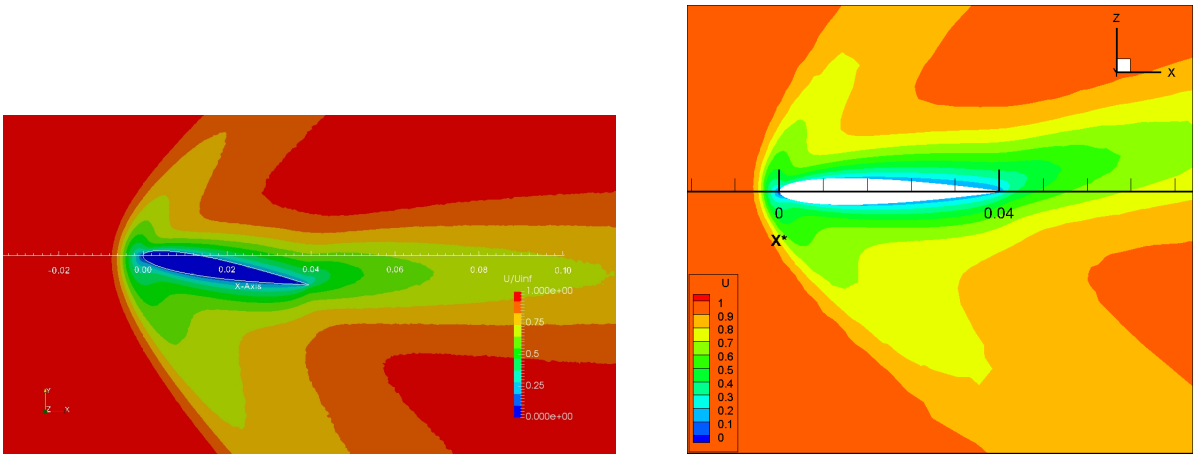
Figure 9: Distribution of pressure coefficient *Cp* around the NACA0012 airfoil (Case A).

additional *40,000* steps, aiming to reduce the statistical scattering error. The aforementioned parameters are outlined in Table 2. Considering that DSMC simulations depend strongly on the employed number of particles, several different numbers of them were tested (between 2×10^6 and 16×10^6) along with the aforementioned grid, as well as with a finer one; the latter was generated by refining each computation cell around the airfoil into *5x5* smaller ones. From this study the optimum *Fnum* (real particles per one simulator particle) of 1.2×10^{15} was derived, which subsequently produced a total number of 8×10^6 simulator particles for the whole simulation domain. As far as the grid is concerned, no difference was identified, hence, the initial mesh was selected. The previous parameters defining actually the DoFs (Degrees of Freedom) of the simulation were proved to produce the desired accuracy, avoiding though any excessive requirements for computational load and storage.

In Figure 7 the extracted velocity contours around the NACA0012 airfoil by both the *Galatea* and SPARTA solvers are presented for the first test case (Case A - angle of attack *0*°). As one can observe a satisfactory comparison is obtained, especially in the bow shock and stagnation regions, indicating qualitatively the potential of the proposed solver for such simulations. Furthermore, a very good agreement can be observed with the available experimental and numerical results, reported in [29] and [13, 19] respectively. The same similarity can be identified in Figure 8, depicting the corresponding density contours. It should be highlighted that dimensionless values are included in both the aforementioned figures. Figure 9 includes the derived distributions, compared with the one reported in [19]; the distribution of *Galatea* code simulation was extracted from the mid-span of the wing geometry.

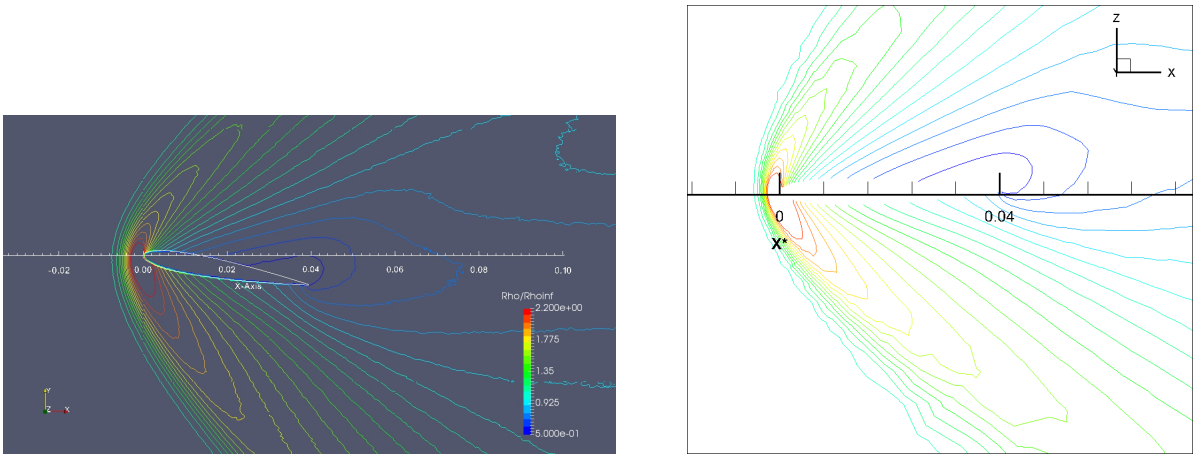
Similarly to Case A, Figure 10 illustrates the extracted velocity contours around the NACA0012 airfoil by both the *Galatea* and SPARTA codes for Case B (angle of attack *10*°). Once more a very good qualitative agreement is succeeded between the results of the employed solvers,

while the same holds in Figure 11 presenting the corresponding density contours. A quantitative comparison for the distributions of the pressure coefficient *Cp* around the NACA0012 airfoil, depicted in Figure 12. It is obvious that for both test cases (with angle of attack *0°* and *10°*) the employed codes produce almost identical results, despite they are based on completely different numerical approaches.



(a) SPARTA. (b) *Galatea*.

Figure 10: Velocity contours (Case B).



(a) SPARTA. (b) *Galatea*.

Figure 11: Density contours (Case B).

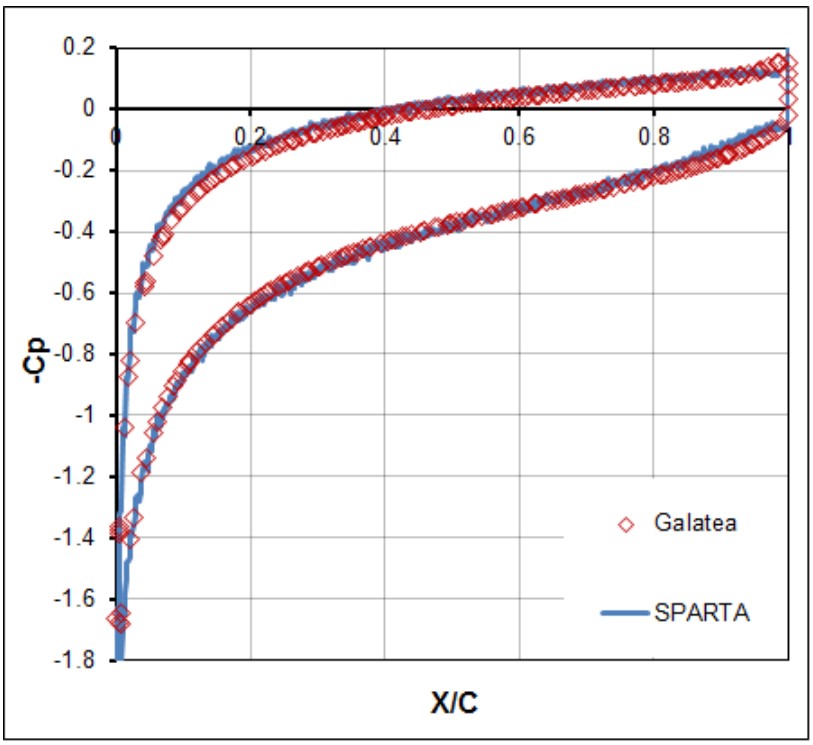


Figure 12: Distribution of pressure coefficient *Cp* around the NACA0012 airfoil (Case B).

5 CONCLUSIONS

In this study the enhancement of the in-house academic CFD solver *Galatea* to simulate rarefied gas flows in slip flow regime (Knudsen number between *1.0E-2* and *1.0E-1*) is reported [28]. Velocity slip and temperature jump models have been incorporated to this end, as the no-slip boundary conditions imposed on solid wall surfaces in flows in continuum regime appear to fail when rarefied gases are examined. The incorporated second-order accurate model didnt produce improved results; this lack of difference is attributed to the relatively fine grid utilized. In addition, the incorporated normalization scheme [18] was revealed a considerable enhancement, as the Dirichlet-type of the slip/jump boundary conditions was found to cause excessive oscillations without it, especially during the initial steps of the iterative solution procedure. The proposed solver was validated against a benchmark test case concerning rarefied laminar flow over a wing with a NACA0012 airfoil in different angles of attack [13, 19, 29]. The parallel open-source kernel SPARTA [27] was used for comparison purposes, while a very good agreement between the results of the two codes was achieved. Finally, ongoing work includes further assessment of the enhanced *Galatea* solver against more complex test cases of rarefied gas dynamics, as well as further examination of the incorporated second-order accurate slip model, utilizing coarser grids.

6 ACKNOWLEDGEMENTS

This work has been co-funded by the European Commission (European Regional Development Fund) and the Greek State, through the Operational Programme ”Competitiveness and Entrepreneurship (OPCE II 2007-2013), National Strategic Reference Framework - Research funded project: Development of a spark discharge aerosol nanoparticle generator for gas flow

visualisation and for fabrication of nanostructured materials for gas sensing applications DE_ SPARK_ NANO_ GEN", 11SYN_5_144, in the framework of the Action "COOPERATION 2011" — Partnerships of Production and Research Institutions in Focused Research and Technology Sectors.

REFERENCES

[1] W.M. Zhang, G. Meng and X. Wei, A review on slip models for gas microflows, *Microfluidics and Nanofluidics,* 2012.

[2] C.M. Ho and Y.C. Tai, Micro-electro-mechanical-systems (MEMS) and fluid flows, *Annual Review of Fluid Mechanics,* vol. 30, pp. 579-612, 1998.

[3] M. Gad-el-Hak, The fluid mechanics of microdevices The Freeman scholar lecture, *ASME Journal of Fluids Engineering,* vol. 121, 1999.

[4] S.A. Schaaf and P.L. Chambre, Flows of rarefied gases, *Princeton University Press,* Princeton, 1961.

[5] G.A. Bird, Molecular gas dynamics and the direct simulation of gas flows, *Clarendon Press,* Oxford, 1994.

[6] C.R. Lilley and J.E. Sader, Velocity profile in the Knudsen layer according to the Boltzmann equation, *Proceedings of the Royal Society A,* vol. 464, pp. 2015-2035, 2008.

[7] J.C. Maxwell, On stresses in rarified gases arising from inequalities of temperature, *Philosophical Transactions of the Royal Society,* vol.170, pp.231-256, 1879.

[8] N. Dongari, R. Sambasivam and F. Durst, Extended Navier-Stokes equations and treatments of micro-channel gas flows, *Journal of Fluid Science and Technology,* vol. 4, pp. 454-467, 2009.

[9] R.S. Myong, A computational method for Eu's generalized hydrodynamic equations of rarefied and microscale gasdynamics, *Journal of Computational Physics,* vol. 168, pp. 47-72, 2001.

[10] R.S. Myong, A generalized hydrodynamic computational model for rarefied and microscale diatomic gas flows, *Journal of Computational Physics,* vol. 195, pp. 655-676, 2004.

[11] R.S. Myong, J.M. Reese, R.W. Barber and D.R. Emerson, Velocity slip in microscale cylindrical Couette flow: The Langmuir model, *Physics of Fluids,* vol. 17, 2005.

[12] H.D. Madhawa Hettiarachchi, Mihajlo Golubovic, William M. Worek and W.J. Minkowycz, Three-dimensional laminar slip-flow and heat transfer in a rectangular microchannel with constant wall temperature, *International Journal of Heat and Mass Transfer,* vol. 51, pp. 5088-5096, 2008.

[13] J. Fan, I.D. Boyd, C.P. Cai, K. Hennighausen and G.V. Candler, Computation of rarefied gas flows araound a NACA 0012 airfoil, *AIAA Journal,* vol. 39, pp. 618-625, 2001.

[14] A. Beskok and G.E. Karniadakis, A model for flows in channels, pipes, and ducts at micro and nano scales, *Microscale Thermophysical Engineering,* vol. 3, pp. 43-77, 1999.

[15] G.E. Karniadakis and A. Beskok, Micro flows: fundamentals and simulation, *Springer,*New York, 2002.

[16] C. Shen, Rarefied Gas Dynamics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[17] N.D. Stevanovic and V.D. Djordjevic, An exact analytical solution for the second order slip-corrected Reynolds lubrication equation, *FME Transactions,* vol. 43, pp. 16-20, 2015.

[18] W.L. Ferras, J.M. Nobregal and F.T. Pinho, Implementation of slip boundary conditions in the finite volume method: new techniques, *International Journal for Numerical Methods in Fluids,* vol. 72, pp. 724-747, 2013.

[19] A. Shoja-Seni, E. Roohi, M. Kahrom and S. Stefanov, Investigation of aerodynamic characteristics of rarefied flow around NACA0012 airfoil using DSMC and NS solvers, *European Journal of Mechanics B/Fluids,* vol. 48, pp. 59-74, 2014.

[20] G.J. LeBeau, A parallel implementation of the direct simulation Monte Carlo method, *Computer Methods and Applications in Mechanical Engineering,* vol. 174, no. 3, pp. 319337, 1999.

[21] M.A. Gallis, K.A. Boyles, G.J. LeBeau, DSMC simulations in support of the STS-107 accident investigation, *Proceedings of the 24th Rarefied Gas Dynamics (RGD) Symposium,* Bari, Italy, July 2004.

[22] M.I.G. Markelovf and S.F. Gimelshein, Statistical simulation of reactive rarefied flows: numerical approach and applications, *Proceedings of the 7th AIAA/ASME Joint Thermophysics and Heat Transfer Conference,* Albuquerque, June 1998.

[23] S. Dietrich and I.D. Boyd, Scalar and parallel optimized implementation of the direct simulation Monte Carlo method, *Journal of Computers and Physics,* vol. 126, no. 2, pp. 328342, 1996.

[24] T.J. Bartel, S. Plimpton, and M.A. Gallis, Icarus: A 2-D Direct Simulation Monte Carlo (DSMC) code for multi-processor computers, *Sandia National Laboratories Technical Report,* 2001-2901, Albuquerque.

[25] D. Gao and T.E. Schwartzentruber, Optimizations and OpenMP implementation for the Direct Simulation Monte Carlo method, *Journal of Computers and Fluids,* vol. 42, no. 1, pp. 7381, 2011.

[26] T.J. Scanlon, E. Roohi, C. White, M. Darbandi, and J.M. Reese, An open source, parallel DSMC code for rarefied gas flows in arbitrary geometries, *Journal of Computers and Fluids,* vol. 39, no. 10, pp. 20782089, 2010.

[27] M.A. Gallis, J.R. Torczynski, S.J. Plimpton, D.J. Rader, and T. Koehler, Direct Simulation Monte Carlo: The quest for speed, *Proceedings of the 29th Rarefied Gas Dynamics (RGD) Symposium,* Xi'an, China, July 2014.

[28] G.N. Lygidakis and I.K. Nikolos, Numerical analysis of flow over the NASA Common Research Model using the academic Computational Fluid Dynamics code Galatea, *ASME Journal of Fluids Engineering,* vol. 137, 2015.

[29] J. Allegre, M. Raffin and J.C. Lengrand, Experimental flow fields around NACA0012 airfoils located in subsonic and supersonic rarefied air streams, *M.O. Bristeau et al. (eds.), Numerical Simulation of Compressible Navier-Stokes Flows,* pp. 5968, 1985.

[30] B.E. Launder and D.B. Spalding, The numerical computation of turbulent flows, *Computer Methods in Applied Mechanics and Engineering,* vol. 3, pp. 269289, 1974.

[31] P.G. Saffman and D.C. Wilcox, Turbulence-model predictions for turbulent boundary layers, *AIAA Journal,* vol. 12, pp. 541546, 1974.

[32] F.R. Menter, Two-equation eddy viscosity turbulence models for engineering applications, *AIAA Journal,* vol. 32, pp. 1598-1605, 1994.

[33] P. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, *Journal of Computational Physics,* vol. 43, pp. 357-371, 1981.

[34] E.F. Toro, M. Spruce and W. Speares, Restoration of the contact surface in the HLL Riemann solver, *Shock Waves,* vol. 4, pp. 25-34, 1994.

[35] G.D. Van Albada, B. Van Leer and W.W. Roberts, A comparative study of computational methods in cosmic gas dynamics, *Astronomy and Astrophysics,* vol. 108, pp. 46-84, 1982.

[36] P.K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, *SIAM Journal on Numerical Analysis,* vol. 21, pp. 995-1011, 1984.

[37] T.J. Barth and D.C. Jespersen, The design and application of upwind schemes on unstructured meshes, *27th AIAA Aerospace Sciences Meeting and Exhibit,* Reno, USA, January 9-12, 1989, AIAA-89-0366.

[38] V. Venkatakrishan, Convergence to steady state of the Euler equations on unstructured grids with limiters, *Journal of Computational Physics,* vol. 118, pp. 120130, 1995.

[39] J.S. Park, S.H. Yoon and C. Kim, Multi-dimensional limiting process for hyperbolic conservation laws on unstructured grids, *Journal of Computational Physics,* vol. 229, pp. 788-812, 2010.

[40] J. Blazek, Computational Fluid Dynamics: Principles and Applications. *Elsevier Science,* Kidlington, 2001.

[41] G.N. Lygidakis, S.S. Sarakinos and I.K. Nikolos, Comparison of different agglomeration multigrid schemes for compressible and incompressible flow simulations, *Advances in Engineering Software,* in press, 2016.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

TOPOLOGY OPTIMIZATION FOR FLUID FLOW EMPLOYING LOCAL OPTIMALITY CRITERIA

P. Sarstedt, G. Kachel, J. Ettrich, and K. Bühler

Hochschule Offenburg
University of Applied Sciences
Badstraße 24, D-77652 Offenburg, Germany
e-mail: philip.sarstedt@hs-offenburg.de

Keywords: Topology optimization, Local optimality criterion, Fluid flow, Navier-Stokes

Abstract. *We herein present a topology design method based on local optimality criteria which has been implemented in an open source Navier-Stokes solver for turbulent flows. Our method aims for the fast generation of geometry proposals in the early conceptual phase. To the best of our knowledge, this is the first local criteria approach utilizing a wall function turbulence model in order to consider turbulent flows. In order to allow for the growth as well as the shrinkage, or even the formation or disappearance of structural features, a topological approach is chosen. By introducing a volume fraction parameter, we distinguish between fluid and solid properties in each control volume. The fluid-solid interface is represented by an immersed boundary method using a piecewise linear surface reconstruction.*

1 INTRODUCTION

The development of fluid flow components, reaching from simple pipe systems to complex geometries, is still a challenging task for engineers and scientists. Usually a compromise between different design goals has to be made. Here CFD simulations coupled with optimization strategies are a valuable tool, but today they still go along with high computational costs. In order to keep these costs low and allow for a fast generation of geometry proposals in the early conceptual phase, we present a method based on local optimality criteria.

The term "optimization" is defined as the determination of design variables that minimize (or maximize) a cost (or objective) function with respect to a state equation possibly subject to constraints. This task can be accomplished by different strategies, for example classical gradient-based methods or Evolutionary Algorithms (EA). The former are fast but lack robustness and show the tendency to fall into local optima. Here recent progress is made in the calculation of the response sensitivities by the adjoint method which outperforms traditional methods like finite differences and has gained currency in commercial CFD software. Evolutionary Algorithms on the other hand are more robust but also computationally more expensive [15]. An example for an EA is the Generic Algorithm (GA) which mimics biological evolution. A third strategy, which is the subject of this paper and sometimes referred to as "bionic optimization", is based on local optimality criteria. Strictly speaking this method performs no optimization in the former defined sense, as a global extremal solution is not ensured. As we aim for a fast generation of appropriate geometry proposals rather than the optimal geometry, this is not considered necessary.

Furthermore we have to distinguish between *shape* and *topology* optimization. Shape optimization considers the parameterized bounding surface of the geometry. Topology optimization on the other hand utilizes all computational elements in a defined design space. This allows for the appearance of new boundaries and results in a much wider spectrum of possible solutions.

Shape optimization has been subject to research in the field of CFD since several decades, mainly in the sector of aeronautical problems, e.g. [13]. Well established in structural mechanics, topology optimization on the other hand is relatively new to CFD. The transfer of topology optimization from structural mechanics to computational fluid dynamics began with the work of Borrvall and Petersson in 2003 [2]. In their paper they considered the minimization of pressure drop in Stokes flow. In order to account for topological change, they adopted a source term of inverse permeability in the Stokes equation. In 2004 Moos et al. [11] presented a bionic topology optimization approach based on a local "backflow" criterion under the axiom that a fluid flow finds the best path under given constraints by itself. The method used the physical information of the flow field to deposit so called "virtual sand" in cells identified by the criterion. This local approach led to an improvement of the duct geometry within a very short computational period of time. Othmer et al. [12] picked up the idea of a backflow criterion in order to compute sensitivities for a climate channel optimization. Recirculation was detected by comparison of the actual velocity vector in a control volume with the potential flow solution. They implemented the optimization as a one-shot method in the CFD solver, which means that the optimization takes place in every iteration. The method was tested in 2D and 3D and characterized as extremely fast as regions with recirculation were inhibited that would usually require additional computing time. Recent effort was done by Wang et al. [16] who implemented a topology optimization method based on local optimality criteria in a lattice Boltzmann (LBM) framework. Prior to our investigations a project was carried out at Hochschule Offenburg by Gottlieb [6] which focused on the topological optimization of turbulent flows around blunt bodies. A

method employing the backflow criterion for the recirculation region "behind" the body and a criterion of homogenization of gradients for the front part was developed. In contrast, our work focuses on internal flows which are viscous-dominated, whereas blunt bodies are dominated by pressure drag.

2 NUMERICAL METHODS

2.1 Governing equations

The governing equations express mass and momentum conservation. Steady viscous incompressible flows without volume forces can be described by the following set of equations, namely continuity and Navier-Stokes momentum equations

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$(\mathbf{u} \cdot \nabla)\mathbf{u} = \nu \Delta \mathbf{u} - \frac{1}{\rho} \nabla p. \tag{2}$$

For turbulent flows, according to the Boussinesq hypotheses, the kinematic viscosity ν composes of two terms, the molecular and the turbulent kinematic viscosity $\nu = \nu_l + \nu_t$. Herein, the turbulent viscosity is calculated by a turbulence model. In industrial applications two-equation models like the k-ϵ or k-ω-SST model are state of the art for Reynolds-averaged Navier-Stokes (RANS) simulations.

2.2 Local optimality criteria

The idea behind local optimality criteria is the utilization of heuristic correlations between local flow parameters and their global impact. In this way the calculation of sensitivities with respect to a global cost function is circumvented. In this section different optimality criteria are presented.

Following the idea of Moos et al. [11] we implement a backflow criterion. In order to avoid losses caused by recirculation, such regions are identified and set to solid. Therefore a comparison of the local velocity vector with a reference vector can be performed. If a specified angle of deviation, e.g. 90 degrees, is exceeded, backflow is present. The angle of deviation λ is calculated from the scalar product of the actual velocity vector $\mathbf{u}$ and a reference vector $\mathbf{u}_{ref}$ as follows:

$$\lambda = \frac{\mathbf{u}}{|\mathbf{u}|} \cdot \frac{\mathbf{u}_{ref}}{|\mathbf{u}_{ref}|}. \tag{3}$$

The reference vector can be obtained through different approaches. In a straight channel the direction from inlet to outlet is a sensible choice for $\mathbf{u}_{ref}$. In complex flow domains the potential flow field is a fast way to obtain a sensible reference vector field, as it contains no backflow regions by definition.

Wang et. al [16] proposed a different optimality criterion algorithm based on the natural behavior of erosion and deposition of sand in river basins. They introduced two local optimality criteria: At the fluid-solid interface, vanishing solid cells are chosen by maximum wall shear stress

$$\tau_w = \mu \left(\frac{\partial u}{\partial y} \right)_w, \tag{4}$$

with the dynamic viscosity μ. On the other hand solidifying fluid cells are chosen by lowest dynamic pressure

$$p_{dyn} = \frac{\rho}{2} |\mathbf{u}|^{\mathbf{2}}. \tag{5}$$

Based on the work of Bourot [3] we derive another criterion based on the homogenization of gradients. Bourot observed a correlation between vorticity and optimal body shapes in Stokes flow. A homogenization of the vorticity magnitude on the surface of the body led to optimal shapes with respect to flow loss. The vorticity is defined as the curl of the velocity vector (cf. Eq. (6)) and can be interpreted as a measure of the rate of rotation of a fluid element [9].

$$\omega = \nabla \times \mathbf{u}. \tag{6}$$

An often encountered issue in flow optimization is the relocation of loss inducing geometrical features to the proximity of the outlet. This results in a seemingly reduced loss inside the considered domain as the loss takes place behind the outlet. An example would be a sharp edge at the outlet of the design space resulting in a flow separation in the adjacent pipe. In order to avoid such a behavior, we introduce an adaption region (AR) for the inlet and outlet, cf. figure 1. Herein the adapted reference vectors are obtained by interpolation between the reference vector field $\mathbf{u}_{ref}$ and the boundary normal vector (e.g. $\mathbf{u}_{ref_{out}}$). In addition we limit the angle of deviation λ more strictly as we approach the boundary. Also a *fixed* fluid region in the vicinity of the inlet and outlet is introduced where no geometrical change is performed by the optimization algorithm. For this purpose the inverse permeability coefficient α is fixed to zero. Altogether this leads to an almost orthogonal outflow and the prevention of outlet blockage.

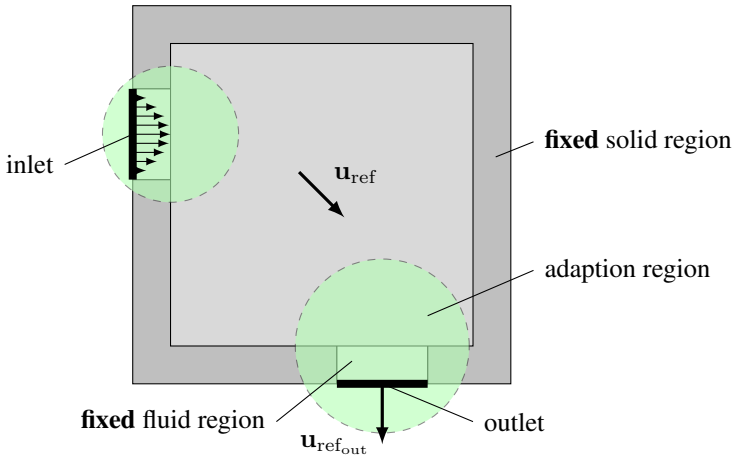


Figure 1: Adaption regions and fixed regions.

2.3 Optimization algorithm

As discussed earlier, we utilize a topology optimization with all control volumes serving as parameters. In order to distinguish between fluid and solid cells, we introduce the scalar volume fraction field Ψ, defining the ratio of the cell-volume occupied by solid to the total cell-volume:

$$\Psi = \frac{V_{solid}}{V_{total}}. \tag{7}$$

The volume fraction serves as the parameter for the optimization algorithm and is manipulated according to the method of steepest descent with a constant step size δ.

$$\Psi_{new} = \Psi_{old} - \delta\lambda. \tag{8}$$

Basically this is done in a “one-shot” fashion for every timestep [12].

3 TOPOLOGY REPRESENTATION

The representation of topological features of a time dependent flow domain is subject of this section. In the literature different approaches are found that can be classified into two categories: *Mesh morphing (MM)* and *immersed boundary methods (IBM)*. In mesh morphing a body fitted computational grid is distorted according to the displacement of the boundary. In immersed boundary methods the simulation is carried out on a Cartesian grid where the effect of the boundary is imposed on the flow through a source term in the governing equations. For an extensive overview of the different IBM methods we refer the reader to Mittal and Iaccarino [10]. As we aim for a fast optimization method a recomputation of the mesh for each iteration is too cost intensive. Therefore, following the work of Angot et al. [1] and Khadra et al. [8], we assume the design space as a porous medium and the flow to be governed by the Navier-Stokes/Brinkman equations, also called penalized Navier-Stokes equations [4] or penalization method (PM) [5]. For a steady viscous incompressible flow the equations read

$$(\mathbf{u} \cdot \nabla)\mathbf{u} = \nu \Delta \mathbf{u} - \alpha \mathbf{u} - \frac{1}{\rho} \nabla p, \tag{9}$$

with the inverse permeability coefficient α. If $\alpha = 0$, the Navier-Stokes equations are solved and the flow is a fully liquid phase. If α tends to infinity, the medium is solid and the equations tend to the Darcy equation. For numerical reasons α is set to $\alpha_{max} = 10^5$ in the solid domain. This approach can be seen as an immersed boundary method with the main advantage that an imposition of a boundary condition at the fluid-solid interface is not necessary. For laminar flows it appropriately mimics the effects of solid boundaries on the flow as will be shown in the subsequent section.

In order to account for the technically important fields of application, the topology representation approach has to be capable of considering highly turbulent flows. Here the boundary layer at walls is critical for the accuracy of the simulation as it contains large gradients of the flow quantities. In RANS simulations a very fine mesh to the wall is generally used in order to resolve these gradients sufficiently. This approach is known as low Reynolds number approach (Low-Re). It is not suitable for a fast topology optimization as the position of the walls is not known a priori and the mesh would have to be very fine everywhere or refined around the boundaries for every change in topology. In order to circumvent such an excessive grid resolution, a high Reynolds number turbulence model approach (High-Re), also called wall function approach, can be used to skip the boundary layer via a universally valid wall function.

In simulations with high Reynolds number flow the Navier-Stokes/Brinkman approach leads to large errors in the flow field prediction, caused by the step-wise representation of the fluid-solid interface and a missing treatment of the turbulent quantities. This affects the optimization and leads to unrealistic geometry proposals. Therefore a more advanced immersed boundary method is used to guarantee proper fluid-solid interface treatment. Here, the discrete forcing approach with direct imposition of boundary conditions is the most promising [10]. By modification of the discretized equations near the fluid-solid interface, the immersed boundary acts like a body fitted mesh. This leads to a higher accuracy which is critical for high Reynolds number flows. Such an IBM was recently implemented in the OpenFOAM Extend Project [7].

Currently we are working on the adaption of the immersed boundary method to handle moving and deforming boundaries. The idea is, that the optimization algorithm still manipulates the volume fraction field Ψ according to the local optimality criteria, where Ψ is used to extract the geometry through a linear surface reconstruction [14]. The resulting surface geometry data is then fed to the immersed boundary method. As this procedure is more cost intensive, it is

no longer coupled to the solver in a "one-shot" fashion but performed at a defined time step interval. A drawback in terms of topology optimization is the exclusion of the solid region from the solution. Once blocked out, the removal of solid cells is only possible at the fluid-solid interface. Here criteria like magnitude of shear stress or vorticity are applicable. Therefore the method can be classified partly as topology and partly as shape optimization.

4 NUMERICAL EXAMPLES

4.1 Plane channel

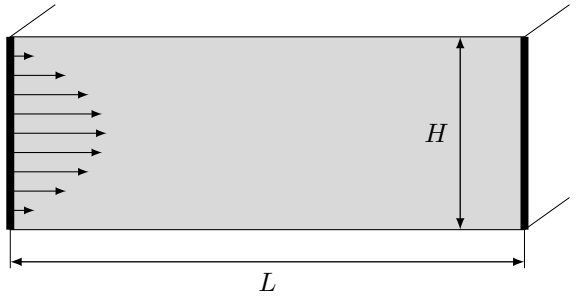
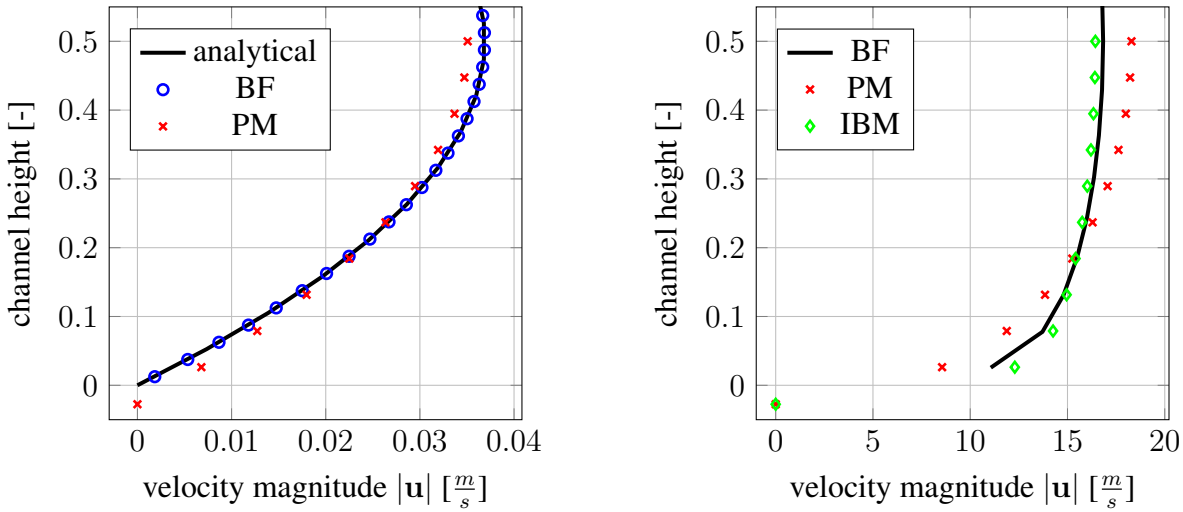


Figure 2: Dimensions of the "plane channel" simulation domain.

To show the validity of the penalization method as well as the immersed boundary method with discrete forcing approach and direct imposition of boundary conditions, we consider a 2D steady viscous incompressible channel flow. In order to achieve a fully developed flow, a periodic boundary condition is applied, which maps the flow quantities from the outlet to the inlet. For a laminar flow at $Re = 40$ the solution calculated with a body fitted mesh (BF) matches the analytical solution. In order to apply the penalization method, an additional layer of control volumes is added to the top and bottom of the flow domain. Here the inverse permeability coefficient α is set to 10^5, driving the velocity towards zero in the cell centers. This does not exactly enforce the no-slip boundary condition at the cell faces which leads to an artificially increased channel height and a lower maximum velocity, cf. figure 3a. For a turbulent flow at $Re = 25,000$ the penalization method shows an artificially increased boundary layer thickness, cf. figure 3b. This originates from the missing treatment of the turbulent quantities at the fluid-solid interface. The immersed boundary method however shows a much better agreement with the body fitted mesh solution, since the boundary condition is applied sharply at the fluid-solid interface and the turbulence is treated by a wall function approach through a high Reynolds number turbulence model. This study suggests the adoption of the immersed boundary method with discrete forcing approach and direct imposition of boundary conditions for the optimization of turbulent flows. The implementation is still under development at the moment and will be presented in subsequent publications.

4.2 Right angle elbow

In order to demonstrate the optimization method, a 2D rectangular simulation domain with a velocity inlet on the upper left boundary and a zero-pressure outlet on the lower right boundary is considered, cf. figure 4. At the inlet and outlet no-adaption zones are applied. The inner rectangular domain represents the so called design space and is bounded by a solid region, enforced by the penalization method with $\alpha = 10^5$. For the numerical calculations the design space is divided into 200×200 unit square cells, forming an equidistant Cartesian grid. The



(a) $Re = 40$ (b) $Re = 25,000$

Figure 3: Comparison of velocity profiles obtained with body fitted mesh (BF), penalization method (PM) and immersed boundary method (IBM) for laminar and turbulent flows.

fluid is assumed as air with a kinematic viscosity of $\nu = 1.56 \times 10^{-5} \frac{m^2}{s}$. Depending on the inlet length $L_{in} = 0.2L$ with $L = 0.05m$ the mean inlet velocity $\overline{u}_{in}$ is adjusted to match the desired Reynolds number $Re = \frac{L_{in}\overline{u}_{in}}{\nu}$. In order to assess the geometry proposed by the optimization methods, the total pressure loss $\Delta\tilde{p}_t$ is calculated from the mass flow averaged total pressures $\tilde{p}_t$ at inlet and outlet.

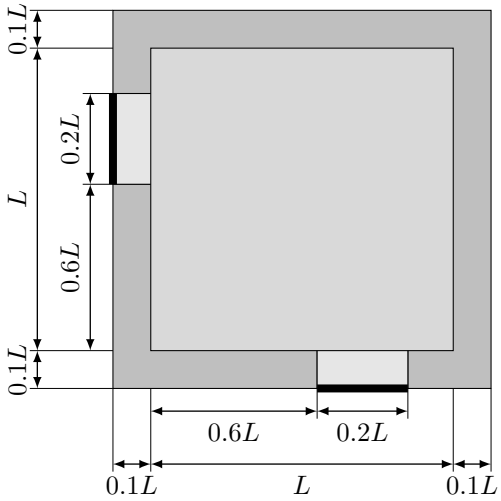
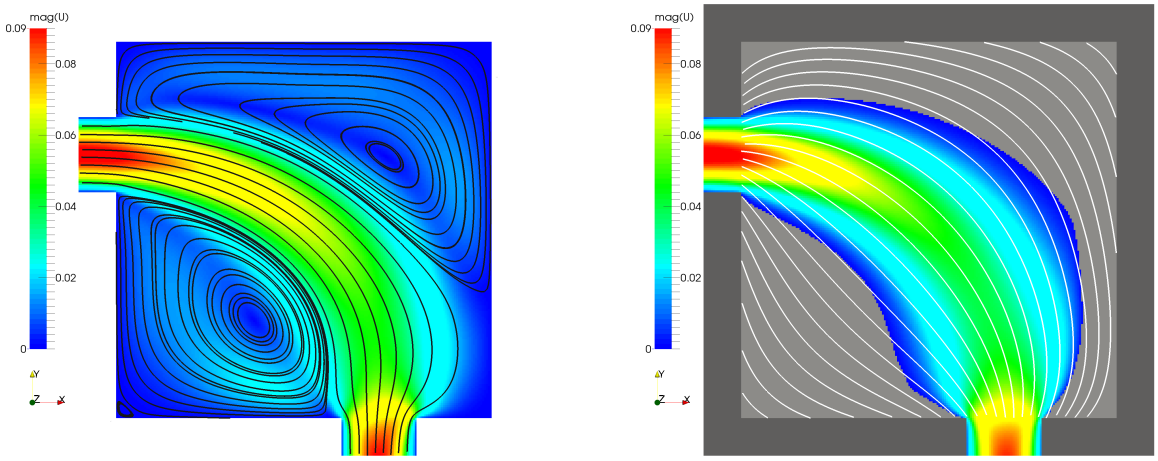


Figure 4: Dimensions of the "right angle elbow" simulation domain.

We consider the laminar case where the fully developed flow enters the domain at a Reynolds number of $Re = 40$. The steady-state solution of the flow shows two recirculation regions in the upper right and lower left corner of the design domain, cf. figure 5a. These regions are undesirable as they cause losses, therefore the aim is to generate a geometry that leads to no such regions. Figure 5b shows the solution of the optimization based on the backflow criterion. In this case the potential flow, represented as white colored streamlines, was used as reference vector field. As the simulation progresses the recirculation regions are blocked by the

penalization method, resulting in an improved total pressure loss of 93.34%.



(a) BF mesh, no optimization

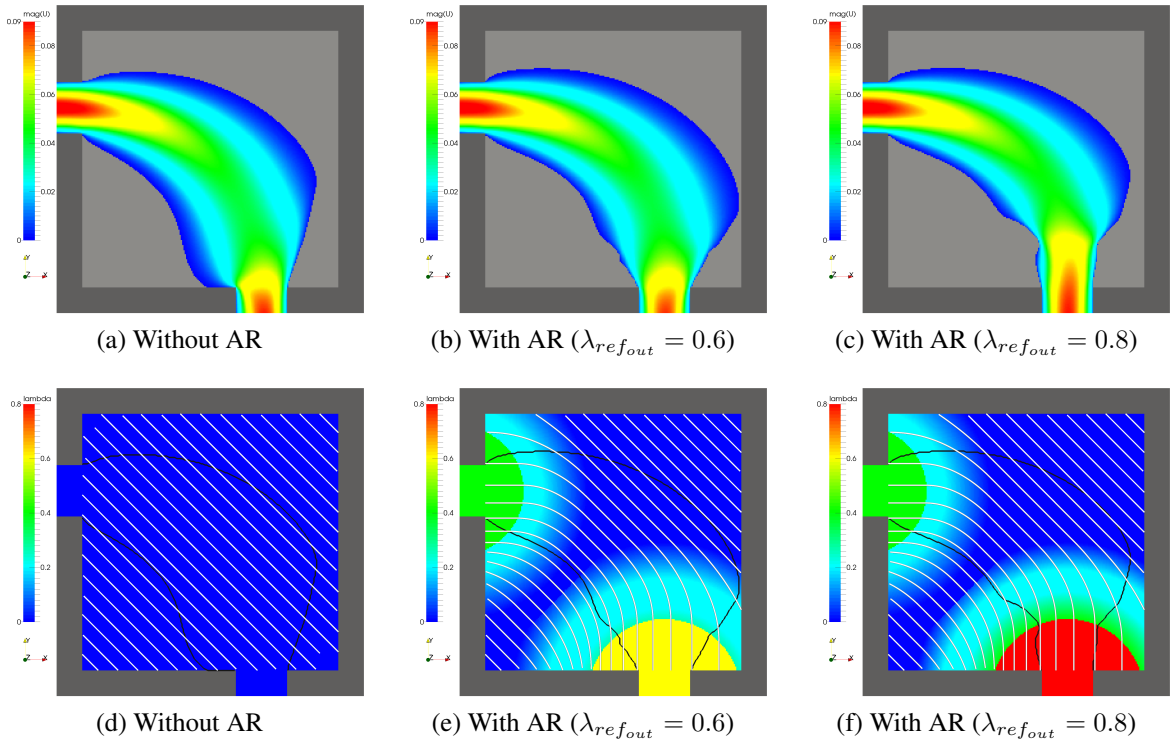
(b) PM, backflow optimization

Figure 5: Bionic optimization of a right angle elbow with the potential flow solution used as reference vector field for the backflow criterion ($Re = 40$).

In order to demonstrate the functionality of the adaption regions, an optimization case of the right angle elbow with a reference vector set to $\mathbf{u}_{ref} = (1, -1, 0)$ is considered, cf. streamlines in Fig. 6d. Figure 6 shows the obtained geometry proposals. The reference vector as well as the allowed angle of deviation λ is manipulated inside the adaption regions through interpolation. The aim is a geometry proposal that leads to an orthogonal outflow at the outlet. If this is not the case the proposed solution can not be considered as a proper one as it would lead to recirculation regions behind the outlet. Figure 7 shows the velocity components u and v at the outlet. The increased x-velocity u indicates that the geometry obtained without adaption regions has a higher probability of forming recirculation regions behind the outlet. The cases with adaption regions show a reduction of the x-velocity depending on the adaption region parameters.

Besides velocity vectors, other flow quantities can be considered as local optimization criteria. The shear stress as well as the vorticity criterion are examples for criteria that are applicable at the fluid-solid interface. Here the optimization algorithm checks for the highest absolute values of the respective quantity and gradually converts the near solid cells to fluid cells, leading to a homogenization of gradients. To demonstrate this functionality, a sharp 90 degree bend is used as initial geometry, cf. figure 8a. Figure 8 shows two geometry proposals obtained by the shear stress criterion and the vorticity criterion. The sharp corners of the initial solution were smoothed and the channel width has increased, leading to a reduced total pressure loss of 51.21% for the shear stress criterion and 37.68% for the vorticity criterion, cf. figure 9.

An advantage of the penalization method is its seamless extension to 3D cases without additional modification. Figure 10 shows examples of 3D flow optimization with the backflow criterion. The first case is a flow around a circular disc inside a rectangular channel at a Reynolds number of $Re = 800$. Through the deposition of solid cells in front of as well as behind the circular disc, flow separation is avoided, leading to a more aerodynamic shape. The second case shows a turbulent flow at $Re = 20,000$ through a rectangular channel with a cavity. Although we know that the penalization method lacks the correct treatment of the turbulence quantities at the fluid-solid interface, the solution can be considered as an improvement as the geometry is



(a) Without AR (b) With AR ($\lambda_{ref_{out}} = 0.6$) (c) With AR ($\lambda_{ref_{out}} = 0.8$)

(d) Without AR (e) With AR ($\lambda_{ref_{out}} = 0.6$) (f) With AR ($\lambda_{ref_{out}} = 0.8$)

Figure 6: Velocity field of three different bionic optimizations of a right angle elbow at $Re = 40$ without (a) and with adaption regions (b, c). Streamlines of the reference vector field $\mathbf{u}_{ref}$ and the angle of deviation λ (d, e, f).

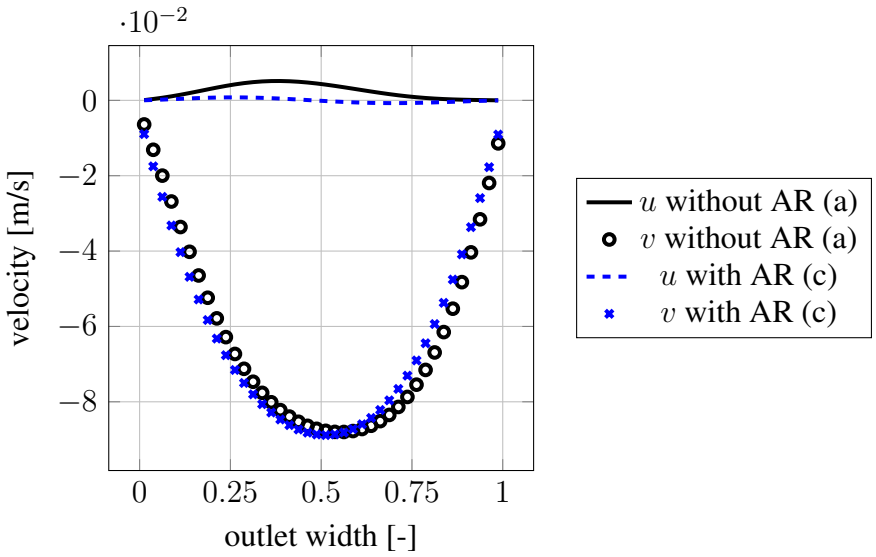
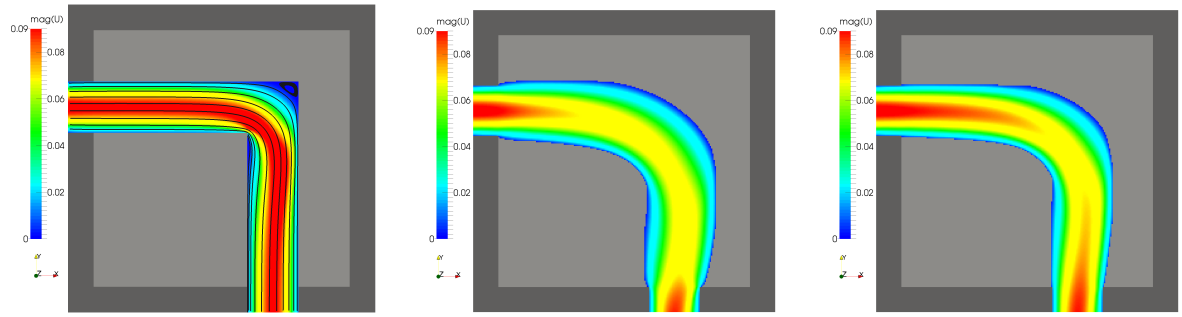


Figure 7: Velocity component u and v at the outlet of case 6a without and case 6c with adaption regions.

smoothed, reducing the likeliness of flow separation.



(a) No optimization (b) Shear stress crit. (c) Vorticity crit.

Figure 8: Bionic optimization of a sharp 90 degree bend ($Re = 40$).

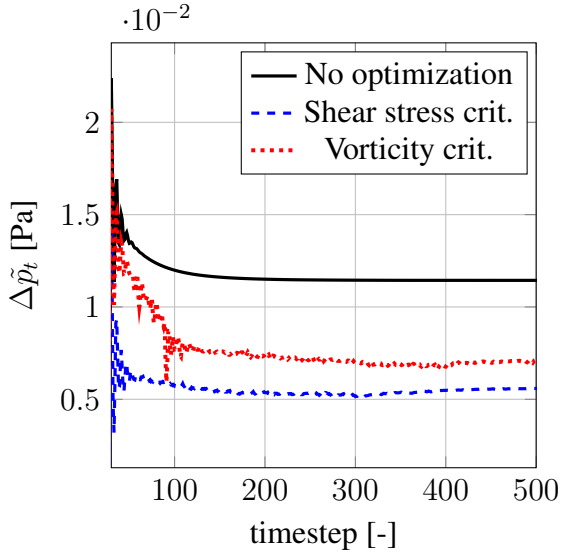
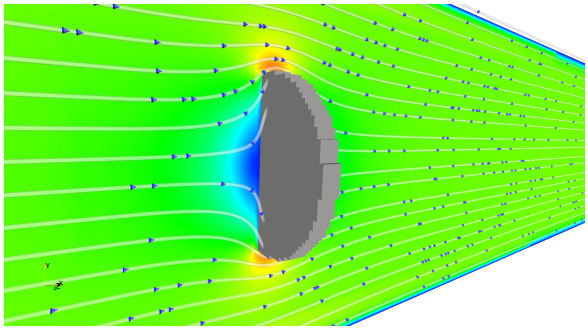


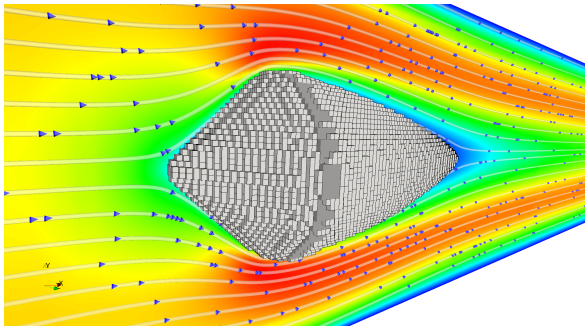
Figure 9: Mass flow averaged total pressure loss over timesteps for a sharp 90 degree bend ($Re = 40$).

5 DISCUSSION AND OUTLOOK

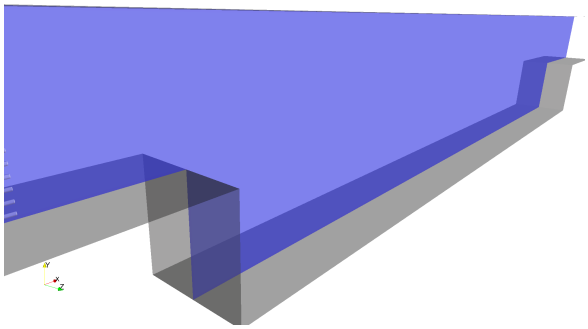
A design method that aims for the fast generation of appropriate geometry proposals in the early conceptual phase based on local optimality criteria was presented for the Finite-Volume Navier-Stokes CFD software OpenFOAM. Several criteria based on backflow, shear stress and vorticity magnitude, were implemented and investigated. Furthermore a method working with adaption regions was proposed in order to avoid recirculation regions behind the outlet. It was shown that hereby the velocity component perpendicular to the outlet normal vector can be reduced in the outlet plane. In order to realize a fast way of adapting the boundary during the simulation, a penalization method was adopted and its applicability successfully shown for 2D and 3D cases. Numerical examples were presented for a 2D right angle elbow at Reynolds number $Re = 40$, a circular disk in a 3D channel at $Re = 800$ and a 3D channel with cavity at $Re = 20,000$. For turbulent flows an immersed boundary method with discrete forcing approach and direct imposition of boundary conditions was adopted in order to consider correctly the turbulent quantities at the boundary. A 2D channel flow at $Re = 25,000$ calculated with this immersed boundary method shows a good agreement with the body fitted mesh solution. The integration of the immersed boundary method into the optimization algorithm is still under progress and results will be shown in subsequent publications. Future efforts will focus on



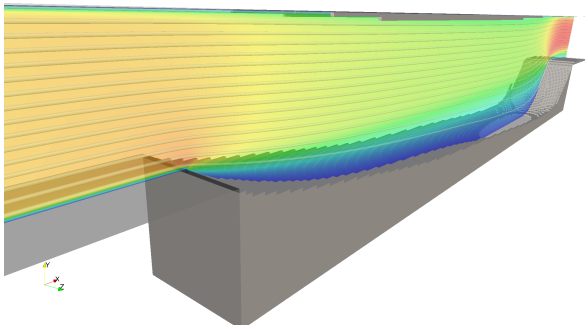
(a) Circular disk in a 3D channel



(b) Optimized circular disc ($Re = 800$)



(c) 3D channel with cavity



(d) Optimized cavity ($Re = 20,000$)

Figure 10: Examples of 3D bionic optimization.

the final implementation of the immersed boundary method to handle moving and deforming boundaries as this is vital for the correct treatment of turbulent flows. Furthermore the development of the optimization method with the main focus on the definition and evaluation of further optimality criteria is planned.

ACKNOWLEDGEMENT

The reported work was carried out as part of the FHprofUnt founding project, project number FKZ 03FH037PA2 (FoEn). The authors greatly acknowledge the financial support provided by the Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] P. Angot, C.H. Bruneau, P. Fabrie, A penalization method to take into account obstacles in incompressible viscous flows. *Numerische Mathematik*, **81**, 497–520, 1999.

[2] T. Borrvall, J. Petersson, Topology optimization of fluids in Stokes flow. *International Journal for Numerical Methods in Fluids*, **41**, 77–107, 2003.

[3] J.M. Bourot, On the numerical computation of the optimum profile in Stokes flow. *Journal of Fluid Mechanics*, **65**, 513–515, 1974.

[4] H.C. Brinkman, A calculation of the viscous force exerted by a flowing fluid on a dense swarm of particles. *Applied Scientific Research*, **1**, 27–34, 1947.

[5] C.H. Bruneau, I. Mortazavi, Numerical modelling and passive flow control using porous media. *Computers & Fluids*, **37**, 488–498, 2008.

[6] A. Gottlieb, Vorauslegung umströmter Bauteile durch Anwendung lokaler Optimalitätskriterien. PhD thesis, Universität Stuttgart, 2013.

[7] H. Jasak, D. Rigler, Ž. Tuković, Finite volume immersed boundary method for turbulent flow simulations. *9th OpenFOAM Workshop*, Zagreb, 2014.

[8] K. Khadra, P. Angot, S. Parneix, J.P. Caltagirone, Fictitious domain approach for numerical modelling of Navier-Stokes equations. *International Journal for Numerical Methods in Fluids*, **34**, 651–684, 2000.

[9] H.J. Lugt, *Introduction to Vortex Theory*, Vortex Flow Press, Potomac, MD, 1996.

[10] R. Mittal, G. Iaccarino, Immersed Boundary Methods. *Annual Review of Fluid Mechanics*, **37**, 239–261, 2005.

[11] O. Moos, F.R. Klimetzek, R. Rossmann, Bionic Optimization of Air-Guiding Systems. *SAE Technical Paper 2004-01-1377*, 2004.

[12] C. Othmer, T. Kaminski, R. Giering, Computation of topological sensitivities in fluid dynamics: Cost function versatility. *ECCOMAS CFD*, Delft, 2006.

[13] O. Pironneau, On Optimal Design in Stokes Flow. *Journal of Fluid Mechanics*, **59**, 117–128, 1973.

[14] R. Scardovelli, S. Zaleski, Analytical Relations Connecting Linear Interfaces and Volume Fractions in Rectangular Grids. *Journal of Computational Physics*, **164**, 228–237, 2000.

[15] D. Thévenin, G. Janiga, *Optimization and Computational Fluid Dynamics*, Springer, Berlin, Heidelberg, 2008.

[16] L. Wang, Y. Fan, L. Luo, Heuristic optimality criterion algorithm for shape design of fluid flow. *Journal of Computational Physics*, **229**, 8031–8044, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL SIMULATION OF THE CONJUGATE HEAT TRANSFER IN THE COOLING SYSTEM OF THE COMBUSTION CHAMBERS OF THE AVIATION RAMJET ON THE ENDOTHERMIC FUELS

V.M. Volokhov, P.D. Toktaliev and S.I. Martynenko

The Institute of Problems of Chemical Physics of the Russian Academy of Sciences
Academician Semenov avenue 1, Chernogolovka, Moscow region, 142432 Russian Federation
e-mail: vvm@icp.ac.ru; Toktaliev@icp.ac.ru; Martynenko@icp.ac.ru

Keywords: Endothermic Fuel, Numerical Simulation, Aviation Ramjet Engines, Thermal Destruction.

Abstract. *The paper represents 3D mathematical model of conjugated heat exchange in the cooling systems of the aviation ramjet engines. The systems use cooling capacity of the turbulent endothermic fuel flows in heated curvilinear rough channels surrounding the combustion chamber. Simplest one-step model based on substitution of the real hydrocarbon substances by some fictitious medium is used for description of thermal decomposition of the endothermic fuels. The one-step pyrolytic reaction mechanism describes the fictitious medium decomposition without intermediate reactions. In addition to the standard model based on the Navier-Stokes equations, the mathematical model includes a partial differential equation for computation of a local decomposition degree of the fictitious medium in curvilinear channels to avoid computation of residence time. Energy equation has an extra source term describing the endothermic effect of the hydrocarbon substance pyrolysis. The paper represents results of the numerical experiments with pentane under supercritical pressure as the endothermic fuel.*

1 INTRODUCTION

Cooling of an aircraft and a combustion chamber of its engine is the most important problem for flight Mach number higher than 6 [1]. Now the main trend is to use the cooling capacity of the engine fuel for thermal protection of the combustion chamber. Prior to injection and burning, the hydrocarbon fuel flows through the cooling system surrounding the combustion chamber. The fuel consumption is relatively small because the engine thrust must compensate the aerodynamic friction losses on high flight altitudes. As a result, minimum fuel can be used for thermal protection of the combustion chamber, average specific heat flux from the combustion products to the chamber wall is ~ 1 MW/m^2 and higher.

Some natural hydrocarbons (ethane C_2H_6, propane C_3H_8, butane C_4H_{10} and pentane C_5H_{12}) can be decomposed into mixture of simpler hydrocarbons with low molecular weights at high temperature. The thermal decomposition is an endothermic process, i.e. it takes place with heat absorption. Hydrocarbons absorbed heat at the decomposition and used as a fuel for high speed aircrafts are called as "endothermic fuels" (EF) [2, 3]. Attractive property of EF is an opportunity to use an additional cooling effect [4]. This fact results in intensive experimental and numerical studies of heat transfer in EF with endothermic pyrolytic reactions under supercritical pressure.

The simplest model of EF thermal decomposition is the most preferable for the engineering applications. Instead of real EF, the model uses some fictitious substance that decomposes without any intermediate reactions (so-called one-step pyrolytic mechanism) [5, 6]. Empirical constants and functions described the fictitious substance decomposition are chosen for obtaining good agreement between computed and experimental data for real EF [7, 8]. Advantage of the simplest model is the least computational efforts as compared with more complicated models. It should be emphasized that the combustion chamber and its cooling system should be computed in the coupled manner. So numerical simulation of aviation engines require not only accurate mathematical models, but also very efficient solvers for minimizing of the computational work.

Cooling system of the combustion chamber walls is shown in Fig. 1. EF comes to the cooling system and absorbs heat through enthalpy variation. Once the EF temperature exceeds a threshold level, the pyrolytic process can enhance the heat-absorbing capacity through chemical energy changes of EF. Note that reactivity of decomposition products (DP) is higher than reactivity of EF. This fact makes it possible to improve organization of the combustion process in the chamber. After that DP comes to the combustion chamber through pylons.

The aim of this work is: 1) development of 3D mathematical model of turbulent flows of thermal decomposed EF under supercritical pressure in curved rough channels of the cooling system; 2) numerical study of the conjugate heat transfer in panel of the cooling system.

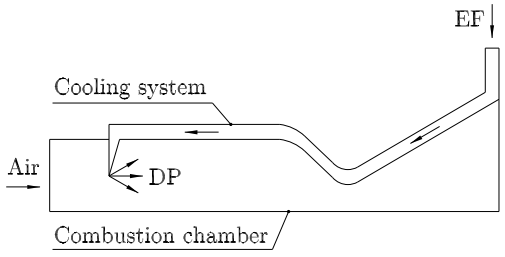


Figure 1: Cooling system of the combustion chamber walls of the aviation ramjet engines.

2 MATHEMATICAL MODEL OF ENDOTHERMIC FUEL DECOMPOSITION

The simplest mathematical model of conjugated heat exchange in the cooling systems is based on the following assumption: EF is replaced by a fictitious medium (FM), thermal properties of FM (density, viscosity, thermal conductivity and heat capacity) depend on the pressure, temperature and function Ψ, hereinafter referred as a local decomposition degree. In addition to the equations of continuity, momentum, turbulence, energy (for FM flow) and thermal conductivity (for channel), the mathematical model should include an equation for the local decomposition degree Ψ. Energy equation should include a source term accounting the endothermic effect of the EF decomposition.

2.1 Computation of the decomposition degree

The decomposition degree is a quantitative measure of the EF conversion into mixture of the simpler hydrocarbon compounds. Assume that a vessel is filled by of EF and N_0 is the number of EF molecules in the initial time τ_0. The equation described reduction of the number of EF molecules at the thermal decomposition in the vessel is

$$\frac{\partial N}{\partial \tau} = -fN, \tag{1}$$

where τ is a residence time and $f \geqslant 0$ is some nonnegative function. Eq. (1) reminds the main law of radioactive decay, however here the function f has a more complicated form. Integration of Eq. (1) gives

$$\ln \frac{N}{N_0} = -\int\limits_{\tau_0}^{\tau} f\, dt.$$

Definition. The function

$$\Psi = 1 - \frac{N}{N_0} = 1 - \exp\left(-\int\limits_{\tau_0}^{\tau} f\, dt\right) \tag{2}$$

is called a local decomposition degree of EF (LDDEF). The function $0 \leqslant \Psi < 1$ defines a local activity of the EF decomposition.

A mathematical model of a cooling system of a combustion chamber wall of an aircraft engine requires an equation for computation of LDDEF. In order to obtain this equation, we divide both sides of Eq. (1) on N_0

$$\frac{\partial}{\partial \tau}\left(\frac{N}{N_0}\right) = -f\frac{N}{N_0}. \tag{3}$$

Eq. (2) gives

$$\frac{N}{N_0} = 1 - \Psi,$$

and Eq. (3) can be rewritten as

$$\frac{\partial \Psi}{\partial \tau} = (1 - \Psi)f, \tag{4}$$

where $\Psi = \Psi_0$ at $\tau = \tau_0$ (i.e. in the reference time τ_0).

For purpose of convenience, we introduce an auxiliary function $\hat{\Psi} = \ln(1 - \Psi)$. There are different approaches for approximation of the right hand side function f of Eq. (4). The most popular approximation takes the Arrhenius form

$$f = A \exp\left(-\frac{E}{RT}\right), \tag{5}$$

where A (s^{-1}) and E (J/mole) are some empirical constants (which are different for different EF), $R = 8.318$ J/(kmol·K) is absolute gas constant and T is absolute temperature (K).

Using the approximation (5) and Eq. (4), the auxiliary function $\hat{\Psi}$ coupled with LDDEF as $\hat{\Psi} = \ln(1 - \Psi)$ can be computed by solving the problem:

$$\frac{\partial \hat{\Psi}}{\partial \tau} = -A \exp\left(-\frac{E}{RT}\right), \qquad \hat{\Psi}(\tau_0) = \hat{\Psi}_0. \tag{6}$$

Now we consider an adaptation of the problem (6) for computation of LDDEF in curvilinear channels of the cooling system. At first, we consider the EF flow between the parallel heated plates. Remember that τ in (6) is the residence time and it coincides with the physical time t only for motionless EF. However for EF flow between plates, the residence time τ depends on local speed and passed distance, i.e. it is possible to assume that

$$u = \frac{dx}{dt},$$

where u is a velocity component of EF flow in the axial direction (along plates). Accounting $dt = dx/u$, Eq. (6) can be rewritten as

$$\frac{\partial \hat{\Psi}}{\partial x} = -\frac{A}{u} \exp\left(-\frac{E}{RT}\right). \qquad \hat{\Psi}(\tau_0) = \hat{\Psi}_0. \tag{7}$$

Eq. (7) predicts that the most intensive thermal decomposition of EF takes place at high temperature and/or slow fluid velocity. Therefore application of the rectilinear channels in the cooling systems seems to be problematic because of the EF decomposition will take place mainly in the near wall region. It is clear that full usage of the endothermic cooling capacity can be possible only at uniform decomposition in the channel cross-sections.

Obtained Eq. (7) has the following disadvantages:
1) the turbulent flows are characterized by random fluctuations of velocity, pressure and temperature. As a result, the residence time computed by the Reynolds averaged values will be less than the time computed by instantaneous values. It is expected that the constants A and E in Eq. (7) should be functions of the Reynolds number;
2) Eq. (7) gives inaccurate prediction of LDDEF in near wall region. Really the non-slip conditions lead to

$$u \to 0 \Rightarrow \hat{\Psi} \to -\infty \Rightarrow \Psi \to 1,$$

i.e. full decomposition of EF on the heated walls is expected.

Computation of the residence time of the reacting medium in different technical devices is sufficiently difficult problem [9, 10, 11, 12]. The modern methods of the residence time computation require impressive computational works. Obtained results can be inaccurate for the engineer applications.

Let us rewrite Eq. (7) as

$$\rho u \frac{\partial \hat{\Psi}}{\partial x} = -\rho A \exp\left(-\frac{E}{RT}\right),$$

where ρ is FM density. It is easy to see that this equation is similar to convective "part" of some transport equation written in nonconservative form. It is expected that the equation for the function $\hat{\Psi}$ can be written as

$$\frac{\partial(\rho\hat{\Psi})}{\partial t} + \nabla\left(\rho\vec{V}\hat{\Psi}\right) = -\rho A \exp\left(-\frac{E}{RT}\right),$$

where t is physical time, $\vec{V}$ is a velocity vector and ∇ is the Hamilton operator. It is expected that this equation can be rewritten in the following general form

$$\frac{\partial(\rho\hat{\Psi})}{\partial t} + \nabla\left(\rho\vec{V}\hat{\Psi}\right) = \nabla\left(D_{\hat{\Psi}}\nabla\hat{\Psi}\right) - \rho f, \tag{8}$$

where the function f is taken from Eq. (1) and $D_{\hat{\Psi}}$ is a "diffusion" coefficient.

Boundary conditions for Eq. (8) are given by:

a) inlet (inlet Ψ is given)

$$\hat{\Psi}\Big|_{\text{inlet}} = \ln\left(1 - \Psi\Big|_{\text{inlet}}\right),$$

b) outlet

$$\frac{\partial\hat{\Psi}}{\partial n}\Big|_{\text{outlet}} = 0,$$

where n is outlet normal.

c) channel wall ($\tau = t$)

$$\frac{\partial\hat{\Psi}}{\partial t}\Big|_{\text{wall}} = -A\exp\left(-\frac{E}{RT_w}\right), \tag{9}$$

where T_w is the wall temperature.

Thermal decomposition of EF is non-stationary process: on each time layer first computed $\hat{\Psi}_w$ on the heated walls of the channel (9) and after $\hat{\Psi}$ is computed in the domain (8). Verification of the model requires large computationat efforts and now the simplified model based on assumption that $D_{\hat{\Psi}} = 0$ will be used for EF simulation. The simplified model has a stationary solution because of Eq. (8) does not use the boundary conditions on the heated walls in case $D_{\hat{\Psi}} = 0$.

Note that Eq. (4) makes it possible to predict the nonfailure operation time of the cooling system. Integration of Eq. (4) gives

$$\frac{1-\Psi(\tau)}{1-\Psi(\tau_0)} = \exp\left(-\int_{\tau_0}^{\tau} f(t,T)\,dt\right).$$

Assuming that $\Psi(0) = 0$ and $f(t,T) = const$, we have

$$1 - \Psi(\tau) = \exp\left(-\tau f(T)\right).$$

Let is $\Psi_{\max}(\tau_{\max})$ maximum permissible value of the LDDEF. If $1 \geqslant \Psi \geqslant \Psi_{\max}$ then formation of solid deposits on the heated wall can be expected. This equation

$$1 - \Psi_{\max} = \exp\left(-\tau_{\max} f(T_{\max})\right)$$

results in the estimation of the nonfailure operation time $\tau_{\max}$

$$\tau_{\max} = -\frac{\ln(1-\Psi_{\max})}{f(T_{\max})}.$$

If $f(T)$ is taken as (5), finally we have

$$\tau_{\max} = -\frac{\ln(1-\Psi_{\max})}{A\exp\left(-\dfrac{E}{RT_{\max}}\right)}.$$

2.2 Source term in energy equation

The energy equation accounting the endothermic reaction is written as

$$\frac{\partial(\rho i)}{\partial t} + \nabla\big(\rho\vec{V}i\big) = \nabla\big(\lambda\nabla T\big) - S, \tag{10}$$

where i is FM enthalpy. If the function f is taken from Eq. (1), the source term S depends on FM decomposition degree of and temperature T as follows

$$S = AB\rho\exp\left(\hat{\Psi} - \frac{E}{RT}\right),$$

where A and B (J/m^3) are some empirical coefficients depending on EF. The source term $S > 0$ accounts the extra cooling capacity arising at thermal decomposition of EF.

2.3 State equation of the fictitious medium

Thermophysical properties of EF (density ρ_{EF}, dynamic viscosity μ_{EF}, heat conductivity λ_{EF} and heat capacity $c_{p\,\text{EF}}$) depend on the temperature and pressure, i.e. $\rho_{\text{EF}}(P,T)$, $\mu_{\text{EF}}(P,T)$, $\lambda_{\text{EF}}(P,T)$, $c_{p\,\text{EF}}(P,T)$. Then state equation of FM can be formulated by interpolation, for example

$$\rho_{\text{FM}}(P,T,\hat{\Psi}) = \exp(\hat{\Psi})\rho_{\text{EF}}(P,T) + \big(1-\exp(\hat{\Psi})\big)\rho_{\text{DP}}(P,T,\hat{\Psi}), \tag{11}$$

where $\rho_{\text{DP}}(P,T,\hat{\Psi})$ is density of the decomposition products (DP).

2.4 Mathematical model of the conjugate heat transfer

The 3D mathematical model of the cooling system consists of the following equations:
1) for FM flow:
   a) continuity and momentum equations (Navier-Stokes equations),
   b) equations for ($k-\varepsilon$)–model of turbulence,
   c) energe equation with extra source term (10),
   d) state equation (in form of (11)),
   e) equation for computation of the decomposition degree (8) .
2) for the cooling system construction:
   a) heat conductivity equation.

The model coincides with standard model of the conjugate heat transfer at low temperatures ($T < 600$ K $\Rightarrow \Psi \to 0$).

3 NUMERICAL EXPERIMENTS

A panel of the cooling system for aviation ramjet engine is a rectangle plate manufactured from a high-temperature steel with an internal channel for EF supply. Fragment with Π-like channel will be named as a heat-exchange section. Fig. 2 represents the overall sizes of the typical heat-exchange section with the channel for EF supply. The cooling panel consisting of three heat-exchange sections is shown on Fig. 3.

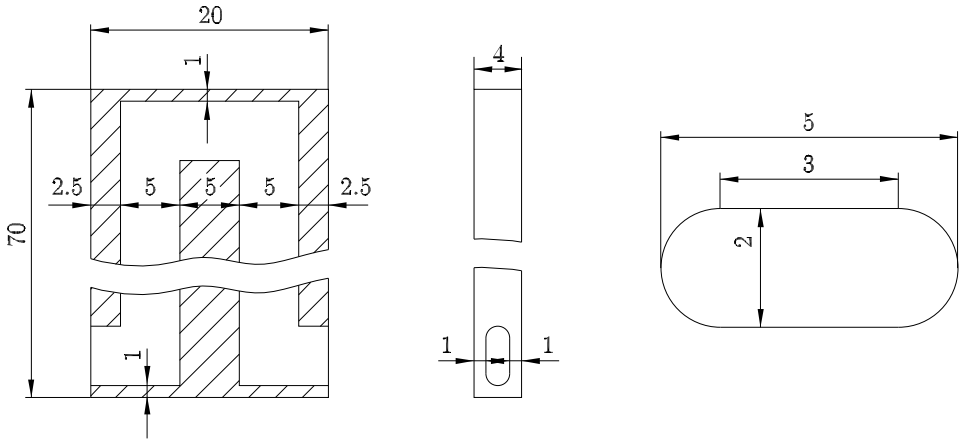


Figure 2: Geometry of the heat-exchange section with the internal channel for EF supply (all sizes are given in millimeters).

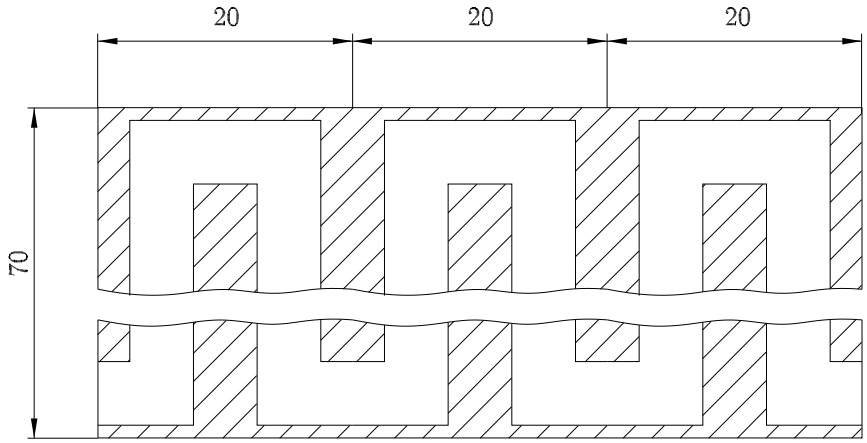


Figure 3: The cooling panel consisting of three heat-exchange sections (all sizes are given in millimeters).

Heat-insulated manifold equips the last heat-exchange section as shown on Fig. 4. The manifold is needed for the correct formulation of the outlet boundary conditions for momentum, energy, $(k-\varepsilon)$–equations and Eq. (9). It is assumed that zero Neumann conditions are given on the manifold outlet.

Unstructured grid in the panel (633047 control volumes in each section) and structured grid in the channel (191862 control volumes in each section) are generated for approximation of the basic partial differential equations (Fig. 5).

It is well known that roughness has strong influence on convective heat transfer in the channels (especially for small equivalent diameters). In our numerical experiments the uniform microroughness height is taken to be $4 \cdot 10^{-5}$ m.

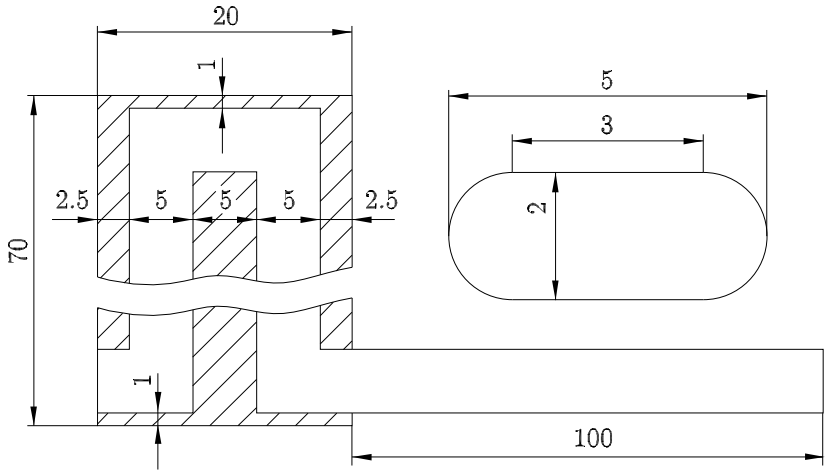


Figure 4: Last heat-exchange section and the heat-insulated manifold.

Pentane (C_5H_{12}) under supercritical pressure $p_0 = 5$ MPa is taken as the EF. Values of the empiric constants for the pentane are $E = 231268$ J/mole, $A = 1.75 \cdot 10^{12}$ c^{-1}, $B = 1.68 \cdot 10^6$ J/m^3. Inlet temperature of the pentane is $T_0 = 300$ K, mass flow rate is 0.0067 kg/s.

To analyze results of the computations, we use the following parameters:
1) maximum temperature ($T_{\max}$) and maximal LDDEF ($\Psi_{\max}$)

$$T_{\max} = \max_{\Omega_{ijk}} T_{ijk}, \qquad \Psi_{\max} = \max_{\bar{\Omega}_{ijk}} \Psi_{ijk}, \tag{12}$$

where Ω_{ijk} and $\bar{\Omega}_{ijk}$ are used computational grids in metal part of the panel and channel, respectively.
2) mass-averaged flow temperature $\langle T_f \rangle$ and mass-averaged LDDEF $\langle \Psi \rangle$

$$\langle T_f \rangle = \frac{\int\limits_s \rho u T_f \, d\sigma}{\int\limits_s \rho u \, d\sigma}, \qquad \langle \Psi \rangle = \frac{\int\limits_s \rho u \Psi \, d\sigma}{\int\limits_s \rho u \, d\sigma}, \tag{13}$$

where s is an area of the channel cross-section and u is velocity component.
3) mass-averaged LDDEF $\langle \bar{\Psi} \rangle$ in the manifold outlet (Fig. 4)

$$\langle \bar{\Psi} \rangle = \frac{\int\limits_m \rho u \Psi \, d\sigma}{\int\limits_m \rho u \, d\sigma}, \tag{14}$$

where m is an area of the outlet cross-section of the manifold.

Goal of the numerical experiments is to computational study the cooling panel from point of view of intensification heat transfer, minimization of hydraulic resistance and maximum usage of the endothermic effect at EF decomposition. Some EFs have tendency to coke deposit

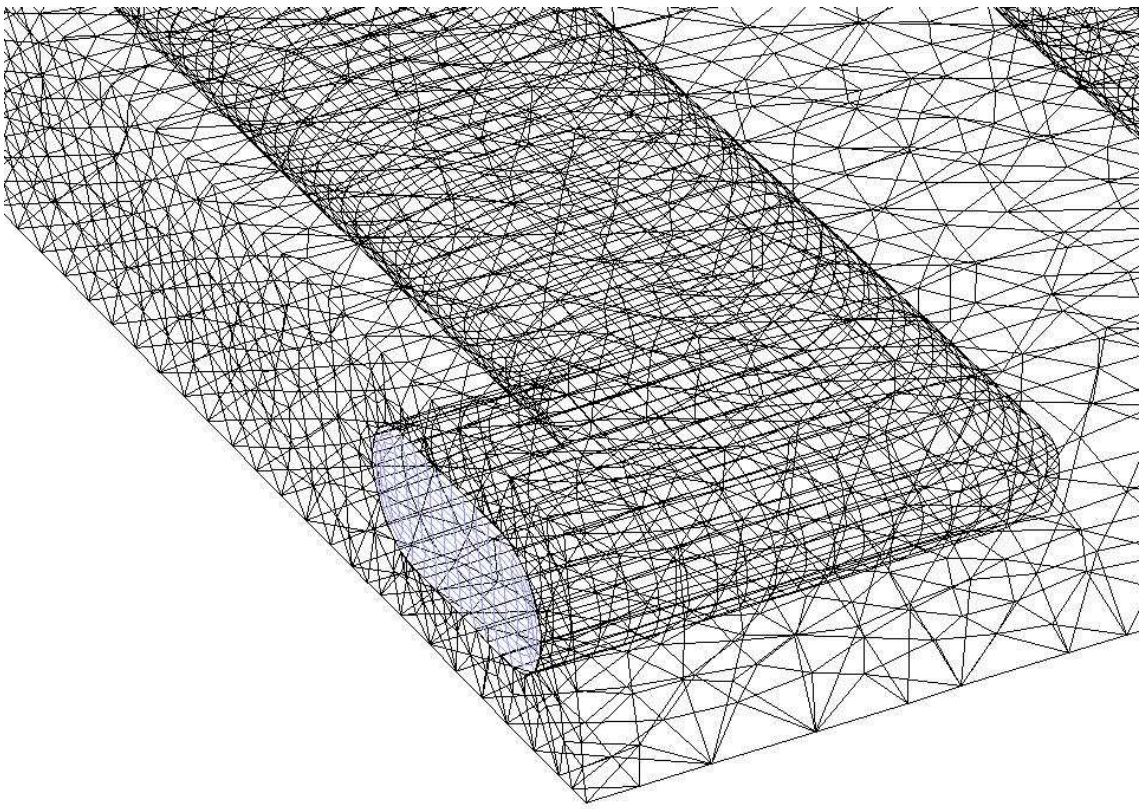


Figure 5: Computational grid in the cooling panel.

formation on overheated channel walls. Layer of the solid coke deposits results in additional heat resistance and reduces the cooling system efficiency. To overcome this problem, operating conditions of the cooling system should guarantee that maximum LDDEF is limited by some value, for example $\Psi_{\max} < 0.8$.

We summarize the main features of the cooling system:
a) Near inlet region. Heat of cold EF leads to remarkable change of their thermal properties (density, viscosity, heat conductivity etc.). If the channel cross-sectional area is the same, reduction of EF density results in increase of hydraulic resistance. The channel configuration should ensure intensification of convective heat transfer without unpractical hydraulic losses.
b) Near outlet region. EF is overheated up to their thermal decomposition. The decomposition starts in the channel subdomains with high temperature and high residence time (or small linear velocity), i.e. it is expected that the decomposition will take place in small overheated sublayer in near wall region. As a result, all endothermic cooling capacity of EF cannot be used in such system because of coke deposit hazard near walls and cold flow core. In addition, the channel configuration should ensure sufficient residence time for required decomposition value.

It is expected that the channel configuration shown on Fig. 3 results in continuous mixing of EF flow. In near inlet region the mixing leads to the convective heat transfer intensification and in outlet region the mixing leads to temperature balance in channel cross-sectional area and increase of residence time for needed decomposition value. Disadvantage of the cooling system is high hydraulic resistance as compared with straight channels.

Fig. 6 represents influence of the pentane decomposition on the outlet temperature of the

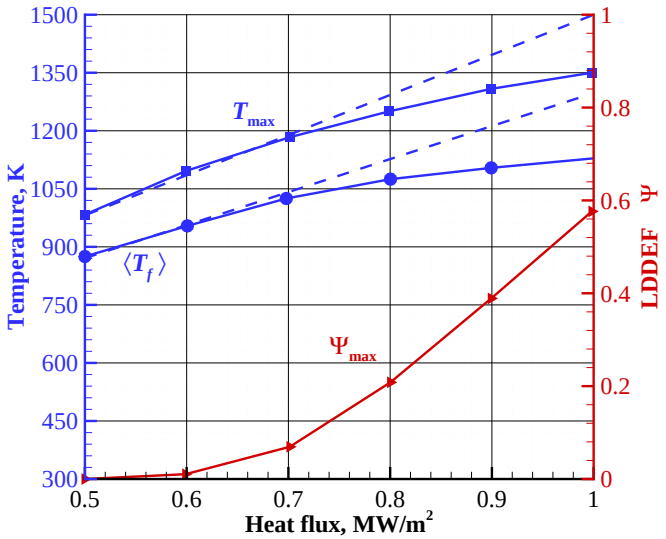


Figure 6: Influence of the endothermic effect on temperature of twenty sectional panel of the cooling system: without (dashed line) and with (solid line) taking the pentane decomposition into account at uniform heat flux.

cooling panel at different values of heat flux. Here $T_{\max}$, $\Psi_{\max}$ and $\langle T_f \rangle$ are given by (12) and (13) in outlet section, respectively. Maximum value of LDDEF $\Psi_{\max} \approx 0.6$ predicts absence of solid deposit formation on the channel surface under these conditions.

Results of numerical simulation of the heat transfer in twenty sectional panel of the cooling system are given in Table 1. The table represents maximal values of LDDEF $\Psi_{\max}$ (12) and temperature $T_{\max}$ (12), mass-averaged outlet values of LDDEF $\langle \bar{\Psi} \rangle$ (13) and EF temperature $\langle T_f \rangle$ (13) and mass-averaged outlet values of LDDEF $\langle \bar{\Psi} \rangle$ (14) in the outlet manifold (Fig. 4) at different heat fluxes. Note that the hot EF continues to decompose in the outlet heat-insulated manifold: $\langle \Psi \rangle < \langle \bar{\Psi} \rangle$. This fact should be accounted at the cooling system designing since the solid depositions formation in the outlet manifolds can take places at high values of LDDEF.

Fig. 7 shows distribution of temperatures, LDDEF, averaged Reynolds number and averaged pressure in the panel at $q = 1$ MW/m^2. In spite of strong change of the pentane thermal properties, pressure drop is 0.5 MPa (from inlet pressure 5 MPa downto outlet pressure 4.5 MPa). Computational results predicts that the panel (size 0.07×0.4, coolant (pentane) mass flow rate is 0.0067 kg/s, $T_{\text{inlet}} = 300$ K, $p_{\text{inlet}} = 5$ MPa) can be used for the combustion chamber

Heat flux	Two last sections		Outlet		Manifold
q, MW/m^2	$\Psi_{\max}$	$T_{\max}$, K	$\langle \Psi \rangle$	$\langle T_f \rangle$, K	$\langle \bar{\Psi} \rangle$
0.5	0.00	983.75	0.00	868.71	0.00
0.6	0.01	1088.98	0.01	953.04	0.01
0.7	0.07	1181.17	0.04	1025.02	0.05
0.8	0.21	1248.60	0.15	1070.53	0.17
0.9	0.39	1300.98	0.31	1098.90	0.34
1.0	0.58	1349.14	0.48	1125.04	0.52

Table 1: Result of the heat transfer simulation in twenty sectional panel of the cooling system.

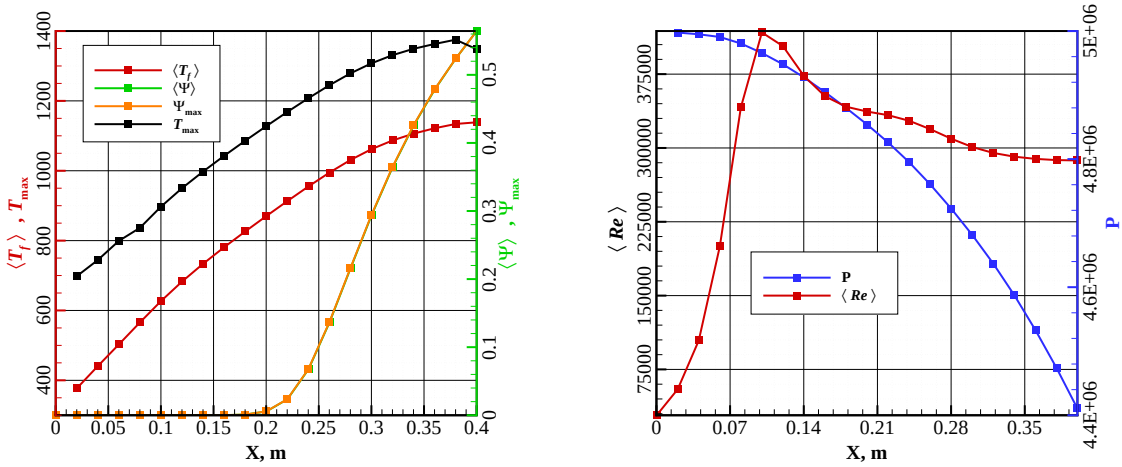


Figure 7: Distribution of LDDEF Ψ and temperature T (K) (left), Reynolds number Re and pressure p (Pa) (right) in twenty sectional panel of the cooling system at heat flux $q = 1$ MW/m^2.

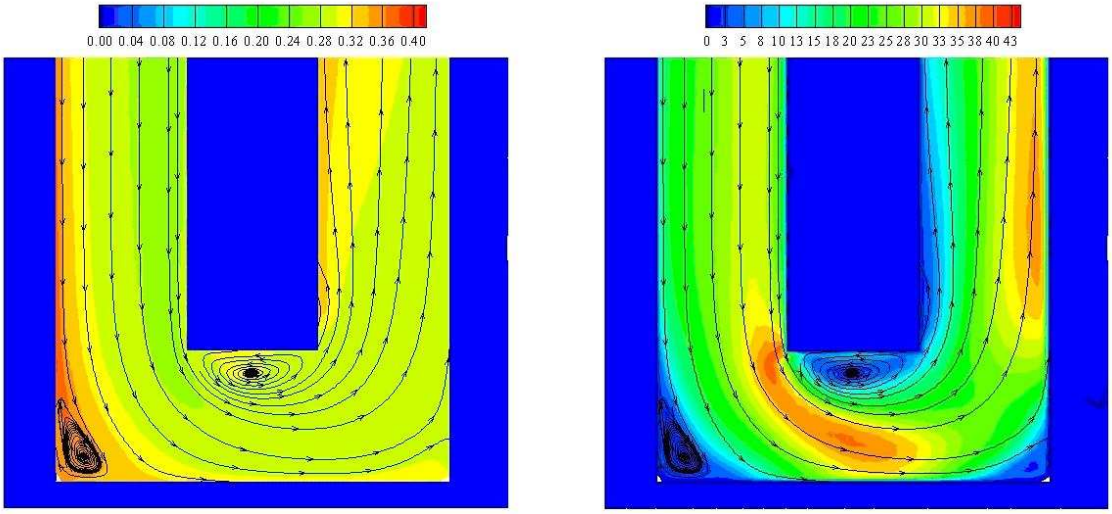


Figure 8: Distribution of LDDEF Ψ (left) and velocity modulus (m/s) (right) on boundary between penultimate and last heat exchange sections at heat flux $q = 1$ MW/m^2.

protection.

Fig. 8 represents the planar distribution of LDDEF and absolute value of EF velocity between penultimate and last heat exchange sections ($q = 1$ MW/m^2). The panel geometry gives intensive EF mixing in the channel. As a result, almost uniform EF temperature distribution in the channel cross-section is observed. In this case thermal decomposition of EF depends only on the local velocity. Recirculating zones lead to increase EF residence time and LDDEF [13].

4 FUTURE WORKS

Proposed mathematical models have advantages in computational efforts. The model allows simulate physical and chemical process in multidimensional reacting EF flows and burning in the combustion chamber in coupled manner. However real hydrocarbon fuels consist of many compounds and their thermal decomposition has complex multistage nature. Detailed mechanism of thermal decomposition of simple hydrocarbon fuels consists of many elementary reac-

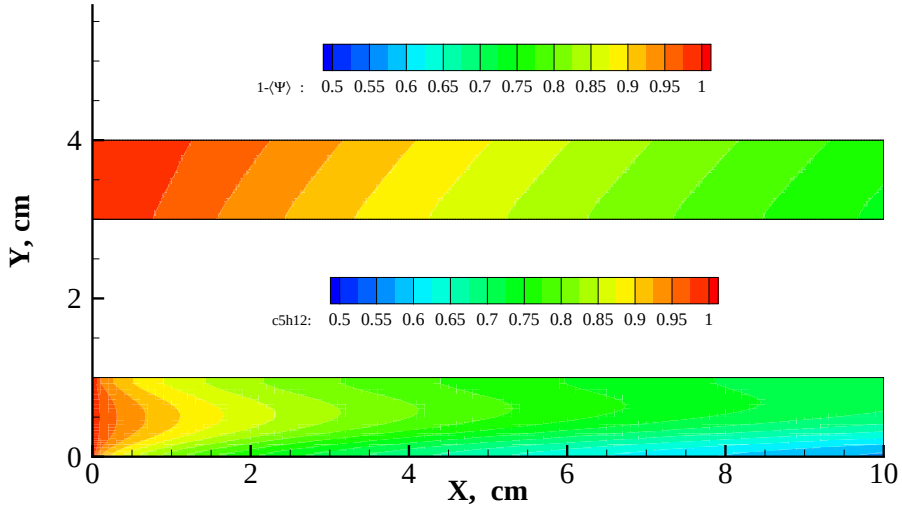


Figure 9: Mass fraction of pentane, laminar flow between parallel plates at $T_{\max} = 700$ K: brutto-approximation (upper figure) vs. detailed kinetics of the decomposition (lower figure).

tions. As a result, sufficiently accurate mathematical models can include thousands equations and it difficult to use for practical computations. The most promising approach is reduction to the simplified description. The simplified model should describe such the main features of full model as heat sources and outlet substances concentrations. Fig. 8 represents comparison of results obtained by using full model (more than 1000 reactions and 200 components [14]) and proposed model. Fig. 8 demonstrates simulation of 2D laminar flow between parallel plates, lower plate is heated and upper plate is heat heat-insulated. Inlet temperature is close to temperature of the decomposition start (700 K). Average cross-sectional values of the decomposition degree are similar to each other (maximum difference is no more than 10%), but their spatial distributions are differ.

Following development of proposed model is hybrid approach based on chain-radical mechanism of the thermal destruction and global brutto-correlations for components of the reacting system.

ACKNOWLEDGEMENT

The activity is a part of the work “Supercomputer simulation of physical and chemical processes in the high-speed direct-flow propulsion jet engine of the hypersonic aircraft on solid fuels” supported by Russian Science Foundation (project no. 15-11-30012)

5 CONCLUSIONS

3D mathematical model have been proposed to simulate the scramjet active cooling system and to determine the mixture composition at the cooling channel outlet. To minimize the computational efforts, the one-step pyrolytic reaction mechanism has been used for the pentane decomposition. As a result, it is possible to compute physical and chemical processes in the combustion chamber and its cooling system in coupled manner. At present time a mathematical model with more accurate description of EF decomposition based on complete detailed mechanism is developed.

REFERENCES

[1] L.S. Yanovski, Endothermic Fuels for Hypersonic Aviation in "Fuels and Combustion Technology for Advanced Aircraft Engines" *AGARD Propulsion and Energetic Panel 31st Symposium*. Fiuggi, Italy, May 10-14, 1993.

[2] L.S. Yanovski, E.P. Fedorov, G.I. Kovalev, M.L. Nemchikov, Investigation of the endothermic fuel thermal destruction under increased pressure. *Problems of aviation science and technology. Aviation propulsion engineering*, **3**, 41-51, 1987. (in Russian)

[3] L.S. Yanovski, T.N. Shigabiev, F.M. Galimov, V.F. Ivanov, *Endothermic fuels and working mediums of the power units*. Kazan, 1996. (in Russian)

[4] L.S. Yanovski, Features of heat transfer to the endothermic fuel on the initial segment of the heated channel under conditions of the liquid-phase autoxidation. *Problems of the chemical heat regeneration in aircraft and propulsion*, 1990. (in Russian)

[5] R. Hazlett, *Progress Report on Advanced Hydrocarbon Fuel Development*, Naval Research Laboratory, 342, 1975.

[6] F.Q. Zhong, X.J. Fan, G. Yu G. Thermalcracking of aviation kerosene for scramjet application. *Science in China, Series E: Technological Sciences*. 52, 2644-2652, 2009.

[7] T.A. Ward, J.S. Ervin, R.C. Striebich, Simulation flowing mildly-cracked normal alkanes incorporating proportional product distribution. *Journal of Propulsion and Power*, 20, 394-402, 2004.

[8] T.A. Ward, J.S. Ervin, S. Zabarnick, Pressure effect on flowing mildly-cracked n-decane. *Journal of Propulsion and Power*, 21, 344-355, 2005.

[9] P.V. Danckwerts, Local residence-times in continuous flow systems. *Chemical Engineering Science*. 9, 78-79, 1958.

[10] J.N. Baleo, P. Le Cloirec, Validating a prediction method of mean residence time spatial distributions. *AIChE Journal*, 46, 675-683, 2000.

[11] G.F. Froment, K.B. Bischoff, *Chemical reactor analysis and design*. New York: John Wiley & Sons, Inc., 1979.

[12] V. John, I. Angelov, A.A. Oncul, D. Thevenin, Techniques for the reconstruction of a distribution from a finite number of its moments. *Chemical Engineering Science*. 62, 2890-2904, 2007.

[13] P.D. Toktaliev, S.I. Martynenko, Mathematical model of the cooling system of combustion chambers of the aviation ramjet engines on the endothermic fuels. *Herald of the Bauman Moscow State Technical University, Series "Natural Science"*, **1**(58), 84-98, 2015. (in Russian)

[14] C.K. Westbrook, W.J. Pitz, O. Herbinet, H.J. Curran, E.J. Silke, A Detailed Chemical Kinetic Reaction Mechanism for n-Alkane Hydrocarbons from n-Octane to n-Hexadecane, *Combustion and Flame*, **156**(1), 181-199, 2009.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SENSITIVITY STUDY OF THE NUMERICAL SETUP FOR AN AUTOMATIC OPTIMIZATION PROCEDURE FOR A HYDRAULIC MACHINE

Alexander Tismer[1], Markus Schlipf[1] and Stefan Riedelbauch[1]

[1]Institute of Fluid Mechanics and Hydraulic Machinery
Pfaffenwaldring 10, 70550 Stuttgart, Germany
e-mail: alexander.tismer, markus.schlipf, stefan.riedelbauch@ihs.uni-stuttgart.de

Keywords: CFD, turbo-machine, structured-unstructured mesh, OpenFOAM, machine design

Abstract. *For an automatic optimization using Computational Fluid Dynamics, all design iterations of a turbo-machine require the generation of geometry and computational mesh as well as the simulation of the flow and the evaluation of one or more characteristic properties. All steps can be expensive regarding time or manpower. In addition, all meshes of the design iterations differ in total number of cells, percentage of a particular element type, local refinement sections or structured mesh block thickness being summarized as mesh characteristics. An object-oriented framework focused on hydraulic machinery is used to generate different geometry variants. Furthermore, the framework is extended to create hybrid meshes consisting of a structured part surrounding the blades and an unstructured part within the remaining flow channel. The structured-unstructured mesh coupling is realized with the Pyramid Open Method. In order to take care of the boundary layer flow effects along the meridional contour upstream and downstream of the runner blade rows a prismatic layer creation algorithm is also implemented. OpenFOAM software is applied to simulate the flow through the hydraulic machine. Each step in the process chain is implemented as part of the framework. The complete study is done using open source applications, libraries and tools. In terms of performing an automatic optimization procedure an assessment for the numerical behavior of the hybrid mesh in a real-engineering application is desired. To limit the number of optimization parameters and the amount of computation time the sensitivity analysis tool supports the identification of relevant geometry input parameters with significant influence on the flow field. On the one hand it is important to discover the influence of geometry variations on the predicted physical behavior (e.g. flow losses, torque and head) of the simulation results. On the other hand it is investigated if the uncontrollable changes of mesh characteristics or user controlled numerical discretization parameters, such as flux gradients affect the physical quantities on the same level as the geometric variations. The goal is to achieve a dominating geometry influence and a minor effect of mesh characteristics on the predicted flow field solution. For both, a typical Francis and a Propeller turbine runner, the fluctuations and absolute values of physical properties are shown for a number of varying meshes, numerical and geometrical setup parameters.*

1 INTRODUCTION

In almost all cases hydraulic parts of power plants are constructed individually for each power plant and hence for special operating conditions. From an engineer's point of view this means finding a geometry that fits best to the desired operating conditions. Each design iteration requires generating a geometry, meshing the domain, simulating the flow field and investigating the solution. Iterations or variants of a hydraulic machine have to be evaluated against each other to distinguish if a variant is better than the others. In addition a design process is normally limited by a project duration and in that sense it is worthwhile to automate all steps as much as possible.

The very first step of the process is to parameterize a geometry to be able to create different variations. For hydraulic machines, e.g. Francis or Propeller runners, the blade angles at leading and trailing edge, blade thickness, meridional and circumferential blade length serve as possible design degrees of freedom (DDoF). A geometry of a machine can be seen as a function that takes the DDoF as inputs and generates a geometry. Frequently, geometry and mesh generation applications act on different file formats which produces the necessity to convert output of a first application to the correct input for a second application. Meshing a hydraulic machine with a conservative block-structured grid is a highly intuitive and manpower intensive step in the optimization loop. A promising approach to reduce manual effort is the use of unstructured or partly unstructured meshes. A hybrid mesh fits more easily a wider range of geometry variants due to its variable unstructured part. On the other hand it is known that unstructured meshes also produce more inaccuracy regarding numerical errors [1]. Hybrid meshes with hexahedral block-structured elements around the blade and prisms in the boundary layer at hub and shroud are used to reduce numerical errors. The flow field in the hydraulic machine is obtained from a numerical solution of the incompressible Navier-Stokes equations on the mesh, e.g. using a finite volume approach. In the last step of a design iteration an obtained flow field has to be evaluated and classified. Characteristic properties of a design (CPoD) with either scalar or vector character have to be defined. Common physical CPoD are e.g. torque, efficiency and velocity distribution downstream the runner. The question, if a design of the hydraulic machine fits better to the operating conditions than others, is answered by evaluating those parameters.

2 FRAMEWORK

Hydraulic design environments based on applications that have a set of fixed DDoF are meaningful approaches for new designs of hydraulic components [2, 3]. Refurbishment projects need tools that are customizable and extensible. The authors believe that a framework can provide both and hence is able to be applied to both scenarios. Adaptability from the user's perspective is achieved by the use of a modifiable, human-readable construction plan of the machine. For a developer an object-oriented code structure can reduce the work for an implementation of a new feature. The already developed framework is able to handle fairly different machine types and is extended for this study [4]. A lot of open-source libraries reduce the amount of new code. The framework can be divided in different sections: geometry, mesh, pre-processing, flow simulation and post-processing. All steps of the workflow are part of the framework and thus, can be extended by modifying the source code. In order to extend the possibilities regarding geometrical modifications the underlying geometry engine of the framework is adjusted to use the well-established open source project OPENCASCADE [5]. Further used libraries remain the same (see [4] for a list).

2.1 STRUCTURE

Figure 1 shows the general structure with its main parts. The left column lists parts that a "common" user or programmer interacts with. These are implementations for geometrical objects e.g. one-, two- and three-dimensional mappings or mesh objects. The framework or rather OPENCASCADE uses non-uniform rational B-Splines (NURBS) to handle free-form geometrical shapes. It is also possible to handle compositions of NURBS or other geometrical objects [4]. The *solverKernel* provides the interface to a simulation tool that computes a numerical solution of the flow field. In order to provide a mechanism for developers to extend the framework without recompiling or modifying the original source code, a simple plugin system is also provided [6]. For this study the post-processing steps are implemented as plugins. Parts in the right column of figure 1 are generalized objects, e.g. handling string objects, exception handling or logging functions. The user interacts with the framework via a Hypertext Markup Language file that serve as a construction plan for the machine. Additionally, a very first and simple stage of parallel support using MPI (Message Passing Interface) and OpenMP (Open Multi-Processing) is implemented. The connection to the used underlying open source libraries is bundled in the *libraryKernel*. The right column of figure 1 contains an incomplete list of used libraries. The last part *applicationKernel* provides some basic methods to write new applications that use the library and as a consequence become part of the framework. For this study an application that executes geometry generation, mesh generation, simulation and post-processing for one design sample is provided. The black box that surrounds all steps in the figure visually represents the mentioned application.

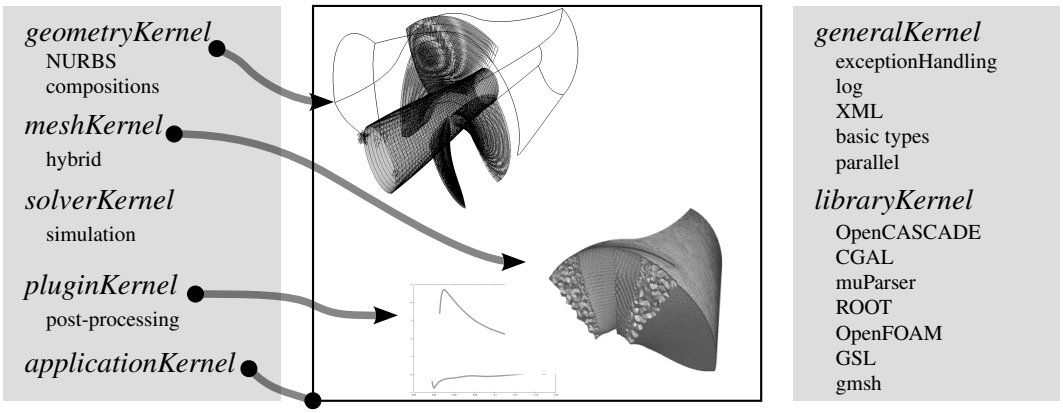


Figure 1: General framework structure

The mesh generation part uses the open source software GMSH [7]. Already implemented meshing steps of GMSH such as triangulation, tetrahedralization and creation of structured mesh blocks using a transfinite interpolation are used without modifications. The framework creates the mesh that depends on different d-dimensional geometrical objects G^d with the parameter space coordinates u, v and w in the same sequence as GMSH does:

(1) Vertices G^0 that are entities of dimension 0,
(2) Edges $G^1(u)$ that are entities of dimension 1,
(3) Faces $G^2(u, v)$ that are entities of dimension 2 and
(4) Regions $G^3(u, v, w)$ that are entities of dimension 3.

GMSH provides a bidirectional data structure of all G^d entities and, thus, enables to have all

upward $(d+1)$-dimensional and downward $(d-1)$-dimensional adjacencies of a geometrical object. Table 1 gives a list and short description of new implemented operators that are all involved in creating the hybrid meshes. The mesh is created in the same order for steps (1) - (3). Step (4) is subdivided into:

(a) create boundary layer elements on wall faces and take transfinite layers into account,
(b) add pyramid elements on quadrangles that are not part of boundary layers and
(c) fill unstructured part of regions.

Operator name	Description
dtMeshFreeGradingEdge	edge with user defined vertex distribution
dtMeshGFaceWithTransfiniteLayer	four sided face with a layer at two opposing edges
dtMeshGRegionWithBoundaryLayer	region with faces that can extrude (create a boundary layer), where vertices can slide and where vertices remain fixed (see section 2.3)

Table 1: Additionally implemented mesh operators

2.2 EDGE- AND SURFACE MESH

dtMeshFreeGradingEdge meshes an edge of length L and takes a predefined function

$$g : \mathbb{R} \to \mathbb{R} : g(x) = y \text{ with } g(0) = 0 \text{ and } g(1) = L \tag{1}$$

into account. The function space x is equidistantly divided into an user defined number of points. The i-th geometrical length l_i of the i-th mesh point is then given by

$$g(x_i) = l_i \ .$$

For this study a two dimensional B-Spline function

$$g_B : \mathbb{R} \to \mathbb{R}^2 : g_B(u) = (x_u, y_u)$$

is defined with three control points. In the sense of equation (1) g_B is forced to be a function by finding a root for the first dimension of

$$g_B(u) = (x_u - x_g, y_u) \tag{2}$$

with a given x_g. It is up to the user to make sure that the solution of equation (2) is unique. Controlling the grading by control points of B-Splines is assumed to be more intuitively from the user's perspective. But grading is not fixed to B-Spline functions only and hence hyperbolic tangent, geometric series, polynomial functions or others can be used.

The mesh operator *dtMeshGFaceWithTransfiniteLayer* acts on two dimensional mesh entities $G^2(u, v)$ according to figure 2. As first step the structured dark gray layer is created using a transfinite interpolation scheme. The second step creates the light gray unstructured part. Layers are not limited to have the same number of elements in u- and v-direction. The implementation of the operator is realized by a combination of existing operators of GMSH.

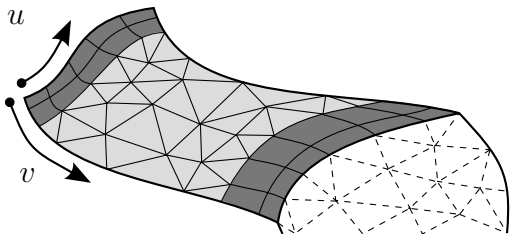


Figure 2: Face with a transfinite layer

2.3 PRISMATIC LAYER

Fluid dynamic boundary layer effects give rise to have special elements near walls. For this study a prismatic boundary layer is created on hub and shroud of the hydraulic machine. Figure 3 shows the dark gray layer as an extrusion of the unstructured light gray surface mesh. A simplified implementation of a generalized advancing layer technique is used to create the prismatic elements [8]. In the following a vertex positioned on a G^2 entity is called ordinary vertex and a vertex located at a G^1 or G^0 entity is called either sliding or fixed vertex. Firstly, a mesh manifold for each mesh vertex is created by finding a face set that shares one common mesh vertex. In figure 3 the manifold of the mesh vertex marked with a black rectangle is defined by all I light gray surface elements s that are touched by the dashed circle around the rectangle. Each surface element s_i has a unit normal vector $\underline{N}_S(i)$. The mesh unit normal $\underline{N}_V$ at mesh vertex j

$$\underline{N}_V(j) = \frac{\sum_i^I c_i \underline{N}_S(i)}{|\sum_i^I c_i \underline{N}_S(i)|} \tag{3}$$

has to be calculated for each ordinary vertex. For all ordinary points c_i is set to one. For each ordinary mesh vertex $\underline{N}_V(j)$ returns its unit normal vector and thus the direction of the extruded boundary layer. Special handling is necessary for sliding and fixed vertices. A sliding vertex is able to move along an internal mesh edge of a surface. A boundary layer that starts from the dashed surface mesh, figure 2, contains sliding vertices at mesh edge $G^1(u) = G^2(u, v = 1)$. All mesh vertices on the geometrical edge slide in negative v-direction of the internal edges located on the dark gray surface mesh of the face. In this study the boundary layer of a face is defined by the transfinite layers of its neighbor faces and consequently the normals for this constrained mesh vertices are initially not necessary to compute. But in order to prevent collisions in the mesh the normals for these vertices are approximately calculated, too. Surrounding normals of a sliding vertex are calculated using equation (3) with $c_i = 2$ if i-th mesh surface s_i contains at least one slidable vertex. A fixed vertex is a vertex that does not move and thus, the normal for this vertex is set to $\underline{0}$. In order to reduce gradients in $\underline{N}_V$ the t-th iterative smoothing step of the unit normal at vertex i with all K direct neighbors

$$\underline{N}_V^{t+1}(i) = \frac{\sum_k^K \underline{N}_V^t(k)}{|\sum_k^K \underline{N}_V^t(k)|}$$

is performed. The final extrusion creates a layer according to the dark gray elements of figure 3. On the one hand the resulting boundary layer consists of prisms that are created by pure ordinary vertex extrusions. On the other hand partly fixed vertex extrusions produce pyramids or tetrahedra for surface elements that contain one or two fixed vertices, respectively.

It is important to mention that the simplified algorithm in the framework does not provide all features mentioned in [8]. Caused by the predefined transfinite layers, it is not necessary to

calculate multiple growth curves of mesh vertices. But it also has be stated that the implemented algorithm is not as general as the algorithm found in the literature.

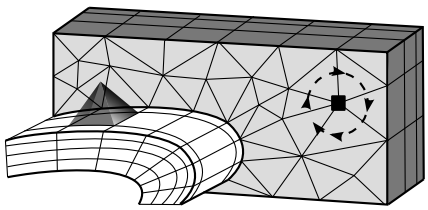


Figure 3: Prismatic layer (dark gray) and Pyramid (black shaded) as coupling element

2.4 STRUCTURED-UNSTRUCTURED COUPLING

The tetrahedral mesh in the resulting flow channel that is not part of the structured mesh block surrounding the blade has to be coupled to the structured hexahedral mesh. The Pyramid Open Method (POM) is a robust coupling algorithm. It is reported that the method is more stable compared to the Tetrahedral Transformation Method, but produces pyramids of lower quality. An advantage is that the POM does not modify or rather divide any existing hexahedra and additionally, the algorithm is easier to implement [9].

Starting with two nonconforming meshes of the tetrahedral and hexahedral part, the method opens pyramids from the coupling face. Figure 3 shows a pyramid (black transparent color) that starts to open from the white structured part. In its original formulation POM inserts a mesh vertex in the center of the quadrangle, subdivides all direct neighboring tetrahedral elements and then opens the pyramid. A quadrangle needs to have two adjacent triangle faces or rather a hexahedron has to share one face to two different faces that belong to two different tetrahedra. The implementation in the framework starts from a hexahedron that shares a face to four different faces of also four different tetrahedra. The shared face already contains the hanging vertex. After the insertion of a zero volume pyramid element the nonconforming mesh is topologically coupled. The iterative procedure takes each of the J quadrangle mesh vertices $\underline{\mathcal{Q}}$ and each from the K vertices that belong to an adjacent tetrahedron $\underline{\mathcal{T}}$ into account. The initial position ($t = 0$) of the apex vertex $\underline{C}$ of the i-the pyramid

$$\underline{\mathcal{C}}_i^{t=0} = \frac{\sum_j^J \underline{\mathcal{Q}}_j + \sum_k^K \underline{\mathcal{T}}_k}{J + K}$$

moves the apex vertex into the tetrahedral mesh and thus increases the pyramid's volume. Due to the mesh deformation it is possible that some adjacent tetrahedra become illegal (negative volume). An iterative procedure checks the volume of all adjacent tetrahedra for the t-th iteration step. If illegal elements occur the apex moving

$$\underline{\mathcal{C}}_i^{t+1} = (1 - r)\underline{\mathcal{C}}_i^t + r\underline{\mathcal{C}}_i^0$$

is constrained using a weighted average of the 0-th (initial) and the t-th vertex position. The relaxation parameter r is a user-defined scalar constant. Following the proposal of the algorithm in [9] a quality check for the shape metric is performed, too. The modified procedure does not check element quality due to multiple mesh optimization steps following the POM and

consequently creates pyramidal elements with a larger volume. Thus, the quality and volume of the deformed tetrahedral elements are lower after the implemented POM compared to the original algorithm.

2.5 PRE- AND POSTPROCESSING

For this study the open source Computational Fluid Dynamics software package OPENFOAM is used to simulate the flow field in the hydraulic machine [10]. All simulation files and data are directly written by the framework. Parts of OPENFOAM are connected to the framework as shared libraries e.g. used boundary conditions, mesh read- and write support or some general input- and output (IO) methods. It prevents avoidable conversion steps due to different formats. Additionally, the direct pre-processing of OPENFOAM via its provided libraries gives the possibility to write all necessary input files and thus is implicitly corrected if the underlying version changes. Furthermore, it is possible to use the large amount of provided functions for field operations in the sense of OPENFOAM. Therefore, also post-processing profits by the direct connection to the simulation library, e.g. the mass flow averaged circumferential velocity distribution on arbitrary user-defined surfaces or a pressure distribution on spanwise constant blade cuts are possible to calculate.

3 MODELS

For this study two typical hydraulic machines are investigated. The geometrical appearance of both, a Francis and a Propeller runner, is fairly different, but the underlying used geometrical parameterization is very similar. The template that is defined by the geometrical parameterization of the machine is either automatically or manually adjusted (change of DDoF) to fit the existing geometry. The intention of this study is not to model the existing geometry as exact as possible and not to "re-simulate" the machines. In fact, the machines serve as examples for the automatic hybrid grid generation method and in addition to discover the peculiarities of this general workflow approach.

3.1 GEOMETRICAL PARAMETRIZATION

In classical refurbishment tasks the meridional shape outside the runner is fixed and therefore not considered as DDoF for the new design. Consequently, the meridional contour needs not to be parameterized and thus, a defined number of points, picked from the contour, is imported into the framework. The blade and its center plane is transferred to the framework, too. It is important to understand that for an existing machine it is initially not possible to find its original parameterization. In reality every machine designer is free to use his own parameterization. Both machines are modeled with a three point mean line represented by a second order B-Spline defined with three control points. It is possible to define the inlet α and outlet angle β but also the meridional length of the blade H. The circumferential length of the blade is defined by the ratio r. The interested reader is referred to [4] for a detailed description. The sum of the thickness distribution $y(u,t)$ for a four-digit NACA profile with maximum thickness t_{max} and parameter coordinate u and the trailing edge thickness t_{te}

$$t = y(u, t_{max}) + t_{te}$$

defines the blade thickness [11]. The coordinate u of the blade surface points in meridional direction, whereby v represents the spanwise direction from hub to shroud. Both machines have a spanwise constant trailing edge thickness ($t_{te} \neq t_{te}(v)$). The Francis runner also has

a spanwise constant ($t_{max} \neq t_{max}(v)$) maximum thickness t_{max}. In contrast the Propeller's maximum thickness

$$t_{max} = t_0(1 - v) + t_1 v$$

linearly depends on the spanwise coordinate v with thickness t_0 and t_1 at hub and shroud, respectively. Both blades of the machines have equal thickness distributions on suction and pressure side. Table 2 lists angles and main dimensions with unit in squared brackets of the two machines. Functions b_F and b_P, with their input in brackets, represent the parameterization set of the Francis and Propeller runner. Increasing numeric index means increasing spanwise position from hub to shroud. DDoF with three positions (1, 2 and 3) are connected via a B-Spline function, two positions of a DDoF (0 and 1) are linearly connected. As already mentioned the values of table 2 are not known or able to be easily calculated for an existing blade. Both machines are approximately reparameterized with the geometrical template by reparameterizing discrete points on the imported initial blade.

	$b_F(\alpha_{1-3}, \beta_{1-3}, H_{1-3}, r, t, t_{te})$	$b_P(\alpha_{1-3}, \beta_{1-3}, H_{1-3}, r_{1-3}, t_0, t_1, t_{te})$
[°]	$(\alpha_1, \alpha_2, \alpha_3) = (88, 59, 45)$	$(\alpha_1, \alpha_2, \alpha_3) = (76, 34, 20)$
[°]	$(\beta_1, \beta_2, \beta_3) = (45, 28, 18)$	$(\beta_1, \beta_2, \beta_3) = (53, 26, 17)$
[m]	$(H_1, H_2, H_3) = (0.27, 0.24, 0.19)$	$(H_1, H_2, H_3) = (0.070, 0.070, 0.063)$
[m]	$r = 0.75$	$(r_1, r_2, r_3) = (0.400, 0.525, 0.500)$
[m]	$t_{max} = 0.005$	$(t_0, t_1) = (0.1, 0.025)$
[m]	$t_{te} = 0.005$	$t_{te} = 0.001$

Table 2: Angles and main dimensions of Francis and Propeller runner

3.2 NUMERICAL SETUP

OPENFOAM (foam-extend version 3.1) provides "MRFSimpleFoam", a solver for systems with a rotating frame of reference using the SIMPLE algorithm [12]. The investigations only include rotating parts of the machine and thus, it is not necessary to have any stator-rotor interface. Additionally, the solver produces a stationary solution. Volume or mass flow is predefined by a given velocity distribution at the inlet patch and outflow is against zero pressure. On the periodical patches the individual grid elements are nonconforming and therefore are coupled by a general grid interface. The outflow of the domain is enforced for the Francis runner by the use of the "inletOutlet" boundary condition. To take care of turbulent flow effects the simulation uses a standard k-ϵ turbulence model with a given turbulent kinetic energy intensity of 10% at inlet. The mixing length L for the calculation of the dissipation rate ϵ can be estimated based on a pipe diameter d_h

$$L = 0.038 d_h$$

for a fully developed pipe flow. Diameter d_h is replaced with the channel height of the machine. All calculations run 1000 iterations in total thereof the first 100 iterations are laminar to get a stable initial solution.

The discretization scheme can be adjusted in OPENFOAM. It is possible to change the approximation of the gradients and divergence terms in a numerical simulation. It is reported that the least square discretization scheme tends to be more oscillating on tetrahedral meshes. The gradient approximation is also not as accurate as on structured or rather hexahedral meshes and

thus, it is recommended to use gradient limiters. For this study, unless otherwise specified, a cell limited scheme with blending factor 0.33 is used that tends to be less dissipative than a face limited scheme [13].

4 SENSITIVITY ANALYSIS

For complex engineering models it is a challenging task to get knowledge about the influences of the input to the output. Especially for automatic optimization processes, it is important to ensure that the inputs like geometry variations that are adjusted by an optimizer have a predominant influence to the hydraulic properties, e.g. torque, head and efficiency. Other uncontrollable influences such as grid variations need to have less dominance on the output. A sensitivity analysis (SA) is a useful tool to investigate this behavior [14]. For the sampling and evaluation of the study the open source MATLAB/Octave toolbox "Sensitivity Analysis For Everybody" (SAFE) is used [15]. For all studies the samples are created by the Morris method that has a better economy of the sampling plan [16].

4.1 BASIC EQUATIONS

For a numerical model $\mathbb{M}$ that has an input $\underline{x}$ with K elements and an output y

$$y = \mathbb{M}(\underline{x}) = \mathbb{M}(x_1, ..., x_K)$$

an elementary effect (EE) is defined as

$$\mathrm{EE}_i(y) = \frac{\mathbb{M}(x_1, ..., x_{i-1}, x_i + \delta, x_{i+1}, ..., x_K) - \mathbb{M}(\underline{x})}{\delta} \,. \tag{4}$$

For each input i one or more EE of the output y can be determined between x_i and $x_i + \delta$. Equation (4) can be interpreted as a discrete gradient and gives the variation of the output y with respect to the i-th input x_i. A desired number of R EE for each of the K inputs require for a simple sampling method $2RK$ evaluations of $\mathbb{M}$. In the following the distribution of EE for the output y of an input i is denoted as $\mathrm{EE}_i(y)$. The chosen Morris sampling method has a better economy and requires $R(K + 1)$ evaluations of the model [16]. It is the main aim of the method to determine which input is

(a) negligible,
(b) linear or additive,
(c) non-linear or
(d) involved in interactions with other inputs.

At first, this study is not distinguishing between (c) and (d). The other distinguishing criteria are explained later. Common practice is to calculate the standard deviation $\mathrm{std}(\mathrm{EE}_i(y))$ and average $\mathrm{mean}(\mathrm{EE}_i(y))$ for i-th input. In its simplest interpretation a high mean value indicates inputs with a high overall influence and a high standard deviation is a hint for a highly dependent input that may also interact with other inputs [16]. For the application of the Morris sampling method the standard error of the mean (SEM)

$$\mathrm{SEM} = \frac{\mathrm{std}(\mathrm{EE})}{\sqrt{R}} \tag{5}$$

gives an additional indication whether an input is considered as significant or not [16]. Output values with dimensions are designated with $\mathscr{Y}$. Mean values of all samples of one run, e.g. $\mathrm{mean}(\mathscr{Y}(M))$, are used to generate dimensionless output values Y, e.g. $Y(M) = \frac{\mathscr{Y}(M)}{\mathrm{mean}(\mathscr{Y}(M))}$.

5 RESULTS

A first investigation uses the Francis runner according to table 2 and only varies numerical and mesh parameters while keeping the DDoF fixed. Table 3 gives all parameters and ranges. The symbol "⇔" means that this parameter is adjusted from the lower to the upper bound according to its column of the table.

$n_H[\%]$	$p_H[\%]$	$n_{dT}[\%]$	$n_{dH}[\%]$	$g[-]$	$g_P[-]$	$g_U[-]$	$g_{\phi,U}[-]$	$l[-]$	$s[-]$
0	0	0	0	0.20	0.20	0.20	0.20	0.20	0.20
⇔	⇔	⇔	⇔	⇔	⇔	⇔	⇔	⇔	⇔
50	3	30	50	0.40	0.40	0.40	0.40	0.40	0.40

Table 3: Parameters with ranges for a run with $R = 20$ and $K = 10$

In order to simulate different mesh variations that occur in an automatic optimization process an implemented mesh operator perturbs an initial given mesh. Perturbations are inserted in the mesh in two variations. On the one hand each edge of n_H percentage of all hexahedra is shrunk by p_H percentage of its initial length. On the other hand n_{dT} percentage of tetrahedra are divided in four new tetrahedra or n_{dH} hexahedra are split in six pyramids. Both division strategies are done by inserting a new mesh vertex in the barycenter of the initial tetrahedron or hexahedron. A random generator creates indices for mesh elements that are perturbed. Both hexahedron perturbation techniques influence the structured part and thus influences the mesh near the blade. The splitting of tetrahedrons is a local refinement and in most cases a mesh quality modification in the unstructured far field. By the use of a random generator two samples having the same values of n_{dT} differ in the position of local refinement sections while the number of elements remains unchanged. Element perturbation is only done once per element and elements that are connected to prisms at the wall or pyramids at the structured-unstructured coupling are kept unchanged. Table 3 gives the percentage and intensity of all inserted perturbations.

All g, s and l parameters of table 3 are blending factors for the gradient limiters of the discretization scheme used to stabilize the numerical flow field simulation on tetrahedral meshes. Furthermore, it is not unusual to compensate a low mesh quality with high blending values for the limiters. The blending parameters in table 3 are replaced in the discretization file

```
gradSchemes {
  default          cellLimited Gauss linear g;
  grad(p)          cellLimited Gauss linear gP;
  grad(U)          cellLimited Gauss linear gU;
}
divSchemes {
  div(phi,U)       Gauss linearUpwindV cellLimited Gauss linear gφ,U;
}
laplacianSchemes {
  default          Gauss linear limited l;
}
snGradSchemes {
  default          limited s;
}
```

of OPENFOAM. A detailed description of all limiters can be found in [13].

The fluctuations of torque $\widetilde{M}$ and head $\widetilde{\Delta_H}$ within the last 200 iteration steps can be interpreted as a physical convergence criteria. Figures 4a and 4b give the EE plots of the fluctuations. According to [16] the dashed line with slope $2\,\mathrm{SEM}$ of equation (5) is plotted. According to equation (5), SEM is the standard deviation of the distribution divided by the square root of the

sample size. In general, the mean value of a random sample is not correct or rather can only be estimated within a range of uncertainty. SEM is a confidence interval of the estimated mean value and thus, a quantity for the uncertainty. The prefactor 2 gives the 95-th percentile of the estimated mean value based on the assumption of a normal distribution. In other words, the estimated mean value $\mathrm{mean(EE)}$ varies with a probability of 95% between

$$\mathrm{mean(EE)} - 2\mathrm{SEM}$$

and

$$\mathrm{mean(EE)} + 2\mathrm{SEM}\ .$$

In that sense, a calculated mean value $\mathrm{mean(EE)}$ that lies on the left side of the dashed line can also be a mean value of zero that spreads caused by taking only a random sample into account. For that reason input parameters fulfilling the inequality

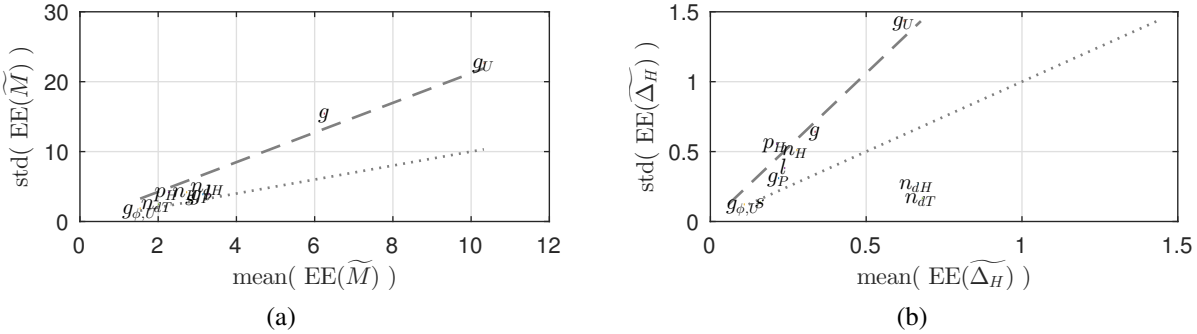
$$\mathrm{std}(EE) > 2\mathrm{SEM} \tag{6}$$

could have a mean value of zero.

The dotted line is the bisect line and hence indicates points that have equal standard deviation and mean. EE of input parameters that lie on the right side of the dotted line and additionally fulfill

$$\mathrm{std}(EE) \ll \mathrm{mean}(EE) \tag{7}$$

have a linear influence to the output and no interaction effects with other inputs. Whereby points located on the left side have non-linear and/or interaction effects [17]. Inputs that are very close to the origin are assumed to have no influence. It is obvious that g_U is separated in both figures 4a and 4b and additionally, it is located on the left side of the dashed line. The value g_U is the blending of the limiter of the $\mathrm{grad}(U)$ terms for the numerical discretization. Additionally, figure 4b gives the indication that mesh perturbations due to hexahedron and tetrahedron division are more critical regarding $\widetilde{\Delta_H}$. Or in other words $\widetilde{\Delta_H}$ seems to be more sensitive to the mesh. However, it has also to be stated that overall fluctuations of both physical quantities are quite small. The mean value of $\mathscr{Y}(\widetilde{M})$ and $\mathscr{Y}(\widetilde{\Delta_H})$ is 0.02% and 0.41%, respectively. Parameters n_{dH} and n_{dT} are on the right side of the dotted line and satisfy equation (7). For that reason $\widetilde{\Delta_H}$ linearly depends on both perturbations. All other numerical setup parameters lie close to the origin or close to the dashed line.



(a) (b)

Figure 4: EE plot for Francis runner of (a) $\mathrm{EE}(\widetilde{M})$ and (b) $\mathrm{EE}(\widetilde{\Delta_H})$

The hydraulic losses H_V are given by the energy equation

$$H_V = \Delta_H - \frac{M\omega}{\rho Q g}$$

with head Δ_H, torque M, angular velocity ω, density ρ, volume flow Q and gravitational constant g. Figure 5a gives the plot of the distribution of hydraulic losses $Y(H_V)$ versus the number of divided hexahedra n_{dH}. The gray solid lines are linear regressions that subtend the sample $\min(Y(M))$ and $\max(Y(M))$. The H_V coordinate of the intersection points with the bottom and the upper linear regression are given in figure 5a as b_{min} and b_{max}, respectively. The slope of the regression is named m and given in the figure, too. The lines are not necessarily bounding lines of all samples. Figure 5b is the equivalent plot with respect to the number of divided tetrahedra n_{dT}. Table 4 gives the number of overall, tetrahedra, pyramids and hexahedra mesh elements.

	min	max
overall $[10^3]$	621	940
tetrahedra $[10^3]$	345	446
pyramids $[10^3]$	21	338
hexahedra $[10^3]$	129	182

Table 4: Number of overall elements and different element types

It is obvious that the division of hexahedra produces more spread and lower slope compared to division of tetrahedra. It is a well known phenomena in simulation that refining a mesh may cause changes in the physical quantities. Both figures 5a and 5b show the same trend. A higher number of divided elements resulting in a finer grid resolution increases the hydraulic losses. The comparison of points with equal abscissa values gives an indication of output fluctuations, i.e. H_V, caused by an uncontrollable change of mesh characteristics due to the randomly positioned refinement of tetrahedral elements. Having this in mind the bounded distance in H_V direction of the section, or equivalent to the difference between b_{min} and b_{max}, gives the amount of fluctuations that are uncontrollable inserted due to mesh modifications. According to figures, the fluctuations of hydraulic losses H_V for hexahedra and tetrahedra splitting are approximately 9% and 3%, respectively. It has to be mentioned that hexahedra splitting is not an uncontrollable mesh change, because under normal circumstances this does not occur. But in terms of getting an indication for the behavior of unstructured meshes the observations show that physical properties are sensitive to the structured part of the mesh.

Summarizing, the influence of the uncontrollable mesh characteristics on the output results, caused by different orderings of tetrahedrons, is small and in addition of linear character. Consequently, a highly non-linear and fluctuating behavior of the mesh is not observed.

A second run using the Francis runner is created by an additional variation of geometrical DDoF namely blade inlet (α_1, α_2 and α_3) and outlet angles (β_1, β_2 and β_3). The mesh parameter λ is the thickness of the structured mesh block that surrounds the blade. In addition, the structured part that consists of four hexahedral blocks is varied. Only the hexahedral block located on the blade tip is changed by moving mesh vertices on hub ($d_{0,0}$ and $d_{1,0}$) and shroud ($d_{0,1}$ and $d_{1,1}$). Last column "[...]" in table 5 has to be seen as a link to columns g, g_P, g_U, $g_{\phi,U}$, l and s of table 3 resulting in a run that contains 17 input parameters. The symbol "$\pm$" means that a

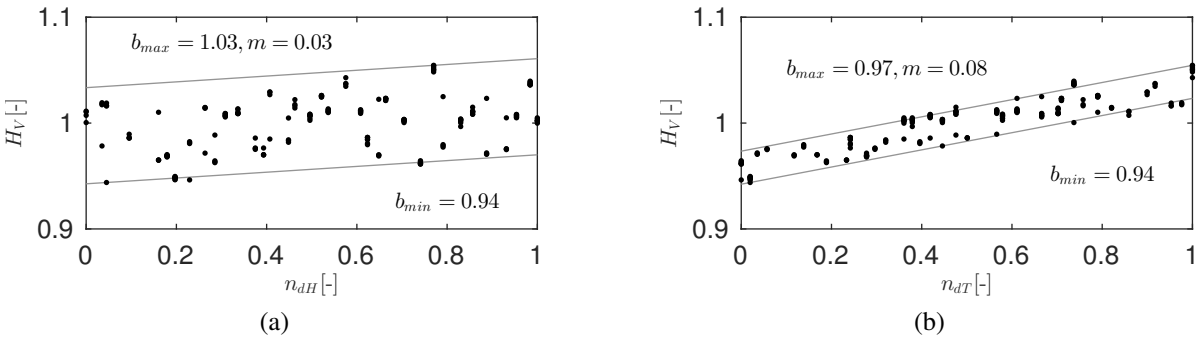


Figure 5: Hydraulic losses for Francis runner versus normalized input (a) n_{dH} and (b) n_{dT}

parameter is adjusted by addition or subtraction by the given bounds to the initial value listed in table 2. If samples fail it may be necessary to remove more than only the failed samples, caused by the Morris sampling method. Following a recommendation of [17] the value of R should be between 4 and 10. For the sake of completeness it is mentioned that each failed sample can, but must not necessarily, reduce R by one. It depends on the distribution of failed designs. A failed sample can reduce the number of samples in this case by $M + 1 = 18$ samples. In its worst case only 5% failed samples are able to reduce R to zero. A miserable distribution of 4.72% failed designs reduces R to 8 for this run. It is observed that many failed samples have small values of $d_{0,0}$ and high values of $d_{0,1}$. The combination represents a strongly distorted mesh block at the blade tip, regarding the shape from hub to shroud.

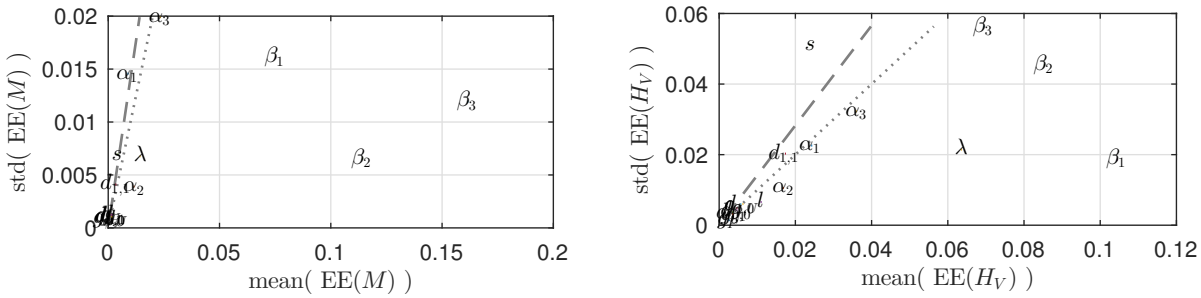
The mean value of $\mathscr{Y}(\widetilde{M})$ and $\mathscr{Y}(\widetilde{\Delta_H})$ are 0.07% and 0.25%. It is important to have in mind that for this study also geometrical changes are allowed and thus, it is stated that some samples may not have a stationary solution of the flow field. Furthermore, varying blade angles at three points for inlet and outlet may also result in an uncommon geometrical shape of the runner. In fact, it is possible that some samples have a set of DDoF that produce a geometry that does not make sense from a well skilled engineer's point of view. For that reason a simulation of the flow may produce high fluctuations if non-stationary phenomena occur. The high $\max(\mathscr{Y}(\widetilde{M}))$ and $\max(\mathscr{Y}(\widetilde{\Delta_H}))$ of 1.04% and 1.91%, respectively, may confirm the assumption.

λ[m]	$d_{0,0}[-]$	$d_{0,1}[-]$	$d_{1,0}[-]$	$d_{1,1}[-]$	$\alpha_1[°]$	$\alpha_2[°]$	$\alpha_3[°]$	$\beta_1[°]$	$\beta_2[°]$	$\beta_3[°]$	[...]
0.015	0.45	0.45	0.51	0.51	$+2$	$+2$	$+2$	$+2$	$+2$	$+2$	...
$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	...
0.04	0.49	0.49	0.55	0.55	-2	-2	-2	-2	-2	-2	...

Table 5: Extract of parameters with ranges for a run with $R = 20$ and $K = 17$

Figure 6a shows the EE plot of torque M. It is obvious that the three outlet angles β_1, β_2 and β_3 are separated from the other input parameters. According to equation (7) β_3 shows the strongest linear behavior on M. Inlet angles at shroud α_3 and hub α_1 are also separated. The other geometrical blade parameter α_2 lies close to origin like almost all other numerical parameters and thus, they are considered to have no influence on the output M. The decision whether an input is considered to have influence or not has always been done in relation to other inputs. Angles α_1 and α_3 have a similar amount of $\mathrm{mean}(\mathrm{EE}(M))$ but differ in $\mathrm{std}(\mathrm{EE}(M))$. The location close to or on the left side of the dashed line is an indication that the mean value is

zero. According to figure 6b mesh block thickness λ has a much higher mean value compared to figure 6a. An automatic optimization procedure to reduce hydraulic losses that is not able to distinguish between geometrical and numerical influences may produce an unwanted dependency on the numerical parameters, e.g. mesh block thickness λ. The mesh block thickness has a similar mean value as the blade angle β_3 and a similar standard deviation as β_1. Thus, the structured mesh block surrounding the blade has a significant influence on the physical properties. It is worth noting that performing the same run with an increased mesh block thickness λ of $0.035 \Leftrightarrow 0.04$ compared to table 5 shifts mean value and standard deviation of $\mathrm{EE}_\lambda(H_V)$ below 0.01. Figures 6a and 6b clearly show the essential tendency that the other numerical setup parameters do not vary the output in the same order as the geometry variants. The most critical numerical setup parameters are the surface normal blending factor are s and the mesh block position on shroud $d_{1,1}$. Inequality (6) is nearly true for both parameters.



(a) (b)

Figure 6: EE plot of (a) torque M and (b) head loss H_V for geometry variants

Table 6 gives the spread of an output o

$$\overline{o} = \frac{\max(Y(o)) - \min(Y(o))}{\mathrm{mean}(Y(o))}$$

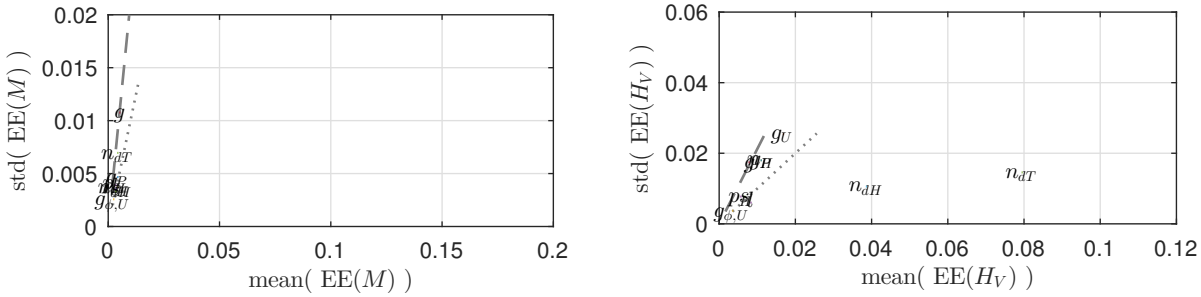
for different physical properties. All spreads are increased if geometric changes are allowed. The spread of hydraulic losses $\overline{H_V}$ is only doubled. According to the observations of figure 5b the estimated uncontrollable fluctuation is added to the table in brackets. An important point to have in mind is that the efficiency η depends on the losses H_V. The above mentioned procedure of 5b applied to η gives a corrected estimated uncontrollable fluctuation of 0.35%.

	without	with
$\overline{M}[\%]$	0.49	24.23
$\overline{H_V}[\%]$	$11.07(3.12)$	21.40
$\overline{\Delta_H}[\%]$	1.31	23.56
$\overline{\eta}[\%]$	$1.27(0.35)$	1.30

Table 6: Comparison of runs without and with geometrical modification

Figures 7a and 7b give the EE plot of torque M and Δ_H for the fixed geometrical run. The axes are equally scaled to figures 6a and 6b. For the output M all numerical or mesh parameters

are located close to the origin and consequently in the presence of allowed geometrical modifications, these inputs are considered to have no influence to the output. As expected the EE plot of Δ_H contains the linear dependent inputs n_{dH} and n_{dT}. The location of n_{dH} and n_{dT} in figure 7b is a confirmation of figures 5a and 5b. Tetrahedral division n_{dT} shows a more linear behavior than hexahedral division n_{dH}. It is equivalent to a higher slope and a lower distance of the bounded section in figure 5b compared to figure 5a.



(a) (b)

Figure 7: EE plot of (a) torque M and (b) hydraulic losses H_V without geometrical changes

The observations again show that influences by uncontrollable changes do not occur. Moreover, the observed dependency of mesh block thickness and physical properties are a controllable influence that shows in addition a linear behavior. It is again a confirmation that a comparison of samples does make sense and is not uncontrollably affected by numerical setup or mesh parameters.

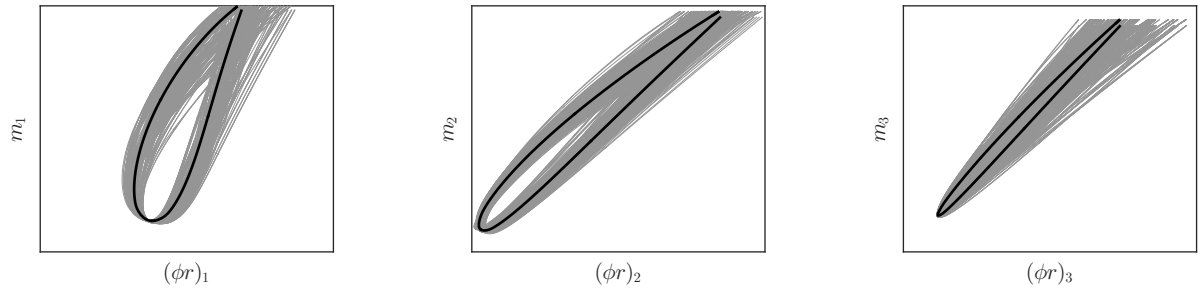
The third investigation uses the Propeller runner with parameterization given in table 2. All upper and lower bounds of the input parameters are shown in table 7. All inputs except the mesh block thickness are geometrical DDoF explained in section 3.1. Overall number of mesh elements is between 555 and 626 thousand cells and the structured part contains 95.4 thousand hexahedra ($15\% - 17\%$). The structured-unstructured coupling consists of 7.95 thousand pyramids ($1.3\% - 1.4\%$).

$\alpha_1[°]$	$\alpha_2[°]$	$\alpha_3[°]$	$\beta_1[°]$	$\beta_2[°]$	$\beta_3[°]$	$r_1[-]$	$r_2[-]$	$r_3[-]$	$t_0[\mathrm{m}]$	$t_1[\mathrm{m}]$	$\lambda[\mathrm{m}]$
$+5$	$+5$	$+5$	$+5$	$+5$	$+5$	0.3	0.3	0.3	0.06	0.0125	0.005
$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$	$\Leftrightarrow$
-5	-5	-5	-5	-5	-5	0.7	0.7	0.7	0.14	0.0375	0.015

Table 7: Parameters with ranges for a run with $R = 20$ and $K = 12$

Figures 8a, 8b and 8c show all spanwise constant blade cuts in the circumferential-meridional ϕr-m coordinate system at three different locations. The solid black line indicates the reference state that is as close as possible to the original geometry. Each gray solid line represents one sample of the run. The figures do not have equal ϕr and m axes and hence are not conformal. For this run meridional blade length H_1, H_2 and H_3 in m-direction is fixed according to table 2. For that reason a change of $r_1 - r_3$, $\alpha_1 - \alpha_3$ or $\beta_1 - \beta_3$ requires an increment or decrement of the circumferential blade length ϕr.

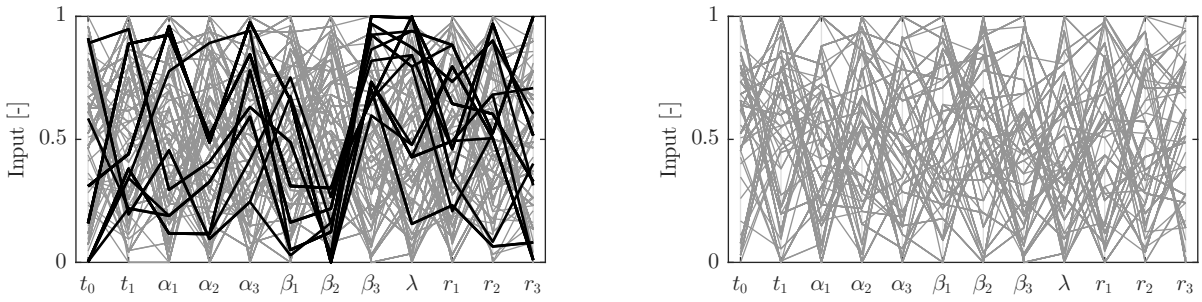
Figure 9a shows the normalized input parameter set of failed samples as black and in the background all samples with gray solid lines. In total less than 9% failed samples are produced



(a) (b) (c)

Figure 8: Spanwise constant blade cuts at (a) hub, (b) mid and (c) shroud

for this run. It is conspicuous in figure 9a that strong differences between outlet blade angles β_2 and β_3 are an indication for failing. In this case R is reduced from 20 to 14. Figure 9b gives an overview of all successful calculated and evaluated samples. It is assumed that the input parameter space is still well covered by the remaining samples shown in figure 9b.

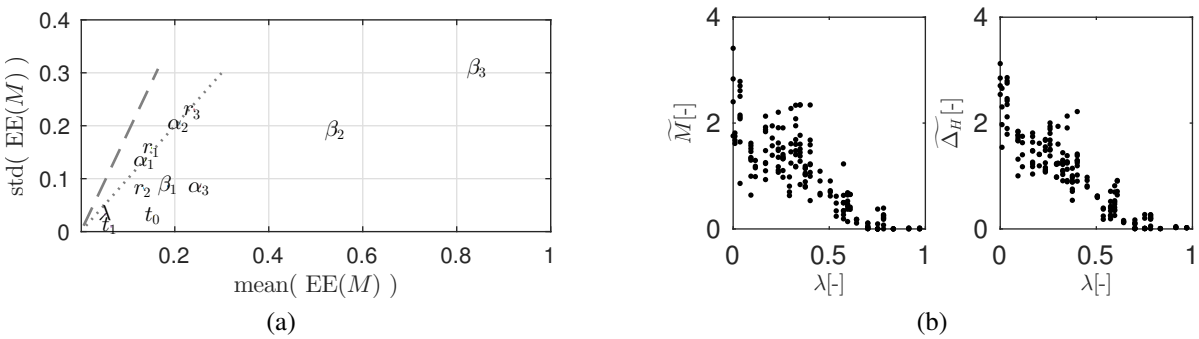


(a) (b)

Figure 9: (a) Calculated (gray) with failed (black) and (b) evaluated (gray) samples

Figure 10a visualizes the EE plot of torque M. The input parameters of figure 10a are on the right side of the dashed line which is an indication that the $\mathrm{mean(EE)}$ of the inputs are not zero. It is also clear from the figure that β_2 and β_3 are the main influencing parameters to the output and according to equation (7) do have linear effect on the output M. Moreover, inputs are clustered in three groups. Input parameters mesh block thickness λ and blade thickness at shroud t_1 are very close to each other. Other mean values of the inputs are located with a distance to origin and thus do influence M. In order to reduce mesh dependency it would be preferable to separate λ from the other inputs. Figure 10b gives $Y(\widetilde{M})$ and $Y(\widetilde{\Delta_H})$ versus normalized mesh block thickness λ. According to the figure the physical convergence criteria for this run strongly depends on the mesh block size. The average of $\mathscr{Y}(\widetilde{M})$ and $\mathscr{Y}(\widetilde{\Delta_H})$ are 4.75% and 7.02% and thus, they are very high compared to the previous observations with the Francis runner. It is important to mention that different samples of figure 10b do not only vary in the input value λ. As already mentioned a geometry modification may prevent a sample to have a stationary flow solution.

The Propeller runner is again simulated with exactly the same SA configuration except the initial range of λ of table 7 is reduced to $0.0125 \Leftrightarrow 0.015$ caused by the observations of figure 10b. The limited range corresponds to abscissa values from 0.75 to 1.00 of figure 10b. In this run R is reduced from 20 to 12 due to 8.8% failed samples. In the following the first



(a) (b)

Figure 10: (a) $\mathrm{EE}(M)$ plot of torque M and (b) fluctuations $\widetilde{M}$ and $\widetilde{\Delta_H}$ of Propeller runner

Propeller run is named "thin" and the second run "thick". According to table 8 average of $\mathscr{Y}(\widetilde{M})$ and $\mathscr{Y}(\widetilde{\Delta_H})$ are reduced to 0.30% and 0.80%. Only the 95-th percentile of $\mathscr{Y}(\widetilde{M})$ is below one percent. In addition, a too thin mesh block and consequently a flow field surrounding the blade that is discretized with more tetrahedral elements may promote smoothing effects. It is a known phenomena that tetrahedra have a more diffusive behavior than hexahedra [1]. According to table 8 mean values and percentiles of torque $\mathscr{Y}(M)$ differ significantly. The number of elements normal to the blade is not changed between the runs and thus y^+_{blade} values are increased at the blade for the thick run. Values of the dimensionless distance to the wall y^+ should be greater than 30 and smaller than 100 to model the boundary layer correctly [18]. For the Propeller runner the y^+ values on the blade are bigger than values calculated on hub and shroud. Thus overall average of y^+ is smaller than 100.

	thin				thick			
	mean	25-th	75-th	95-th	mean	25-th	75-th	95-th
$\mathscr{Y}(\widetilde{M})[\%]$	4.75	1.60	7.16	11.11	0.30	0.01	0.38	1.22
$\mathscr{Y}(\widetilde{\Delta_H})[\%]$	7.02	1.79	10.46	17.12	0.80	0.03	0.93	3.23
$\mathscr{Y}(M)[\mathrm{Nm}]$	15.6	13.2	18.8	22.5	16.7	14.4	19.8	22.8
$\mathscr{Y}(\Delta_H)[\mathrm{m}]$	2.68	2.28	3.20	3.89	2.68	2.31	3.16	3.66
$\mathscr{Y}(y^+_{blade})[-]$	98	76	118	144	131	124	138	144

Table 8: Comparison of runs with thin and thick mesh block of Propeller runner

Figures 11a, 11b and 11c give the obtained values of $\mathscr{Y}(M)$, $\mathscr{Y}(\Delta_H)$ and $\mathscr{Y}(\eta)$, respectively. Gray points represent the thin run and black points the thick run. Torque M is underestimated for the thin run and thus figure 11a gives again a graphical hint of the observations of table 8. The underestimation of M and overestimation of Δ_H leads to an underestimation of runner efficiency η, too. The plot of η in figure 11c gives the assumption that the shape of the curve obtained from the two runs remains quite similar. Hence, only an offset between both curves is assumed.

All samples of $\mathscr{Y}(M)$, $\mathscr{Y}(\Delta_H)$ and $\mathscr{Y}(\eta)$ are sorted for the thin run in ascending order for each output value separately. The ordering is then used for sorting also the thick run to get the same order of parameter sets. A loop over all elements of an output of the thick run, with the fixed ordering obtained from the thin run, checks if the current element is smaller than the next one. It gives the number of correctly ordered elements obtained in both runs. The evaluation

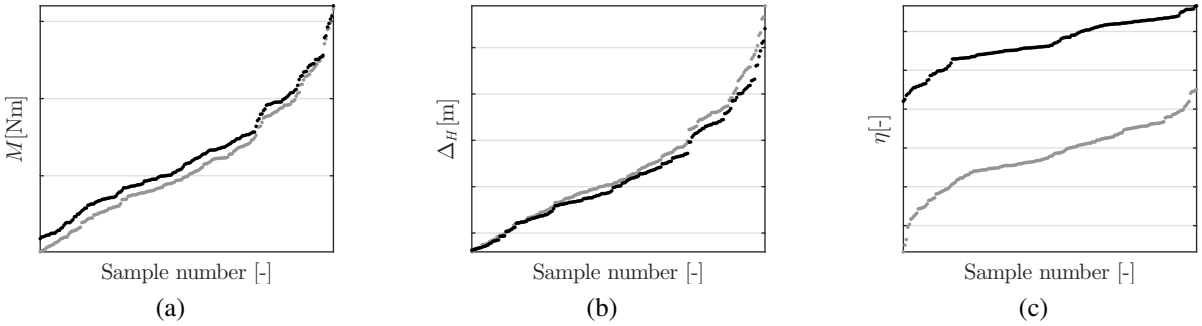


Figure 11: Samples of (a) $\mathscr{Y}(M)$, (b) $\mathscr{Y}(\Delta_H)$ and (c) $\mathscr{Y}(\eta)$ for thin (gray) and thick (black) run of Propeller runner

contains 198 samples and thus 197 checks are performed. All three outputs M, Δ_H, η and, additionally, the hydraulic losses H_V give the maximum number of successful checks. This is interpreted as a strong indication that the shapes of the curves are similar and in addition tendencies between same samples are not affected by different mesh block thicknesses.

Figure 12a, 12b and 12c shows the circumferential u_φ, radial u_r and meridional u_z mass flow averaged velocity components downstream the runner for a point close to the shroud. In order to get an indication of the flow field the velocity curves at three meridional constant positions downstream the runner are calculated and ten equidistant monitor points are used to perform the above mentioned ordering-checking procedure. For each monitor point at least 195 samples ($\approx 99\%$) are correctly ordered. An observed underestimation of the circumferential velocity u_φ for the thin run goes along with the observation of an underestimated torque M.

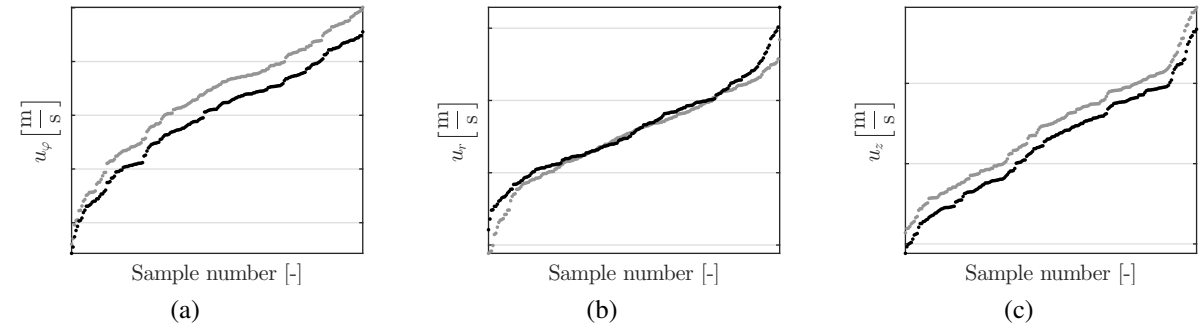


Figure 12: Cylindrical velocity components near shroud of samples (a) $\mathscr{Y}(u_\varphi)$, (b) $\mathscr{Y}(u_r)$ and (c) $\mathscr{Y}(u_z)$ for thin (gray) and thick (black) run of Propeller runner

The ordering of the inputs in figure 13 confirms the assumption of constant tendencies. Inlet angle α_2 and mean line ratio r_1 move downwards compared to figure 10a. Mesh block thickness λ has also moved in the direction of origin. All input values for the thick run are shifted to a lower std(EE(M)) value. This seems to be a consistent observation with figure 11a where torque tends to have a higher spread for the thin run. Consequently, EE have to be higher.

6 CONCLUSIONS AND FURTHER WORK

The developed framework is extended to create a hybrid mesh for different geometries. Besides existing mesh operators a free edge grading operator and a hybrid face operator is used to

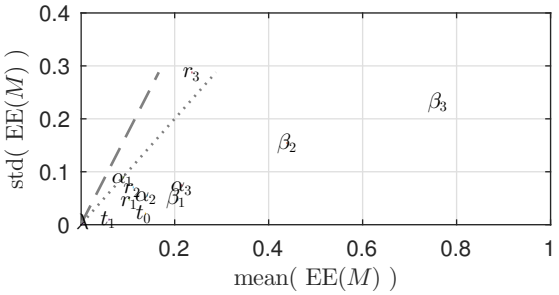


Figure 13: EE plot of torque M of Propeller runner

create the hybrid volume mesh. A modification of the original POM couples the structured with the unstructured mesh part. The prismatic boundary layer is created using a simplified advancing layer technique. Pre-processing as well as post-processing is embedded in the framework. An evaluation chain, completely implemented in the framework, is used to perform an automatic sample evaluation for different runs.

A general geometrical parameterization, that differs only in the thickness distribution and the number of DDoF, defines a typical Francis and Propeller runner. The hybrid mesh is used to perform a stationary, incompressible simulation, using periodic boundaries and a standard k-ϵ turbulence model, to obtain a solution of the flow field. Overall four runs with different modifiable input parameters of the two machines are performed. An open source sensitivity analysis toolbox is used to discover the main influences. Different input values are characterized by their effect on different output parameters. For the Francis runner it is observed that in general the effect of uncontrollable mesh variations, caused by local refinement sections or different positions of the sections in the tetrahedral mesh part, do not affect physical quantities. It is also discovered, that torque seems to be less sensitive to those variations than head or hydraulic losses. Moreover, a linear dependency is observed for the hydraulic losses. A user controlled overall increase of number of mesh elements is a critical aspect regarding the influence to the output. The use of gradient limiters for the numerical solution is more critical regarding head, but also does not affect physical properties of the same order than geometry variations. Evaluating the Propeller runner detects a linear dependency between structured mesh block thickness and fluctuations, too. A comparison of two runs with different mesh block thicknesses shows that for torque, head, runner efficiency, hydraulic losses and mass flow averaged velocities tendencies remain unchanged. In addition, the fluctuations of head are reduced significantly if mesh block thickness is increased. Only controllable, and with it avoidable, changes of mesh characteristics influence physical output values in the same range as geometry variations. The fluctuations caused by uncontrollable mesh changes for different geometry variants are negligible. Therefore, it is believed that a successful automatic optimization procedure using hybrid meshes is possible.

An interesting point is to apply the presented workflow to a whole machine and hence perform a sensitivity analysis. The simulation and evaluation of guide vanes, runner and draft tube in the presence of geometrical modifications may improve the understanding of the machine. Based on the presented observations the application of an automatic optimization procedure is of interest. Moreover, an interesting point is if a preliminary sensitivity analysis is able to reduce the amount of parameters for an automatic optimization.

REFERENCES

[1] T. J. Baker, *Mesh generation: Art or science?*, Progress in Aerospace Sciences **41** (2005), no. 1, 29 – 63.

[2] A. Ruprecht, R. Eisinger, E. Goede, *Innovative Design Environment for Hydro Turbine Components*, HYDRO (Bern), 2000.

[3] R. Eisinger, A. Ruprecht, *Automatic shape optimization of hydro turbine components based on CFD*, Seminar CFD for turbomachinery applications **6** (2001), no. 1, 101–111.

[4] A. Tismer, M. Schlipf, S. Riedelbauch, *An object-oriented approach for a highly customizable framework to design hydraulic machines*, Wasserkraftanlagen (Wien, Österreich), Innovationen und Entwicklungsbedarf für eine nachhaltige Entwicklung der Wasserkraft, vol. 18, Technische Universität Wien, Berger, Horn, 11 2014, Paper, Presentation, 213–223.

[5] MatraDatavision, *OpenCASCADE*, http://www.opencascade.com, accessed 2016-01-19.

[6] T. Bahçecioğlu, *A Simple C++ Framework to load classes from dll files in an OO way*, http://sourceforge.net/projects/pugg, accessed 2016-01-19.

[7] C. Geuzaine, J.-F. Remacle, *Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering **79** (2009), 1309 – 1331.

[8] R. V. Garimella, M. S. Shephard, *Boundary Layer Meshing for Viscous Flows in Complex Domains*, Proceedings of the 7th International Meshing Roundtable, Dearborn, Michigan, USA, October 26-28 1998, 107–118.

[9] S. J. Owen, S. A. Canann, S. Saigal, *Pyramid Elements for Maintaining Tetrahedra to Hexahedra Conformability*, Proceedings on Trends in unstructured mesh generation **220** (1997), 123–129.

[10] H. G. Weller, G. Tabor, H. Jasak, C. Fureby, *A tensorial approach to computational continuum mechanics using object-oriented techniques*, Computers in Physics **12** (1998), no. 6, 620–631.

[11] I. Abbott, A. Von Doenhoff, *Theory of Wing Sections, Including a Summary of Airfoil Data*, Dover Books on Aeronautical Engineering Series, Dover Publications, 1959.

[12] H. Nilsson, *Turbomachinery training at OFW7*, 7th OpenFOAM Workshop (Darmstadt), June 2012, Presentation.

[13] University of Genova, *Tips and Tricks in OpenFOAM*, Introductory OpenFOAM Course (Genova, Italy), February 2015, Presentation.

[14] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Halsted Press, New York, USA, 2004.

[15] F. Pianosi, F. Sarrazin, T. Wagener, *A Matlab toolbox for Global Sensitivity Analysis*, Environmental Modelling & Software **70** (2015), 80 – 85.

[16] M. D. Morris, *Factorial Sampling Plans for Preliminary Computational Experiments*, Technometrics **33** (1991), no. 2, 161–174.

[17] B. Iooss, P. Lemaître, *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, ch. A Review on Global Sensitivity Analysis Methods, 101–122, Springer US, Boston, MA, 2015.

[18] W. Rodi, *Turbulence Models and Their Application in Hydraulics*, IAHR Monographs, Taylor & Francis, 1993.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ADVANCES IN GEOMETRICAL PARAMETRIZATION AND REDUCED ORDER MODELS AND METHODS FOR COMPUTATIONAL FLUID DYNAMICS PROBLEMS IN APPLIED SCIENCES AND ENGINEERING: OVERVIEW AND PERSPECTIVES

F. Salmoiraghi[1], F. Ballarin[1], G. Corsi[1], A. Mola[1], M. Tezzele[1], and G. Rozza[1]

[1]SISSA, International School for Advanced Studies, Mathematics Area, mathLab,
Via Bonomea 265, 34136 Trieste, Italy
e-mail: {filippo.salmoiraghi,francesco.ballarin,giovanni.corsi,
andrea.mola,marco.tezzele,gianluigi.rozza}@sissa.it

Keywords: geometrical parametrization, reduced order methods, model order reduction, multiphysics, computational fluid dynamics, free-form deformation, biomedical applications, naval engineering

Abstract. *Several problems in applied sciences and engineering require reduction techniques in order to allow computational tools to be employed in the daily practice, especially in iterative procedures such as optimization or sensitivity analysis. Reduced order methods need to face increasingly complex problems in computational mechanics, especially into a multiphysics setting. Several issues should be faced: stability of the approximation, efficient treatment of nonlinearities, uniqueness or possible bifurcations of the state solutions, proper coupling between fields, as well as offline-online computing, computational savings and certification of errors as measure of accuracy. Moreover, efficient geometrical parametrization techniques should be devised to efficiently face shape optimization problems, as well as shape reconstruction and shape assimilation problems. A related aspect deals with the management of parametrized interfaces in multiphysics problems, such as fluid-structure interaction problems, and also a domain decomposition based approach for complex parametrized networks. We present some illustrative industrial and biomedical problems as examples of recent advances on methodological developments.*

1 Introduction and motivation: a synopsis

Recent research in Computational Science and Engineering aims at developing and consolidating the capabilities of computational reduction strategies for problems governed by parametrized Partial Differential Equations (PDEs). Parameters might be both physical (material properties, nondimensional coefficients such as Reynolds or Prandtl numbers, boundary conditions, forcing terms) and geometrical (i.e. quantities which characterize the shape of the domain and of the system itself). This research fits into the fields of numerical analysis and scientific computing, with a special interest in computational mechanics, and to applications in the contexts of simulation, optimization and control. In the latter cases iterative minimization procedures entailing several numerical solutions of PDEs (each time with different values of control or design variables, or different physical and geometrical scenarios) are involved, thus requiring high computational efficiency. For this reason, model order reduction techniques, such as reduced basis methods [37, 73], are mandatory to achieve this goal. With the increasing need of real time computing, Reduced Basis (RB) methods have known a remarkable development in the last decade because they make possible a strong reduction of computational times required when solving parametrized PDE problems, owing to a crucial decomposition of the computational procedures. In an offline pre-processing stage, a suitable basis is stored by solving the original problem for a set of parameter values, properly selected in an automatic and optimal way. During an online stage, for each new parameter value the solution is found as a combination of the previously computed basis functions, by means of a Galerkin projection [37, 79]. This problem has a very small size (related with the number of the selected bases, which are typically very few). The resulting procedure is not only rapid and efficient but also accurate and reliable, thanks to residual-based a posteriori error estimators. Research activities in this field have led to a significant development of the reduced order methods for many different problems, and to applications of interest in several real-life scenarios [73]. Moreover, in order to perform efficient numerical simulations in complex and variable geometric configurations, as required for instance in engineering or medical applications, reduced order methods need to be coupled with efficient parametrization techniques for curves and surfaces. Ongoing research aims at deepening the theory and the methodology of reduced order methods for problems in fluid dynamics, characterized by very different physical and temporal scales, but also complex nonlinear problems like bifurcations and instabilities. Another task is devoted in delivering ready tools for applications in naval, nautical, aerospace and mechanical engineering, as well as in medicine (fluid-structure interactions between blood flows and arterial vessels in the human cardiovascular system) [5, 6], biology (motility of cells and micro-organisms) [1, 3], porous media (groundwater flows) [54], and also geophysics (simulation of the Earth's mantle dynamics). The developed methodologies could be properly combined and coupled with novel techniques in data assimilation, uncertainty quantification for the solution of complex inverse problems arising in the multidisciplinary fields previously mentioned. The work is organized as follows: in Section 2 we give an overview about computational challenges to be tackled. In Section 3, we summarize the main idea behind model order reduction with references therein to deepen in the matter. Then, in Section 4, we show some industrial and biomedical applications and challenges in reduced order methods for computational fluid dynamics problems. Finally, in Section 5, some perspectives follow.



Figure 1: Possible examples of CFD applications.

2 Computational Science and Engineering (CSE) challenges

Simulation-based sciences is a quickly emerging field for mathematics and computational modelling. In every specialization of engineering such as aerospace, biomedical, naval, nautical and more generally mechanical field, as well in applied sciences, such as medicine, it is nowadays fundamental to run simulations to understand how a complex system behaves before actually building or operating on it. Typically, these systems depends upon a certain number of parameters, a priori unknown (e.g. to be optimized) or uncertain (e.g. due to experimental errors), so that several simulations need to be run, each time changing the value of the parameters. Present and future efforts consist in simulating complex problems in order to retain more aspects of the reality, such as multiphysics problems, as well as systems characterized by multiple spatial and temporal scales (see Figure 1). This leads to the growing demand of efficient computational tools for many query and real time computations, parametrized formulations, simulations of increasingly complex systems with uncertain scenarios.

One strategy to solve these challenges is to use the brute force and rely on high performance computing (HPC) platforms. However, for complex applications, such as the industrial ones, the solution of the problem can require a great amount of time even in the HPC framework. This leads to some limitations: in many query context, such as optimization, the computational burden can become prohibitive; for on-spot decision one can not wait for the solution of the problem more than a few minutes.

From the analysis of the physical systems surrounding us we can easily see that the behavior of a system very often changes in a smooth way with respect to the value of the parameters. Moreover, if we are interested only in some output of interest, it often happens that the system itself behaves as a filter, smoothing the output according to changes in the parameter values. This leads to the intuition that, instead of restarting from scratch for every new simulation, we can evaluate the behavior of a system in an easy way exploiting the knowledge of the solution for some (already computed) solutions. This is the rationale that drives all the reduced order models and methods formulations [20].

Thus, the master idea is to create a cooperation and a synergy between high performance computing and reduced order methods (see Figure 2). This leads to the well known offline-online splitting: during the offline stage we rely on HPC platforms to compute some very accurate solution for properly selected values of the parameters, and store them in a database of basis functions; during the online stage, we just need to combine basis functions in the database to evaluate the solution at new values of the parameter.

Although in the last decades several efforts have been spent in this field, a lot of challenges still need to be tackled in order to face and overcome limitations of the state of the art. Reduced order methodologies should be improved for more demanding applications in industrial, medical and, in general, applied sciences settings. In particular, multiphysics and coupled problems

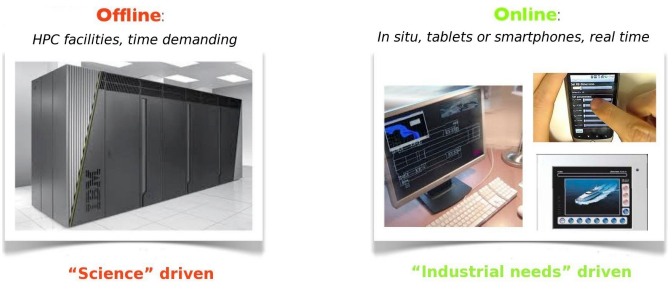


Figure 2: The main idea behind reduced order model: offline-online splitting of computational phases.

shall be considered, as well as a tighter integration with a pre-existing industrial/clinical pipeline (e.g. in data acquisition), in order to export numerical simulations and scientific computing in fields where there is still little exploitation of computational methods. The whole computational science community will need to cooperate to reach these goals, training a new generation of computational scientists. For this reason, SISSA mathLab has released the RBniCS open-source library [10] containing the implementation of several reduced order techniques, based on FEniCS [47].

3 Reduced Order Methods in a nutshell

We present a brief summary about the construction of a reduced order method (ROM) for a general problem at hand, written as:

$$\mathcal{F}(\boldsymbol{u}(\boldsymbol{\mu});\boldsymbol{\mu}) = 0, \tag{1}$$

where the parameters are represented by $\boldsymbol{\mu}$, the state solution is denoted by $\boldsymbol{u}(\boldsymbol{\mu})$ and $\mathcal{F}$ is the operator holding the state equation.

Following the paradigm shown in the previous section, let us denote as $\boldsymbol{u}^{\mathcal{N}}(\boldsymbol{\mu})$ the truth solution, obtained by solving

$$\boldsymbol{F}^{\mathcal{N}}(\boldsymbol{u}^{\mathcal{N}}(\boldsymbol{\mu});\boldsymbol{\mu}) = \boldsymbol{0}, \tag{2}$$

querying an high-fidelity solver characterized by a large number $\mathcal{N}$ of degrees of freedom.

The solver can be of very different type, depending on the application and the know-how gained in the past. For applications related reduced order models based on a finite element high-fidelity discretization see [71, 32, 88, 81, 74, 41]), as well as finite volume [34, 35, 28, 52], finite difference method [29, 17, 51], spectral element method [77, 70], extended finite element method [18, 64], boundary element method [53, 83], isogeometric analysis [53, 84] and discontinuous Galerkin methods [43, 2, 67].

Thus, after having solved (2) for some, properly selected values of the parameter, to build a reduced space, we perform a Galerkin projection over the reduced space and obtain the (general) reduced order model:

$$\boldsymbol{F}_N(\boldsymbol{u}_N(\boldsymbol{\mu});\boldsymbol{\mu}) = \boldsymbol{0}, \tag{3}$$

where $\boldsymbol{u}_N(\boldsymbol{\mu})$ is the reduced order solution and $N \ll \mathcal{N}$ is the dimension of the reduced system to be solved. Several numerical recipes are being studied in the community to guarantee the quality of the approximation and enhance the computational speedup of the evaluation, such as efficient assembly of problem operators [13], efficient treatment of nonlinearities [72], stability of the solution [80, 7], error bounds [88, 81, 62], efficient parametrization of extended systems

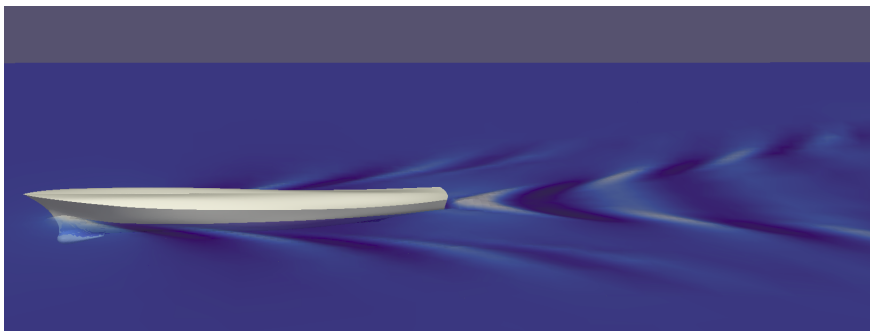


Figure 3: An example of wave pattern simulation around a hull geometry taken from a CAD file.

and complex networks [41, 40], sampling [37] and references therein. After addressing these issues for the particular problem at hand, we obtain an extremely fast system to be solved, which can be queried to obtain real-time input-output evaluation of relevant quantities of interest with an accuracy which is comparable to the high-fidelity solver.

In the following section we will provide a brief overview about some methods and applications we are treating in order to apply ROMs to some real life problems.

4 Some ROM challenges in CFD

In this section we show some, indeed very different, insights about problems arising in the industrial and clinical fields. Moreover, we also provide some ideas on the integration of these frameworks in the context of model order reduction. We first tackle the problem of ship performances prediction using real CAD geometries (Section 4.1), then the problem of the geometrical parametrization (Section 4.2) and the exploitation of isogeometric analysis (Section 4.3) in view of a complete integration of the three aspects in a single pipeline: from geometrical modification of a CAD file to the fast prediction of the wave resistance, thanks to reduced order methods. Moreover, we show new possibilities in parameter studies (Section 4.4) and a multiphysics example regarding fluid-structure interaction (Section 4.5). Finally we provide an insight in biomedical applications (Section 4.6).

4.1 Ship performances: towards the optimization of sea keeping

In the last few years, numerical fluid dynamic simulations have become an increasingly common tool in ship and yacht design. The growing computational resources available have significantly improved the quality of the results obtained. This has naturally generated, among design engineers, the desire to explore new design configurations. Researchers are then faced with the challenge to develop models of higher complexity, which can provide fast and yet accurate prediction of unsteady and nonlinear hydrodynamic loads on the hull, and reliable computations of nonlinear wave-induced ship motions.

SISSA mathLab is currently involved in the development of a model for three dimensional simulations of naval hydrodynamics based on potential flow theory (see Figure 3).

Despite the simplifying assumptions upon which the potential flow theory is based, adopting fully nonlinear free surface boundary conditions on the water free surface makes this reduced model extremely accurate in predicting both water elevation field and hull resistance [58]. In addition, the governing Laplace equation allows for a spatial discretization based on the Boundary Element Method (BEM), which only requires computational grids covering the the flow domain

boundaries. Such surface grids in three dimensions are ideal for the front tracking free surface treatment at hand, as they massively reduce the problems related to mesh deformation and generation. The solver described has been implemented in a stand-alone C++ application, which is only based on open source software libraries. In particular, the deal.II library [11, 12] is used for the spatial BEM discretization of the equation, and the IDA package of the Sundials library [38] is used to time advance the resulting Differential Algebraic Equations (DAE) system. Exploiting the relative simplicity of the surface quadrilateral grids required by BEM, an automatic mesh generation module has been developed to complement the flow solver [59]. Making use of the classes of open source C++ library OpenCASCADE [66], such module is able to import the CAD file describing the hull geometry and use it to generate the computational grid [57, 24]. After its most recent development, the model is also accounting for the rigid ship motions. To this end, the three dimensional rigid body equations of motion for the hull, have been strongly coupled to the fluid dynamic solver, so as to obtain a Fluid Structure Interaction (FSI) model able to compute hydrodynamic equilibrium position and orientation of the ship. At every time instant, the CAD model of the hull is displaced in the current position, and is used to compute the correct positioning of the water nodes in contact with the ship surface.

The model described is able to provide accurate estimations of hull resistance, sink and trim [60] at a computational cost which is significantly lower with respect to models based on the solution of Navier–Stokes equations, such as RANS or LES. For this reason, a very stimulating development of the model would be that of including the effect of non-calm sea conditions on the ship resistance and motions. On one hand, this would allow for the estimation of the added resistance component due to waves characterized by different wave lengths. On the other hand, it would open the door towards time domain manoeuvring and seakeeping simulations able to assess the ship stability and the hydrodynamic loads on the hull in presence of waves. Moreover, considering specific historic databases for local wave conditions, it would be possible to design a hull shape in order to minimize wave resistance, wave loads and motions given the specific sea conditions on the route in which the ship will operate. Naturally, this intriguing scenario requires an extremely high number of calls to the FSI solver, to test the effects of variations of both parameters related to the hull geometry and to the wave field encountered. Thus, the application of reduced order methods to the – already reduced – fully nonlinear potential fluid dynamic solver is currently under study. A first example of reduced order methods applied to BEM has been presented in [83, 53].

4.2 Efficient and accurate geometrical parametrization techniques: the free-form deformation setting

Shape optimization is nowadays a major field of interest in the design community. In this framework, we need to change some (possibly a few) geometrical parameters, compute the output of interest (cost functional) and iterate till we obtain the optimal geometry. To perform these operation, we need to rely on efficient and accurate geometrical parametrization techniques. Among many possibilities [82, 8, 5], we employ the free-form deformation (FFD) [44, 86]. This techniques is made up by three steps, summarized in Figure 4. First of all, the physical domain is mapped to a reference one. Then, design variables are varied (by moving few control points) and the deformed configuration is mapped back to the physical domain.

FFD can be applied to a CAD file or to the mesh directly. Both the approaches have some pros and cons. On the one hand, the application to the CAD allows an intuitive and easy use of the tool by engineer and designers; on the other hand, however, it requires the meshing step for each new configuration. This step can be computational expensive for some, very complex,

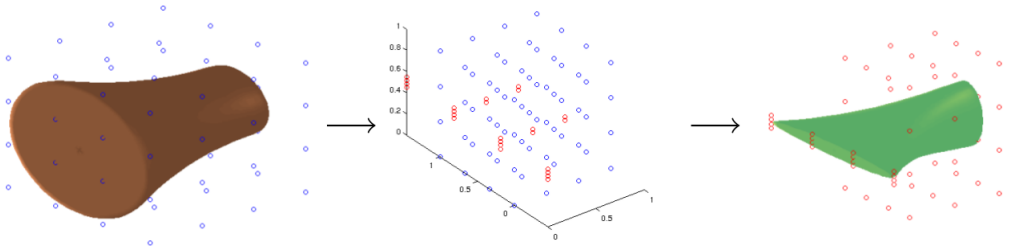


Figure 4: FFD in action: initial configuration, creation and modification of the lattice, final configuration.

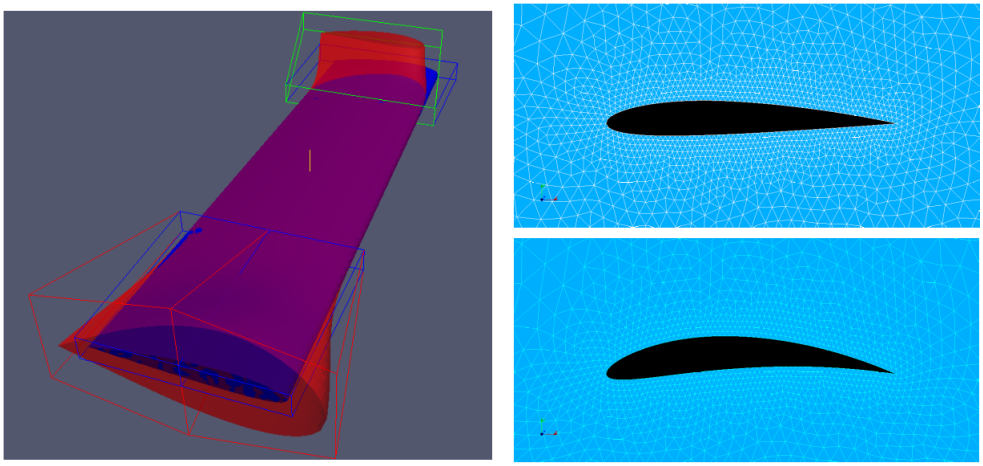


Figure 5: FFD in action: application to a CAD file (left) and a mesh (right) for a simple aeronautical geometry.

engineering problems. In Figure 5 we depict two applications of FFD corresponding to the two different approaches.

The choice of which strategy is the best depends mainly on the specific case. For example, if the high-fidelity software adopted takes as input the CAD file (and performs the mesh step automatically) and it is used as a black-box code, the FFD should be applied to the CAD file itself. Still, if one does not have any particular constraint, and can deeply modify the software, the strategy of applying FFD to the mesh is more efficient and also more suitable for the model reduction. Moreover, starting from this general paradigm, specialized shape parametrization techniques can be tailored for each application [5, 6].

4.3 From CAD to real-time evaluation: isogeometric solvers coupled with ROMs

One of the emerging fields in computational fluid dynamics is the use of isogeometric solvers. These present several features that are very attractive, especially for industrial problems. In fact, all the Computer Aided Design (CAD) geometries, the standard tool for industrial design, are nothing but NURBS patches. The idea is to employ the basis functions describing the geometry also for the analysis of the problem in a fully integrated framework [23]. Isogeometric analysis (IGA) offers also very interesting features from the geometrical parametrization point of view, thanks to its direct interface with CAD files structure. Unfortunately, real CAD files are very complex and contain patches with different parametrization of the geometry, trimmed surfaces and other features difficult to be treated efficiently. The full integration of geometry and anal-

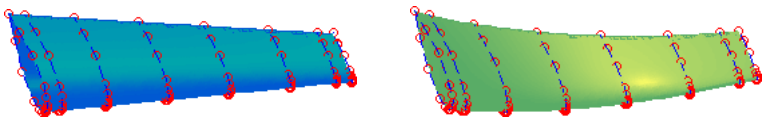


Figure 6: Isogeometric aircraft wings.

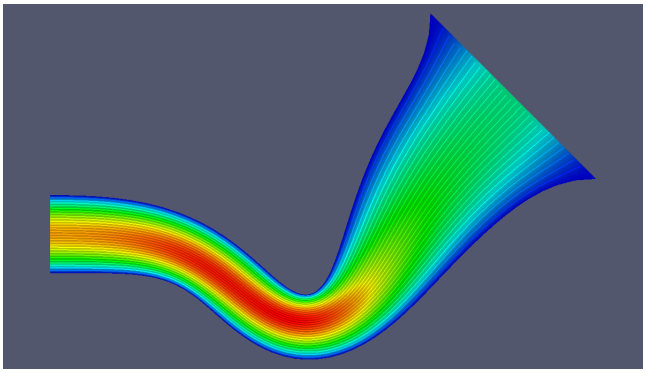


Figure 7: Example of an isogeometric exhaust gasses device.

ysis of the problem is still an open problem, although much work has been done to overcome this gap. See, for instance, Figures 6 and 7 for the case of an isogeometric aircraft wing and exhaust gasses device.

The integration of IGA with reduced order methods is a challenge that can bring together all the aspect of design procedures in an easier way: from the modification of the CAD file to the real time evaluation of the solution for the new configuration.

The work carried out in this field, is intrinsically connected to what has been proposed in section 4.2: both methods treat, from a different point of view, the direct interface with complex geometries, and these two steps can be coupled together to form a complete pipeline from CAD file to ROM solution. For some applications to the ROM-IGA coupling, see [83, 84, 53, 25].

4.4 New opportunities in parameter studies: active subspaces

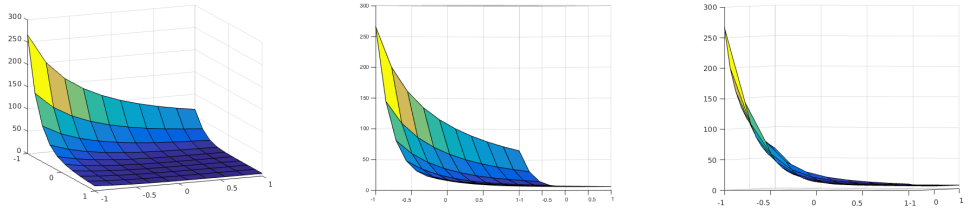
In many cases the dimension of the parametrized problem is only artificially high. To tackle this problem there are several techniques. One of them is manifold learning technique, that is a general cap for many different method (see [46]). They try to represent the manifold of the snapshots from the high fidelity solver in a reduced way, for instance unfolding the manifold along particular directions. Manifold Learning can be thought of as an attempt to generalize linear frameworks like Principal Component Analysis to be sensitive to non-linear structure in data. Though supervised variants exist, the typical manifold learning problem is unsupervised: it learns the high-dimensional structure of the data from the data itself, without the use of predetermined classifications. Engineering applications include manufacturing processes (see [45], where a manifold walking algorithm is used, and [76]) and mechanical tests [56], and structural optimization problems [75]. Among other we mention its application in natural science, for instance [3].

Another emerging idea is the active subspaces method. The active subspaces approach represents one of the emerging ideas for dimension reduction in the parameter studies. The concept was introduced by Constantine in [21] and employed in different real problems. We mention, among others, aerodynamic shape optimization [50], the parameter reduction for the HyShot II scramjet model [22] and active subspaces for integrated hydrologic model [42].

Several recent model reduction techniques attempt to compute comparable predictions at reduced costs for the same inputs [78]. However, using ROMs in a parameter study gives answers for the reduced order model instead of the actual model.

A characteristic of the active subspaces is that instead of identifying a subset of the inputs as important, they identify a set of important directions in the space of all inputs. This is done by exploiting the information given from the gradients of the output function. If the simulation's prediction does not change as the inputs move along a particular direction, then we can safely ignore that direction in the parameter study. We call subspace-based dimension reduction the dimension reduction with linear combinations of inputs. In Figure 8 it is possible to capture the main idea behind the active subspaces approach: we try to rotate the inputs domain in such a way lower dimension behavior of the output function are revealed. When you identify an active subspace for your problem of interest, then it is possible to perform different parameter studies such as response surfaces, integration, optimization and statistical inversion.

Moreover, this method seems to have great potentiality in the model reduction when the input space is very large, for instance with geometrical parameters (see section 4.2) or in shape optimization problem [27]. Moreover it can be coupled with many optimization problems linking together different but correlated outputs.



(a) Original output function (b) Intermediate rotation (c) Output function with respect to the active variable

Figure 8: Example of a bivariate output function (a), intermediate rotation of the domain (b), and the final state (c), where we can see the variation of the function along the active variable.

4.5 Towards multi-physics: an example of fluid-structure interaction problem

Coupled problems are often encountered in the industrial practice, involving different physical system each described by a set of PDEs. An example are fluid structure interaction (FSI) problems, where a mutual interaction between a fluid flow and a deformable structure occurs (see Figure 9). The structural motion or deformation caused by the fluid action is significant, so that it will in turn affect the fluid flow. Such a behavior has to be modeled with proper coupling conditions that need to be treated with adequate numerical techniques. The classic example of fluid structure interaction problem that is most often encountered in the scientific literature is the interaction between an incompressible fluid flow and an hyperelastic solid. Industrial problems, however, often require coupling to further equations to describe a complex multiphysics scenario, characterized for instance by turbulent flows (with boundary layers that need to be accurately solved), thermal analysis, multiphase flows, free surfaces or chemical reactions. Each additional coupled problem severely increases the computational time required for the simulation. Thus, even though complex multiphysics problems are the most challenging for reduced order methods, effective reduction techniques will pave the way to striking computational speedups [4].

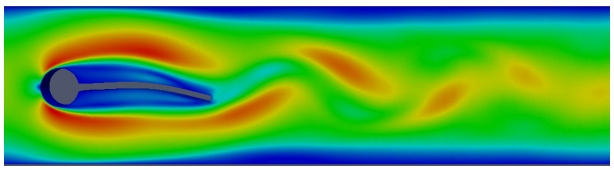


Figure 9: Snapshot taken from the simulation of the Hron and Turek FSI benchmark, proposed in [87].

As an additional source of complexity multiple regions, and thus multiple interfaces, can be present (see for example Figure 10). If multiple couplings have to be solved, it is seldom sufficient to simply concatenate them and solve each interaction sequentially. If the mutual interactions are strong, convergence will not be achieved, and a simultaneous solution of the different interfaces is required [16].

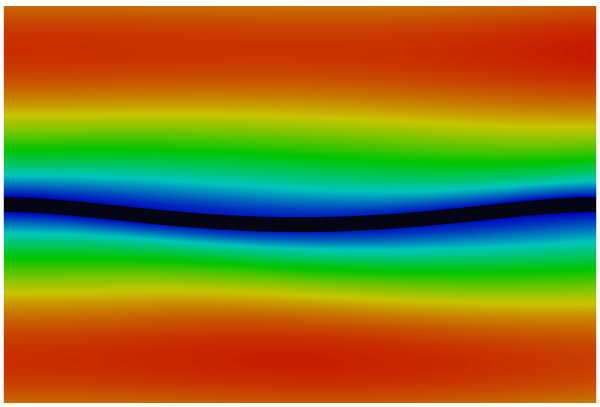


Figure 10: Example of multi-fields problem: a solid plate divides two fluid domains. In this particular example, computations were performed with a finite volume code [61].

At the high-fidelity level, the solution approaches to FSI problems can be classified in two broad categories: partitioned approach (flow and structure equations are solved in sequence) [85], and monolithic approach (the continuum is discretized and solved as a whole) [39]. Reduced order methods built on both approaches should be considered.

The partitioned approach is the most broadly used in the industrial context, because it allows the coupling of existing solvers, treating them as black boxes if the access to the code is not possible. Thus, the already optimized solvers, along with their preconditioners can be readily employed. A partitioned approach often requires to iterate the solution process within each time step of the simulation. This is due to the strength of the coupling as well as the known problem of the artificial added mass [31], that introduces further instabilities. This means that the systems of equations might have to be solved several times every time step. Thus, the computational costs can quickly become prohibitive unless efficient iterative methods to ensure a low number of iterations are employed. Quasi-Newton methods (for example [26]) have proven quite efficient to this regard. In this context, reduction techniques, prove very useful to make the iterative method as efficient and fast as possible. In particular, the reuse of information coming from the results at previous iterations (and previous time steps) has to be optimized. Reusing more iterations allows for a better approximation of the system's Jacobian, but at the expense of the conditioning. The interaction of a fluid and slender structures is an example of a case where the numerical treatment of the problem is very challenging [68]. Proper Orthogonal

Decomposition has been used to preserve the convergence rate of the iterative method when a lot of previous iteration are reused [14]. Singular Value Decomposition is also a very robust method to deal with the ill conditioning (and possible rank-deficiency) of the linear system to be solved in an iterative method [30]. In this latter case, the errors due to ill conditioning can be directly bounded by choosing how many singular values to retain from the decomposition (setting the smaller ones to a zero value).

Monolithic solvers, on the other hand, are more robust for strongly coupled problems. Generally, their performance, in terms of computational times, is superior to that of segregated solvers. Indeed, it has been shown that the monolithic approach can be competitive even for problems that feature a weak coupling, and thus would require a low number of iterations for the segregated approach [36]. For these reasons, preliminary investigation of monolithic reduced order models for FSI problem have been investigated in [9].

4.6 Biomedical applications: from patient-specific clinical data to real-time simulation

Computational fluid dynamics applications in biomedicine, in particular related to haemodynamics, are another example of fields where reduced order methods are required to propose a more widespread usage of mathematical modelling in the daily practice. Indeed, for instance in cardiovascular applications, several studies have shown that computational fluid dynamics indicators actually correlate with relevant medical questions, so that it is nowadays agreed that numerical simulation could, in principle, support medical decisions [49]. Two main requirements are usually identified by clinical partners in order for computational approaches to be really useful in the daily practice: integration with clinical data and real-time simulations. The former takes into account the personalization of the numerical simulation, starting from patient-specific clinical data on the disease (e.g. MRI or CT scans of the organ) or on flow conditions (e.g. flow rate measurements). The latter is dictated by the short time span available to take clinical decisions, especially in cases where the disease may impair the health of the patient. Under these constraints, it is natural to turn to reduced order modelling to provide fast and reliable simulations. We refer to [55, 33, 15, 63, 64, 65] for some applications of ROMs in a biomedical problems. A description of the full reduced order pipeline for the sensitivity analysis of patient-specific configurations of coronary artery bypass grafts (CABGs) is provided in [5, 6].

Moreover, a parametrized formulation of the biomedical problem allows clinicians to compare several different scenarios for what concerns both disease and medical treatment. For instance, ROMs can be a support in decision of whether to pursue medical treatment, by providing accurate indicators of the disease starting from clinical data (e.g. precise identification of degree of stenosis based on experimental pressure drop measurements in coronary arteries). Moreover, when coupled with shape parametrization techniques (Section 4.2), ROMs can also support the comparison of different surgical choices, which are usually reflected in geometrical variations of the computational domain. Figure 11 shows an example of a complete framework for a biomedical simulation. Finally, ROMs can also be used in the design of new medical treatment devices. The efficient treatment of multiphysics problems (Section 4.5) will enhance the applicability of ROM to a broader number of applications, e.g. considering reduced order simulations of fluid-structure interaction phenomena in large arteries or coupling of reduced CFD simulations to drug delivery treatments.

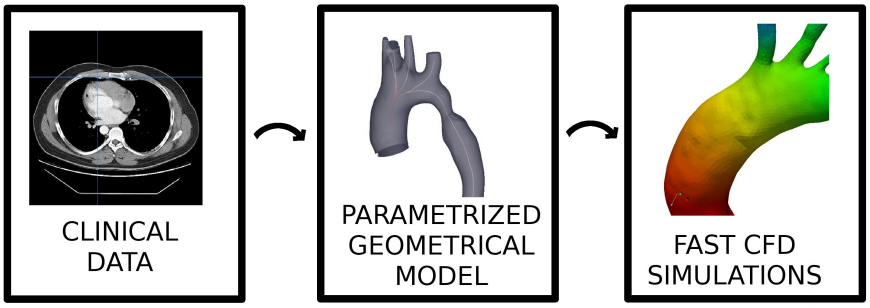


Figure 11: Example of a reduced order framework for biomedical applications.

5 Perspectives

We have provided examples of research activities in industrial and clinical projects where there is a certain degree of complexity and a certain need for the development of reduced models and methods. It is time to better integrate *Data, Modelling, Analysis, Numerics, Control, Optimization and Uncertainty Quantification* in a *new parametrized, reduced and coupled paradigm* to be able to face more and more complex problems, representing complex systems. We need to draw the attention to the fact that *Science and Industry advance with Mathematics* which is a *propeller for Innovation* and technological transfer, as shown in some examples taken from projects in naval and yachts engineering, as well as mechanical and biomedical engineering. Integration of CAD tools and geometrical parametrization, as well as medical data with geometrical reconstruction and exploration of parameter space are only few aspects of a much bigger framework, including proper integration of high order methods and low order ones (IGA-ROM is one of the possible examples).
Among other important issue under exploration, we mention the detection of flow bifurcations and the study of flow stability in CFD problems with ROM techniques [70], with also important applications in cardiovascular flows, such as the Coanda effect in Mitral valves [69]. Reduced order methods for parametrized CFD problems characterized by turbulent patterns are also under investigation for important industrial applications [48]. Last, but not least, we mention important recent advances in model order reduction for uncertainty quantification, see [19] and references therein for recent contributions and a wider framework.

Acknowledgements

The authors want to thank Prof. A. De Simone and Dr. Luca Heltai (SISSA mathLab), Dr. Haysam Telib (OPTIMAD S.r.l), Prof. Piero Triverio (University of Toronto) and Dr Laura Jimenez-Juan (Sunnybrook Hospital, Toronto) for the fruitful collaborations and discussions.
Moreover, the research has been funded by the projects Underwater Blue Efficiency and OpenViewShip, funded by Regione Friuli Venezia Giulia - programme PAR FSC 2007-2013, Fondo per lo Sviluppo e la Coesione, by the INDAM-GNCS 2015 project “Computational Reduction Strategies for CFD and Fluid-Structure Interaction Problems”, by the INDAM-GNCS 2016 projects “Tecniche di riduzione della complessità computazionale per le scienze applicate” and “Numerical methods for model order reduction of PDEs”, by the project “TRIM – Tecnologia e Ricerca Industriale per la Mobilità Marina”, CTN01-00176-163601, supported by MIUR and in partership with CNR-INSEAN, by PRIN project “Mathematical and numerical modeling of the cardiovascular system, and their clinical applications” and by European Union Funding for Research and Innovation – Horizon 2020 Program – in the framework of European Research

Council Executive Agency: H2020 ERC CoG 2015 AROMA-CFD project 681447 “Advanced Reduced Order Methods with Applications in Computational Fluid Dynamics”.

REFERENCES

[1] F. Alouges, A. DeSimone, and L. Heltai. Optimally swimming stokesian robots. *Discrete and Continuous Dynamical Systems - Series B (DCDS-B)*, 18(5):1189–1215, 2013.

[2] P. F. Antonietti, P. Pacciarini, and A. Quarteroni. A discontinuous Galerkin reduced basis element method for elliptic problems. ESAIM: M2AN, in press, 2015.

[3] M. Arroyo, L. Heltai, D. Millán, and A. DeSimone. Reverse engineering the euglenoid movement. *Proceedings of the National Academy of Sciences*, 109(44):17874–17879, 2012.

[4] M. Balajewicz and C. Farhat. Reduction of nonlinear embedded boundary models for problems with evolving interfaces. *Journal of Computational Physics*, 274:489–504, 2014.

[5] F. Ballarin. *Reduced-order models for patient-specific haemodynamics of coronary artery bypass grafts*. PhD thesis, Department of Mathematics, Politecnico di Milano, 2015. `http://hdl.handle.net/10589/102804`.

[6] F. Ballarin, E. Faggiano, S. Ippolito, A. Manzoni, A. Quarteroni, G. Rozza, and R. Scrofani. Fast simulations of patient-specific haemodynamics of coronary artery bypass grafts based on a POD–Galerkin method and a vascular shape parametrization. `http://urania.sissa.it/xmlui/handle/1963/34623`, submitted, SISSA preprint 45/2015/MATE, 2015.

[7] F. Ballarin, A. Manzoni, A. Quarteroni, and G. Rozza. Supremizer stabilization of POD–Galerkin approximation of parametrized steady incompressible Navier–Stokes equations. *International Journal for Numerical Methods in Engineering*, 102(5):1136–1161, 2015.

[8] F. Ballarin, A. Manzoni, G. Rozza, and S. Salsa. Shape optimization by Free-Form Deformation: existence results and numerical solution for Stokes flows. *Journal of Scientific Computing*, 60(3):537–563, 2014.

[9] F. Ballarin and G. Rozza. POD–Galerkin monolithic reduced order models for parametrized fluid-structure interaction problems. Submitted, SISSA preprint 07/2016/MATE, 2016.

[10] F. Ballarin, A. Sartori, and G. Rozza. RBniCS – reduced order modelling in FEniCS, 2015. `http://mathlab.sissa.it/rbnics`.

[11] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II – a general purpose object oriented finite element library. *ACM Transactions on Mathematical Software*, 33(4):24/1–24/27, 2007.

[12] W. Bangerth, T. Heister, L. Heltai, G. Kanschat, M. Kronbichler, M. Maier, B. Turcksin, and T. D. Young. The deal.II Library, Version 8.2. Technical report, 2015.

[13] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Math. Acad. Sci. Paris*, 339(9):667–672, 2004.

[14] E. Bogaers, S. Kok, and T. Franz. Strongly coupled partitioned fsi using proper orthogonal decomposition. In *8th South African Conference on Computational and Applied Mechanics (SACAM 2012)*. SAAM, 2012.

[15] M. Boulakia, E. Schenone, and J.-F. Gerbeau. Reduced-order modeling for cardiac electrophysiology. Application to parameter identification. *International Journal for Numerical Methods in Biomedical Engineering*, 28(6-7):727–744, 2012.

[16] H.-J. Bungartz, F. Lindner, M. Mehl, and B. Uekermann. A plug-and-play coupling approach for parallel multi-field simulations. *Computational Mechanics*, 55, 2015.

[17] Y. Cao, J. Zhu, Z. Luo, and I. Navon. Reduced-order modeling of the upper tropical pacific ocean model using proper orthogonal decomposition. *Computers & Mathematics with Applications*, 52(8–9):1373–1386, 2006. Variational Data Assimilation and Optimal Control.

[18] E. Chahine, P. Laborde, and Y. Renard. A reduced basis enrichment for the extended finite element method. *Mathematical Modelling of Natural Phenomena*, 4(01):88–105, 2009.

[19] P. Chen, A. Quarteroni, and G. Rozza. Multilevel and weighted reduced basis method for stochastic optimal control problems constrained by stokes equations. *Numerische Mathematik, in press*, pages 1–36, 2015.

[20] F. Chinesta, A. Huerta, G. Rozza, and K. Willcox. Model order reduction. Encyclopedia of Computational Mechanics 2016, in press, Elsevier. SISSA preprint 55/2015/MATE.

[21] P. Constantine. Active subspaces: Emerging ideas for dimension reduction in parameter studies, volume 2 of SIAM spotlights. *Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA*, 2015.

[22] P. Constantine, M. Emory, J. Larsson, and G. Iaccarino. Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet. *Journal of Computational Physics*, 302:1–20, 2015.

[23] J. A. Cottrell, T. J. Hughes, and Y. Bazilevs. *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons, 2009.

[24] F. Dassi, A. Mola, and H. Si. Curvature-Adapted Remeshing of CAD Surfaces. In *Proceedings of the 23rd International Meshing Roundtable, Procedia Engineering,*, 2014.

[25] L. Dedè, A. Quarteroni, and S. Zhu. Isogeometric analysis and proper orthogonal decomposition for parabolic problems. Technical report EPFL MATHICSE 44.2014, 2014.

[26] J. Degroote, K.-J. Bathe, and J. Vierendeels. Performance of a new partitioned procedure versus a monolithic procedure in fluid–structure interaction. *Computers and Structures*, 87:793–801, 2009.

[27] M. Diez, E. F. Campana, and F. Stern. Design-space dimensionality reduction in shape optimization by Karhunen–Loève expansion. *Computer Methods in Applied Mechanics and Engineering*, 283:1525–1544, 2015.

[28] M. Drohmann, B. Haasdonk, and M. Ohlberger. Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. *SIAM Journal on Scientific Computing*, 34(2):A937–A969, 2012.

[29] J. Du, J. Zhu, Z. Luo, and I. M. Navon. An optimizing finite difference scheme based on proper orthogonal decomposition for CVD equations. *International Journal for Numerical Methods in Biomedical Engineering*, 27(1):78–94, 2011.

[30] H.-R. Fang and Y. Saad. Two classes of multisecant methods for nonlinear acceleration. *Numerical linear algebra with applications*, 16:197–221, 2009.

[31] C. Förster, W. A. Wall, and E. Ramm. Artificial added mass instabilities in sequential staggered coupling of nonlinear structures and incompressible viscous flows. *Computational Methods in Applied Mechanical Engineering*, 196:1278–1293, 2007.

[32] M. A. Grepl and A. T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: M2AN*, 39(1):157–181, 2005.

[33] R. Guibert, K. McLeod, A. Caiazzo, T. Mansi, M. A. Fernández, M. Sermesant, X. Pennec, I. E. Vignon-Clementel, Y. Boudjemline, and J.-F. Gerbeau. Group-wise construction of reduced models for understanding and characterization of pulmonary blood flows from medical images. *Medical Image Analysis*, 18(1):63–82, 2014.

[34] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 42(2):277–302, 2008.

[35] B. Haasdonk, M. Ohlberger, and Rozza. A reduced basis method for evolution schemes with parameter-dependent explicit operators. *Electronic Transactions on Numerical Analysis*, 32:145–161, 2008.

[36] M. Heil, A. L. Hazel, and J. Boyle. Solvers for large displacement fluid-structure interaction problems: segregated versus monolithic approaches. *Computational Mechanics*, 43(1):91–101, 2008.

[37] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer Briefs in Mathematics, 2015.

[38] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396, 2005.

[39] J. Hron and S. Turek. A monolithic FEM/multigrid solver for an ALE formulation of fluid-structure interaction with applications in biomechanics. In *Fluid-Structure Interaction. Modelling, Simulation, Optimisation*, pages 146–170. Springer Berlin Heidelberg, 2006.

[40] L. Iapichino, A. Quarteroni, and G. Rozza. A reduced basis hybrid method for the coupling of parametrized domains represented by fluidic networks. *Computer Methods in Applied Mechanics and Engineering*, 221–222:63–82, 2012.

[41] L. Iapichino, A. Quarteroni, and G. Rozza. Reduced basis method and domain decomposition for elliptic problems in networks and complex parametrized geometries. *Computers & Mathematics with Applications*, 71(1):408–430, 2016.

[42] J. Jefferson, J. Gilbert, P. Constantine, and R. Maxwell. Reprint of: Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. *Computers & Geosciences*, 2015.

[43] S. Kaulmann, M. Ohlberger, and B. Haasdonk. A new local reduced basis discontinuous galerkin approach for heterogeneous multiscale problems. *Comptes Rendus Mathematique*, 349(23):1233–1238, 2011.

[44] T. Lassila and G. Rozza. Parametric free-form shape design with PDE models and reduced basis method. *Comput. Meth. Appl. Mech. Engrg.*, 199(23-24):1583–1592, 2010.

[45] G. Le Quilliec, B. Raghavan, and P. Breitkopf. A manifold learning-based reduced order model for springback shape characterization and optimization in sheet metal forming. *Computer Methods in Applied Mechanics and Engineering*, 285:621–638, 2015.

[46] J. A. Lee and M. Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.

[47] A. Logg, K.-A. Mardal, G. N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.

[48] S. Lorenzi, A. Cammi, L. Luzzi, and G. Rozza. POD-Galerkin method for finite volume approximation of Navier-Stokes and RANS equations. Submitted. SISSA preprint 2016/MATE, 2016.

[49] F. Loth, P. F. Fischer, and H. S. Bassiouny. Blood flow in end-to-side anastomoses. *Annual Review of Fluid Mechanics*, 40:367–393, 2008.

[50] T. Lukaczyk, F. Palacios, J. Alonso, and P. Constantine. Active subspaces for shape optimization. In *10th AIAA Multidisciplinary Design Optimization Conference*, 2014.

[51] Z. Luo, J. Chen, J. Zhu, R. Wang, and I. M. Navon. An optimizing reduced order FDS for the tropical pacific ocean reduced gravity model. *International Journal for Numerical Methods in Fluids*, 55(2):143–161, 2007.

[52] Z. Luo, H. Li, P. Sun, J. An, and I. M. Navon. A reduced-order finite volume element formulation based on POD method and numerical simulation for two-dimensional solute transport problems. *Mathematics and Computers in Simulation*, 89:50–68, 2013.

[53] A. Manzoni, F. Salmoiraghi, and L. Heltai. Reduced Basis Isogeometric Methods (RB-IGA) for the real-time simulation of potential flows about parametrized NACA airfoils. *Computer Methods in Applied Mechanics and Engineering*, 284:1147–1180, 2015.

[54] I. Martini, G. Rozza, and B. Haasdonk. Reduced basis approximation and a-posteriori error estimation for the coupled Stokes-Darcy system. *Advances in Computational Mathematics*, 41(5):1131–1157, 2015.

[55] K. McLeod, A. Caiazzo, M. Fernández, T. Mansi, I. Vignon-Clementel, M. Sermesant, X. Pennec, Y. Boudjemline, and J.-F. Gerbeau. Atlas-based reduced models of blood flows for fast patient-specific simulations. In O. Camara, M. Pop, K. Rhode, M. Sermesant, N. Smith, and A. Young, editors, *Statistical Atlases and Computational Models of the Heart*, volume 6364 of *Lecture Notes in Computer Science*, pages 95–104. Springer Berlin / Heidelberg, 2010.

[56] L. Meng, P. Breitkopf, B. Raghavan, G. Mauvoisin, O. Bartier, and X. Hernot. Identification of material properties using indentation test and shape manifold learning approach. *Computer Methods in Applied Mechanics and Engineering*, 297:239–257, 2015.

[57] A. Mola and L. Heltai. The step-54 tutorial program. `http://dealii.org/developer/doxygen/deal.II/step_54.html`, 2014.

[58] A. Mola, L. Heltai, and A. DeSimone. A stable and adaptive semi-lagrangian potential model for unsteady and nonlinear ship-wave interactions. *Engineering Analysis with Boundary Elements*, 37(1):128 – 143, 2013.

[59] A. Mola, L. Heltai, and A. Desimone. A fully nonlinear potential model for ship hydrodynamics directly interfaced with CAD data structures. In *The 24th International Ocean and Polar Engineering Conference*, 2014.

[60] A. Mola, L. Heltai, and A. Desimone. Ship Sinkage and Trim Predictions Based on a CAD Interfaced Fully Nonlinear Potential Model, submitted. In *The 26th International Ocean and Polar Engineering Conference*, 2016.

[61] F. Moukalled, L. Mangani, and M. Darwish. *The finite volume method in computational fluid dynamics*. Springer international publishing Switzerland, 2016.

[62] N.-C. Nguyen, G. Rozza, and A. Patera. Reduced basis approximation and a posteriori error estimation for the time-dependent viscous Burgers' equation. *Calcolo*, 46:157–185, 2009.

[63] S. Niroomandi, I. Alfaro, E. Cueto, and F. Chinesta. Real-time deformable models of non-linear tissues by model reduction techniques. *Computer Methods and Programs in Biomedicine*, 91(3):223–231, 2008.

[64] S. Niroomandi, I. Alfaro, D. Gonzlez, E. Cueto, and F. Chinesta. Real-time simulation of surgery by reduced-order modeling and X-FEM techniques. *International Journal for Numerical Methods in Biomedical Engineering*, 28(5):574–588, 2012.

[65] S. Niroomandi, D. Gonzlez, I. Alfaro, F. Bordeu, A. Leygue, E. Cueto, and F. Chinesta. Real-time simulation of biological soft tissues: a PGD approach. *International Journal for Numerical Methods in Biomedical Engineering*, 29(5):586–600, 2013.

[66] Open Cascade S.A.S., OpenCASCADE technology, 2010. `http://www.opencascade.org`.

[67] P. Pacciarini. *Discontinuous Galerkin Reduced Basis Element methods for Parametrized Partial Differential Equations in Partitioned Domains*. PhD thesis, Department of Mathematics, Politecnico di Milano, 2016.

[68] M. Paidoussis. *Fluid-Structure Interactions: Slender Structures and Axial Flow*. Academic Press, 2014.

[69] G. Pitton, A. Quaini, and G. Rozza. Computational reduction strategies for bifurcations and stability analysis in fluid-dynamics: applications to Coanda effect in cardiac flows. Submitted, 2016.

[70] G. Pitton and G. Rozza. A reduced basis method for bifurcation problems in incompressible fluid dynamics. Submitted. SISSA preprint 55/2015/MATE, 2015.

[71] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.

[72] A. Quarteroni and G. Rozza. Numerical solution of parametrized Navier–Stokes equations by reduced basis methods. *Numerical Methods for Partial Differential Equations*, 23(4):923–948, 2007.

[73] A. Quarteroni and G. Rozza. *Reduced Order Methods for Modeling and Computational Reduction*, volume 9. Springer Milano, MS&A Series, 2014.

[74] A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry*, 1(1):1–49, 2011.

[75] B. Raghavan, P. Breitkopf, Y. Tourbier, and P. Villon. Towards a space reduction approach for efficient structural shape optimization. *Structural and Multidisciplinary Optimization*, 48(5):987–1000, 2013.

[76] B. Raghavan, G. Le Quilliec, P. Breitkopf, A. Rassineux, J.-M. Roelandt, and P. Villon. Numerical assessment of springback for the deep drawing process by level set interpolation using shape manifolds. *International Journal of Material Forming*, 7(4):487–501, 2014.

[77] O. Roderick, M. Anitescu, and Y. Peet. Proper orthogonal decompositions in multifidelity uncertainty quantification of complex simulation models. *International Journal of Computer Mathematics*, 91(4):748–769, 2014.

[78] C. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *International Journal of Bifurcation and Chaos*, 15(03):997–1013, 2005.

[79] G. Rozza. Reduced basis approximation and error bounds for potential flows in parametrized geometries. *Comm. Comput. Phys.*, 9:1–48, 2011.

[80] G. Rozza, D. Huynh, and A. Manzoni. Reduced basis approximation and a posteriori error estimation for Stokes flows in parametrized geometries: roles of the inf-sup stability constants. *Numer. Math.*, 125(1):115–152, 2013.

[81] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15:1–47, 2008.

[82] G. Rozza, A. Koshakji, and A. Quarteroni. Free Form Deformation techniques applied to 3D shape optimization problems. *Communications in Applied and Industrial Mathematics*, 2013.

[83] F. Salmoiraghi. Reduced order models for potential flows past parametrized NACA airfoils based on an isogeometric boundary element method. Master thesis, Aerospace Engineering, Politecnico di Milano. `https://home.aero.polimi.it/quadrio/it/Tesi/salmoiraghi/tesi-salmoiraghi.pdf`, 2014.

[84] F. Salmoiraghi, F. Ballarin, L. Heltai, and G. Rozza. Isogeometric analysis-based reduced order modelling for incompressible linear viscous flows in parametrized shapes. Submitted. SISSA preprint 06/2016/MATE.

[85] M. Schäfer, M. Heck, and S. Yigit. An implicit partitioned method for the numerical simulation of fluid-structure interaction. In *Fluid-Structure Interaction. Modelling, Simulation, Optimisation*, pages 171–194. Springer Berlin Heidelberg, 2006.

[86] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *ACM SIGGRAPH computer graphics*, volume 20, pages 151–160. ACM, 1986.

[87] S. Turek and J. Hron. Proposal for numerical benchmarking of fluid-structure interaction between an elastic object and laminar incompressible flow. In *Fluid-Structure Interaction. Modelling, Simulation, Optimisation*, pages 371–385. Springer Berlin Heidelberg, 2006.

[88] K. Veroy and A. Patera. Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-basis a posteriori error bounds. *International Journal for Numerical Methods in Fluids*, 47(8):773–788, 2005.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NEAR-WALL BEHAVIOUR OF IMPLICIT LARGE EDDY SIMULATIONS

Ioannis W. Kokkinakis[1], **and Dimitris Drikakis**[1]

[1] University of Strathclyde
Glasgow, G1 1XQ, UK
e-mail: {ioannis.kokkinakis, dimitris.drikakis}@strath.ac.uk

Keywords: Implicit Large Eddy Simulation (ILES), turbulent boundary layers, compressible flows.

Abstract. *This paper investigates the accuracy of implicit large eddy simulations (ILES) in compressible turbulent boundary layers (TBL). ILES are conducted in conjunction with Monotonic Upstream-Centred Scheme for Conservation Laws (MUSCL) and Weighted Essentially Non-Oscillatory (WENO), ranging from 2^{nd} to 9^{th}-order. The excess artificial dissipation occurring at low Mach numbers is counter-balanced by using low Mach corrections. The study concludes that high-order ILES provide accurate predictions of TBL even on relatively coarse grids.*

1 INTRODUCTION

Design processes in engineering applications require satisfying various degrees of constraints in order to adhere to design quality standards. Through careful planning and availability of instrumentation/apparatus that conform to the necessary specifications, experimental results can indeed assist considerably during the design stage. However, most experiments are deemed cumbersome and require considerable time to plan, execute and later analyse. This is particularly true when a design cycle or optimization approach is required early on in the design phase. Furthermore, availability of diagnostic instrumentation limits the amount of useful information that can be extracted.

Computational methods offer the possibility of a high turnover of results and ample amounts of available data, allowing for a plethora of variations to the initial design of a product to be investigated. Though computational methods are increasingly becoming more popular and widely used, particularly in the early-on design phase of many engineering products, they are still treated with some caution and due care as the solutions provided can contain significant inaccuracies. These are caused mostly by the number of assumptions associated with turbulence modelling as well as the excessive numerical dissipation of schemes particularly when simulations are performed on coarse grids.

Though conducting ILES is deemed too computationally costly for use in most engineering design projects, it is only now, with the availability of evermore increasing computational power, that ILES is in its "infancy" in terms of use in wider industrial applications. Increasing the applicability of ILES requires increasing their accuracy in coarse grid simulations.

Therefore, in this paper, the accuracy of high-order, shock-capturing schemes along with any caveats, are investigated in conjunction with ILES to near-wall turbulent boundary layer (TBL) flows. The effects of numerical dissipation for schemes with accuracy ranging from 2^{nd} to 9^{th}-order are investigated both in subsonic and supersonic TBL. The results show that higher order ILES schemes are particularly well suited for simulating TBL and prove resilient to the excess artificial dissipation in low Mach flow regions.

2 METHODOLOGY

The in-house block-structured grid code CNS3D is used to solve the Navier-Stokes equations using a finite volume Godunov-type method for the convective terms. The inter-cell numerical fluxes of the convective terms are calculated by solving the Riemann problem using the reconstructed values of the conservative variables at the cell interfaces. The reconstruction stencil is a one-dimensional swept unidirectional stencil. The Riemann problem is solved using the so-called "Harten, Lax, van Leer, and (the missing) Contact" (HLLC) approximate Riemann solver [1]. Two different flux limiting approaches have been implemented in conjunction with the HLLC solver, namely the: (i) Monotone Upstream-centred Schemes for Conservation Laws (MUSCL) and (ii) Weighted-Essentially-Non-Oscillatory (WENO). In particular, the following schemes are examined:

- MUSCL piecewise linear 2^{nd} order Monotonized Central (MC) limiter [2];
- MUSCL 3^{rd} (M3) and 5^{th} (M5) order limiters [3];
- WENO 5^{th} (W5) and 9^{th} (W9) order schemes [4].

The accuracy of the above schemes, as well as of any other, can be further improved in the low subsonic region of transitional/turbulent boundary layers by implementing low-Mach corrections [5] (henceforth labelled LM). This essentially involves an additional numerical re-

construction step of the velocity vector via a progressive central differencing of the velocity components. LM corrections ensure a balanced distribution of dissipation of kinetic energy in the limit of zero Mach number, thus extending the validity of compressible flow codes to Mach numbers as low as 10^{-5}, and are particularly required for schemes providing accuracy less than 5^{th}-order [6]. The viscous terms are discretized by a second-order central scheme. The solution is advanced in time by using a five-stage (fourth-order accurate) optimal strong-stability-preserving Runge-Kutta method [7]. Further details of the numerical aspects of the code are given in [6] and [14] and references therein.

3 CHANNEL FLOW

The fully turbulent channel flow test-case has long been established as one of the major "canonical" flow problems used to perform detailed validation of numerical/computational methods. A detailed investigation of the accuracy of a number of popular numerical schemes, originally designed for shock-capturing, was carried out in [6] with respect to a weakly-compressible, turbulent channel flow. The specific objectives were: (i) to investigate the accuracy of the Monotone Upstream-centred Scheme for Conservation Laws (MUSCL) 2^{nd} to 5^{th}-order, and the Weighted Essentially Non-Oscillatory (WENO) 5^{th} to 9^{th}-order accurate flux limiter schemes against DNS data; (ii) to examine the effects of the low Mach correction on the accuracy of the MUSCL and WENO schemes; and (iii) to compare the numerical schemes with respect to their computational cost.

3.1 Case Parameters

The numerical assessment has been made using the incompressible DNS data of Moser et al. [8], corresponding to a friction Reynolds number of $Re_\tau = 395$ based on the friction velocity (u_τ) and channel half-height. Previous studies concerning compressible, turbulent channel flows have been conducted at Mach numbers of 0.4 (quasi-incompressible), however, to examine the effects of the numerical schemes with and without low Mach correction, we consider here a Mach number 0.2.

The size of the non-dimensional domain ($L_x \times L_y \times L_z$) is ($2\pi \times 2 \times \pi$) in the streamwise (x), wall normal (y) and spanwise (z) directions, respectively. In the streamwise and spanwise directions, periodic boundary conditions are employed, while in the wall normal direction an adiabatic no-slip wall condition is applied.

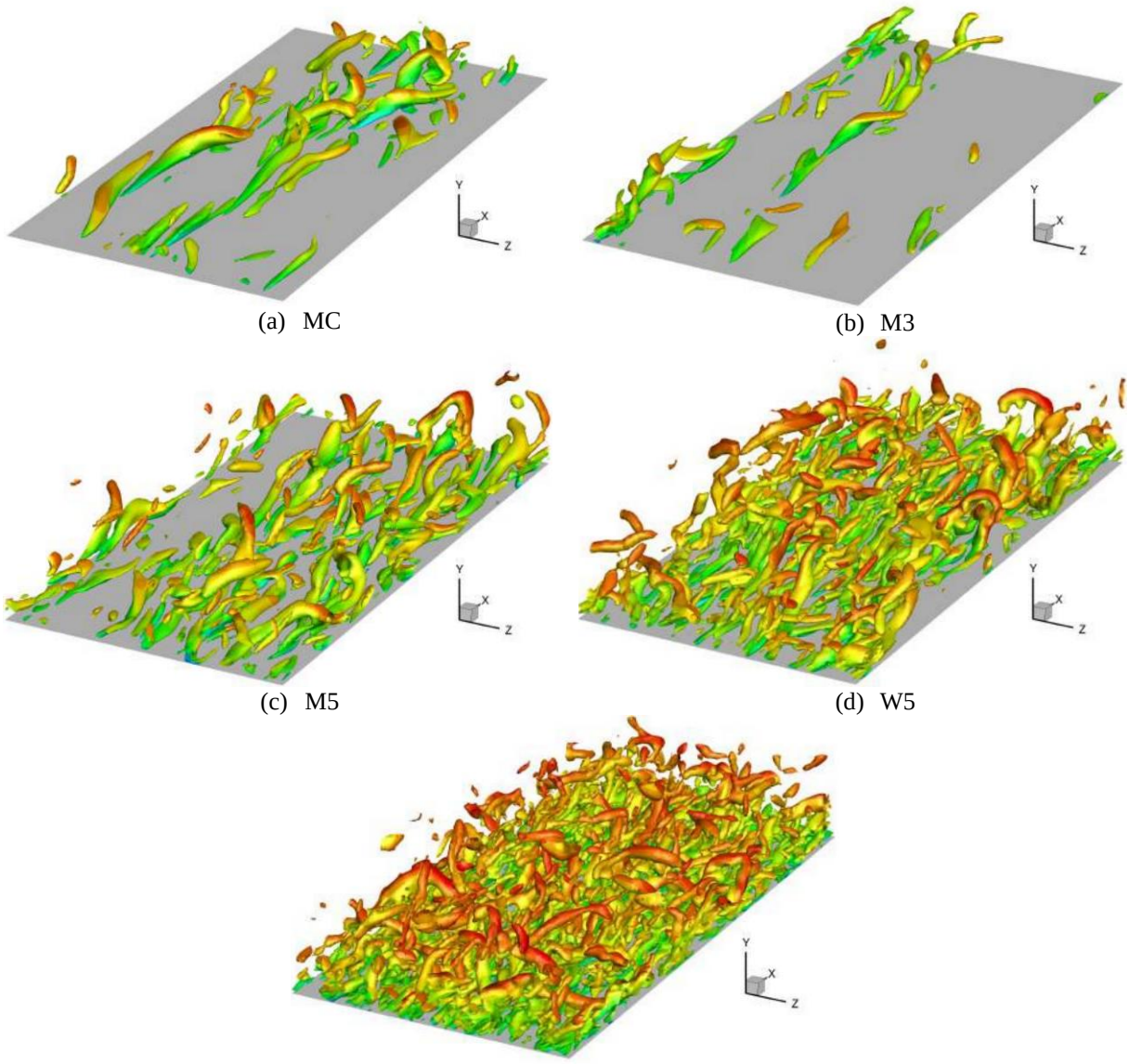
In Duan et al. [9], it was shown that many of the scaling relations used to express adiabatic compressible boundary-layer statistics in terms of incompressible boundary layers hold for non-adiabatic cases too. Wall cooling slightly enhances compressibility effects and increases the coherency of turbulent structures, however, its effects remain insignificant even for a supersonic turbulent channel flow. In the compressible DNS channel flow simulations of [10] and [11], it was shown that decreasing the wall temperature leads to higher skin friction. In the present study, the adiabatic wall condition was employed in order to examine the accuracy of the numerical methods unhindered by external heat transfer effects, thus obtaining a more meaningful comparison to the incompressible DNS. The streamwise velocity profile is given by a laminar (Poiseuille) parabolic profile with a white noise perturbation of 10% superimposed. Here we will summarize the key findings made using results only from the finest grid composed of 128^3 cells. A grid convergence study can be found in [6].

3.2 Results

The three-dimensional turbulent structures obtained from different simulations are shown by plotting the iso-surfaces of *Q*-criterion [12] in Figure 1 and Figure 2 for the schemes with-

out and with the LM correction, respectively. The *Q*-criterion is an indication of vorticity prevailing over strain and thus represents vortex cores.

The iso-surfaces reveal, in a qualitative manner, the ability of the different schemes to resolve turbulent structures. Note that for the calculation of *Q*-criterion, the velocity field is non-dimensionalized by the bulk velocity and the spatial dimensions by the channel half-height. It is clearly evident that as the order of accuracy of the reconstruction increases, more turbulent structures are resolved. For the same (5^{th}) order of accuracy, the W5 scheme resolves more turbulent structures than M5. MUSCL schemes are designed to satisfy positivity-preserving criterion in the framework of the total variation diminishing (TVD) condition, which inevitably leads to more dissipative schemes.

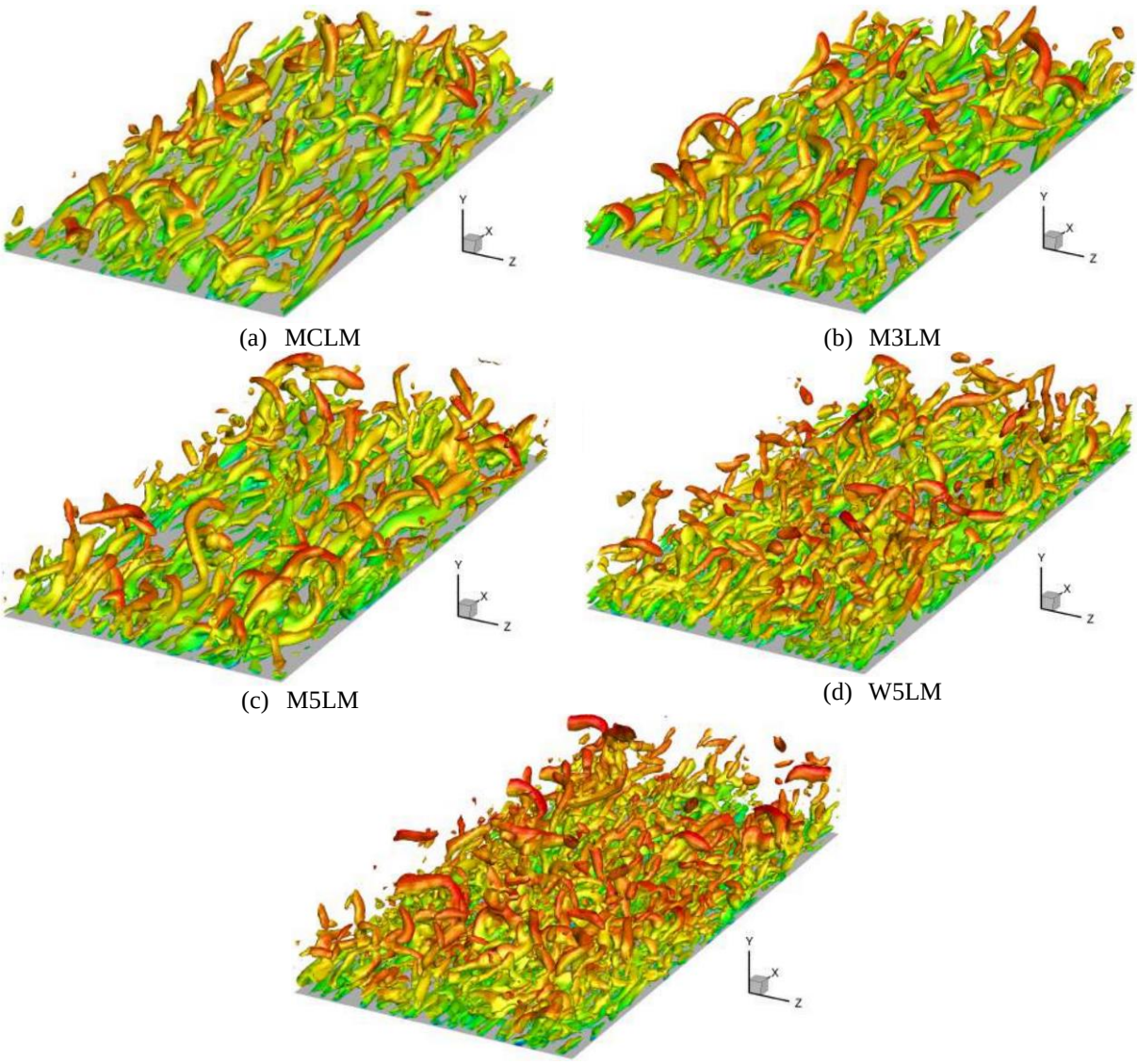


(a) MC (b) M3

(c) M5 (d) W5

(e) W9

Figure 1: *Q*-criterion iso-surfaces on 128^3 grid (iso-value=0.5 coloured by streamwise velocity).



(a) MCLM (b) M3LM

(c) M5LM (d) W5LM

(e) W9LM

Figure 2: *Q*-criterion iso-surfaces on 128^3 grid (iso-value=0.5 coloured by streamwise velocity) using the low-Mach correction [5].

Application of the LM correction results in a remarkable improvement for the 2nd, 3rd and 5th-order schemes (Figure 2), enabling much finer turbulent scales to be resolved. As the order of the scheme decreases, hence the numerical dissipation increases, the greater the improvement the LM correction provides.

The W9 scheme is found to give the most turbulent-like solutions. As the W9 scheme is the least dissipative, the addition of the LM correction does not offer any significant improvement. In fact, it is shown later that it can actually result in less accurate solutions by amplifying the dispersive errors originating from the truncation error terms (odd order terms) of a dispersive dominant scheme.

The results also reveal the mechanism by which the generated vorticity produced in the viscous layer is subsequently "ejected" due to low speed streaks into the outer boundary layer, thus creating and sustaining turbulence. This mechanism is responsible for the production of hairpin vortices that get stretched by the ambient shear. These streamwise elongated vortices have also been reported by previous (incompressible) ILES studies [13].

Figure 3 shows the velocity profiles in wall units obtained by the different numerical schemes examined. The LM correction provides a significant improvement to the MC, M3 and M5 schemes and, to a lesser degree, the W5 scheme. In contrast, the LM correction slightly deteriorates the profile of the W9. The reason is that the W9 scheme is the least dissipative and most dispersive of the schemes examined. As a result, the LM correction triggers locally entropy-violating solutions which tend to reduce the numerical dissipation of a scheme. So, if the numerical dispersion of the scheme becomes significantly greater to the dissipation, it can actually have an adverse effect on the accuracy of the scheme.

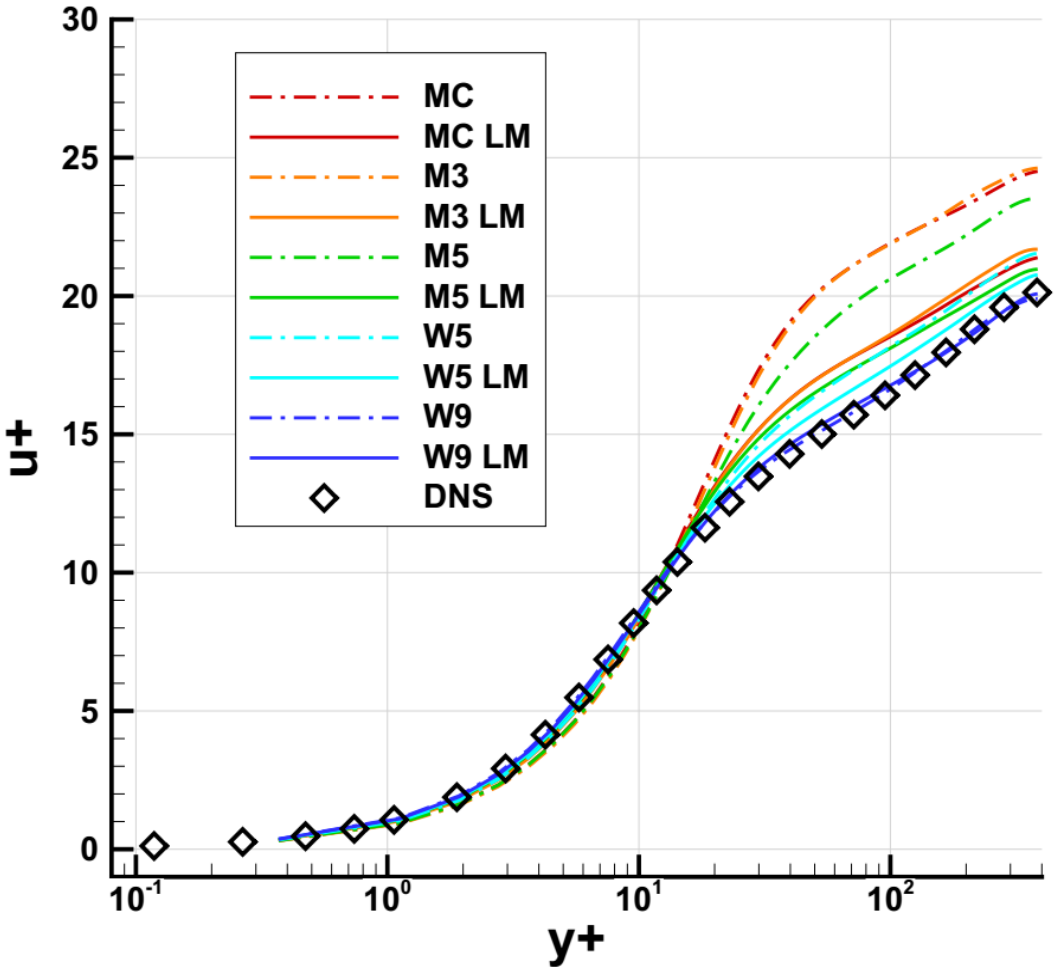


Figure 3: Velocity profiles in wall units for the different schemes.

The results obtained for the Reynolds stresses (RS) (Figure 4) reveal that the most accurate solution is obtained by the W9 scheme. The LM correction significantly improves the accuracy of all schemes apart from W9. The lower the order of accuracy of the scheme is, the greater the effect of the LM correction becomes. The W5 scheme gives better results than M5 for the $RS(u'u')$, $RS(v'v')$ and $RS(u'v')$, while similar results are obtained for $RS(v'v')$ and $RS(w'w')$ when M5 is used with LM corrections (M5LM). Overall, the W5 and W9 perform better than any of the MUSCL schemes and provide extremely accurate results.

The most noticeable result here is the significant over-prediction of the $RS(u'u')$ and under-prediction of $RS(u'v')$ by all schemes apart from W9. Additionally, the W9 was the only scheme capable of accurately resolving $RS(w'w')$ in the vicinity of y = 0.1. Decreasing the

order of accuracy leads to a gradual shift of the peak location towards the midstream. The wall-normal velocity Reynolds stress $\mathrm{RS}(\mathrm{w'w'})$ is the least accurately captured due to the unresolved turbulent scales associated with the small near-wall fluctuations. Note that the local Mach number of the flow reaches the zero limit as the no-slip wall is approached.

Regarding $\mathrm{RS}(\mathrm{u'v'})$, all schemes overall give a good agreement to the incompressible DNS. All Reynolds stress terms gradually converge to the DNS peak values in the proximity of y = 0.1 ($y^+ \approx 40$), an indication of the prevalence of turbulent production located near the end of the buffer layer.

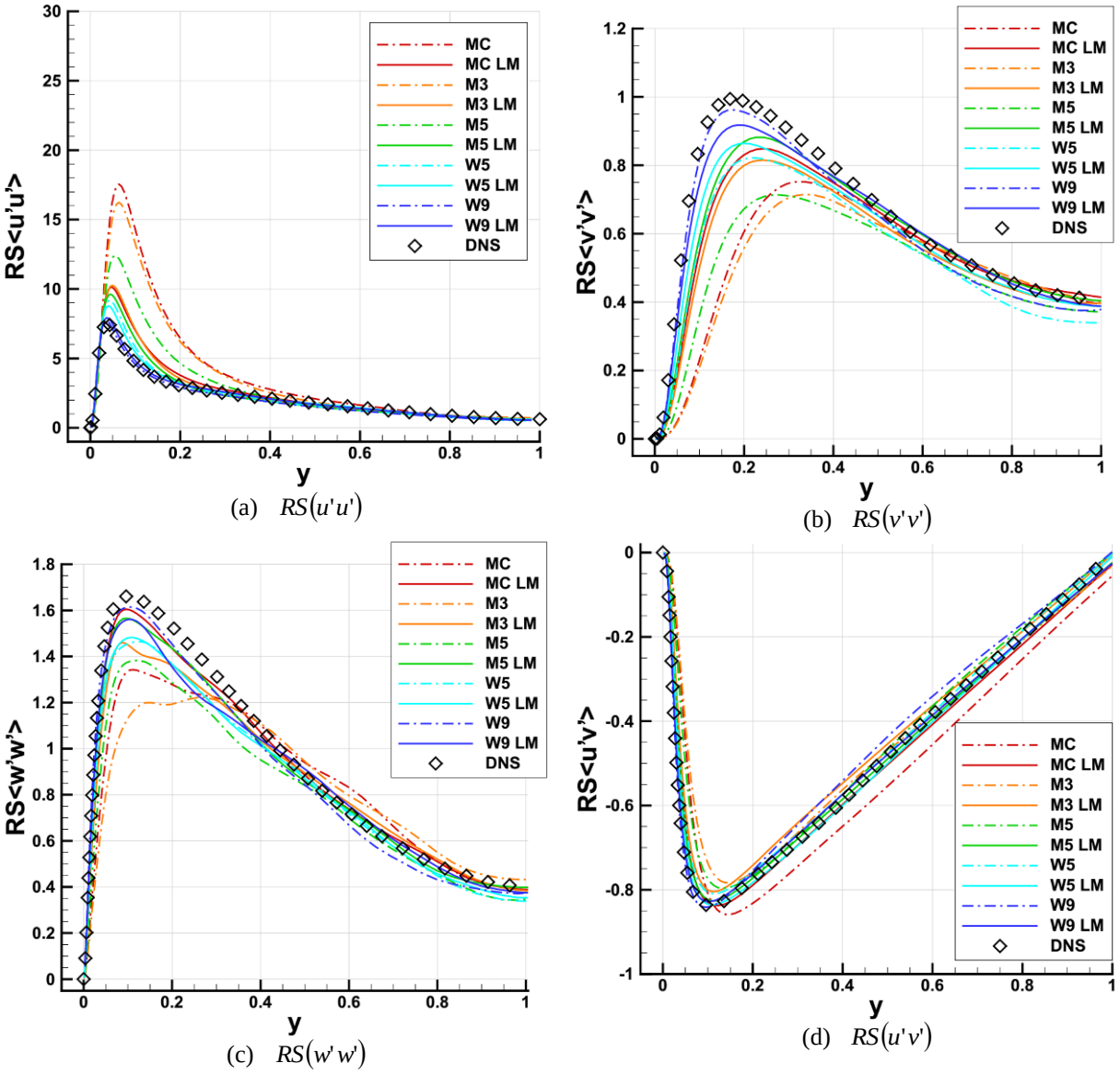


Figure 4: The streamwise velocity Reynolds stresses (RS) calculated by different ILES scheme.

4 COMPRESSION RAMP

4.1 Case parameters

Shock/Turbulent boundary layer interaction (STBLI) over a compression ramp inclined at 24° angle was investigated by Wu and Martin [16] using DNS with free-stream conditions of Mach 2.9 and Re_δ of 38,700. The presence of the inclined surface gives rise to a shock wave, which interacts with the incoming turbulent boundary layer leading to the formation of a separation bubble and a λ-shockwave. The size of the separation region is dictated by the intensity of the incoming turbulent flow and the strength of the formed shock wave. A sketch illustrating the most important physical processes that take place is given in Figure 5.

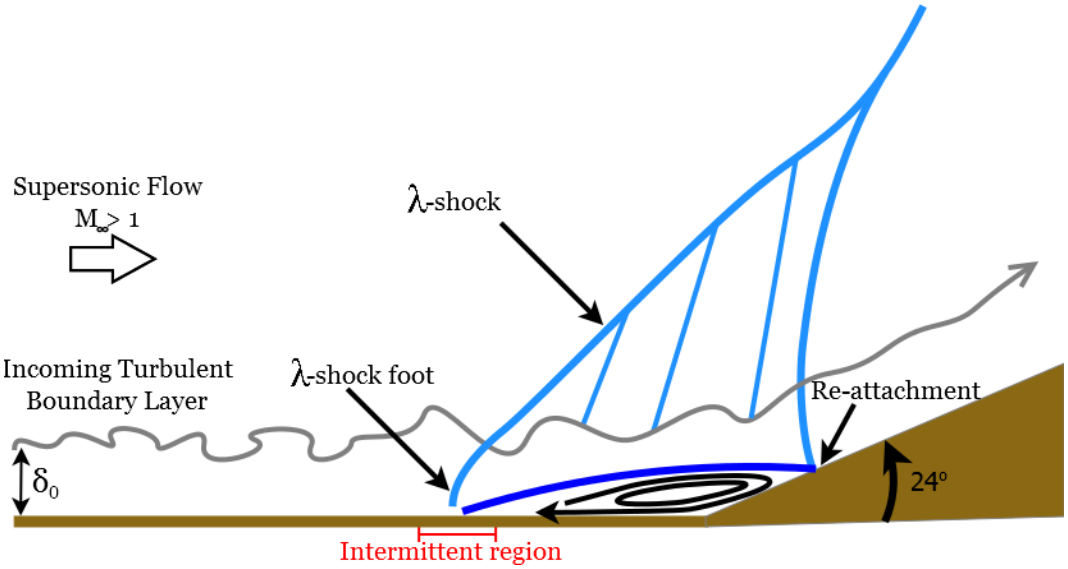


Figure 5: Schematic illustrating the flow over a compression ramp.

The physical properties and computational domain are similar to previous DNS studies and are provided in Table 1 and Table 2, respectively. Since the numerical scheme still needs to capture and resolve the synthetic inflow perturbations that lead to a turbulent flow, the length of the upstream domain is increased relatively to the DNS by approximately five incoming boundary layer heights (δ_0). The spanwise length is also increased in order to investigate possible large-scale structures that may develop in the post-shock region after re-attachment has occurred. Note that henceforth, $x/\delta_0 = 0$ will refer to where the compression corner begins.

δ_0	U_∞	T_∞	Mach	Re_{δ_0}	ρ_∞	T_w
6.4 mm	609.1 m/s	107.1 K	2.9	38,737	0.077 kg/m^3	307 K

Table 1: Flow properties.

Periodic boundary conditions are used in the spanwise (y) direction while in the wall-normal (z) direction a no-slip isothermal wall (T_W=309K) is used. A supersonic outflow condition is imposed at the outlet and upper boundary opposite to the wall. The boundary condition at the inlet requires accurately assigning a turbulent boundary layer. A synthetic turbulent digital filter approach [15], further developed in conjunction with the present numerical study, is employed to generate the incoming turbulent boundary layer data. Apart from the mean turbulent profile data which can be obtained from either previous studies or flat plate simulations,

the digital filter technique requires also knowledge of the integral length scales. Here, an integral length scale of $0.22\delta_0$ is used for the streamwise direction, whereas for the spanwise and wall-normal directions, the integral length scales chosen were $0.2\delta_0$ and $0.5\delta_0$, respectively.

	x_{max}	y_{max}	z_{max}	n_x	n_y	n_z	Δx^+	Δy^+	Δz^+
ILES	21.4	3	5	1128	120	168	41.51	51.92	3.5
DNS[16]	15.4	2.2	5	1024	128	160	31.23	28.56	0.2

Table 2: Mesh parameters for the supersonic ramp flow

4.2 Results

The LM correction is implemented only in conjunction with the 5th-order MUSCL scheme (M5LM). From the WENO schemes examined earlier only the 9th-order WENO variant (W9) will be considered here.

The streamwise distribution of the time and spatially averaged mean wall pressure (Figure 6) shows that the W9 scheme provides more accurate results than the M5LM scheme and is in excellent agreement with both the DNS [16] and experiment [17]. The M5LM scheme (green line) is characterized by a delay in the location of the separation bubble. This is associated with reduced resolution of finer scales compared to W9.

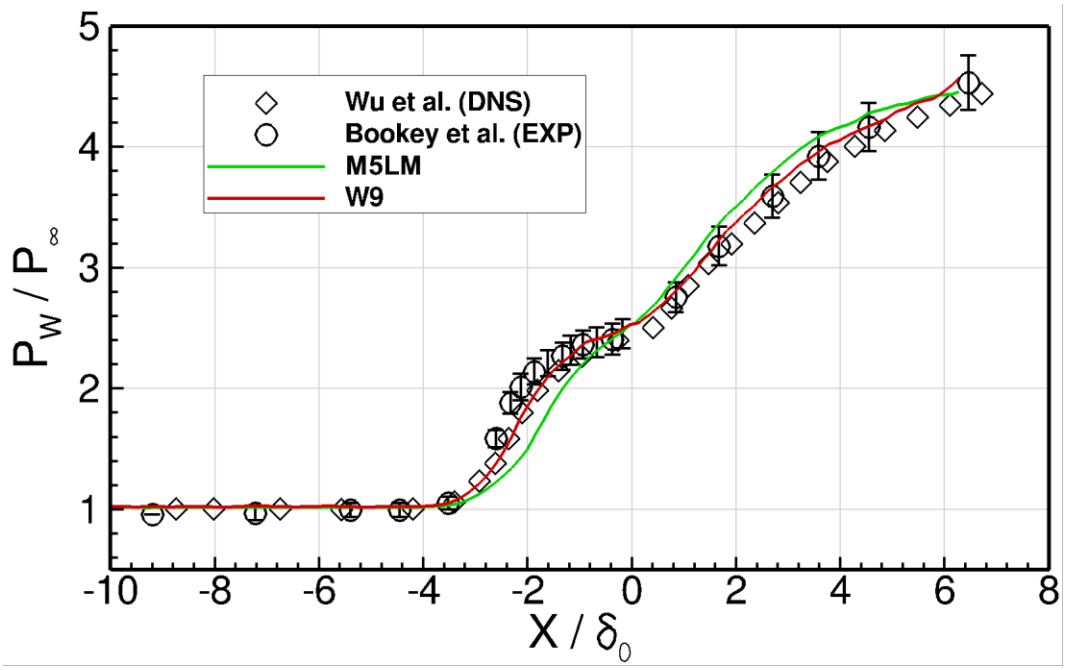
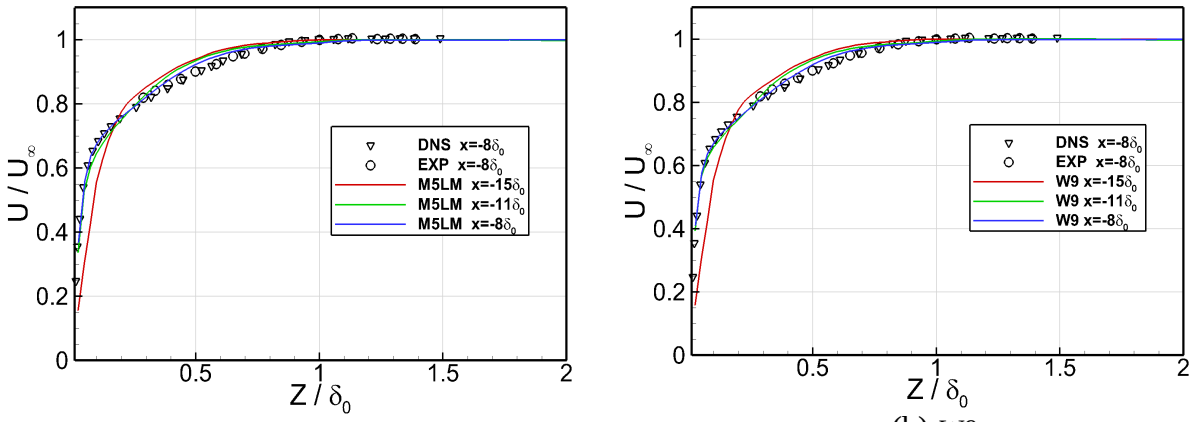


Figure 6: Mean wall pressure distribution in the streamwise direction.

The incoming mean streamwise velocity profile obtained by the digital filter technique at the inflow ($x/\delta_0 = -15$) is compared to a further two downstream locations, positioned at -11 and -8. This allows the observation of the streamwise evolution of the turbulent boundary layer created by the digital filter technique. The position at $x/\delta_0 = -8$ is a common location for both DNS and ILES upstream of the compression ramp corner and STBLI, thus the velocity profiles should closely match. Both M5LM and W9 schemes show very good agreement to the DNS and experiment (Figure 7), however the W9 results are marginally more accurate near the wall ($z/\delta_0 \approx 0.1$) as well as around $z/\delta_0 = 0.5$. Thus, the synthetic turbulent field created by the digital filter technique requires approximately $6\delta_0$ in the streamwise direction to

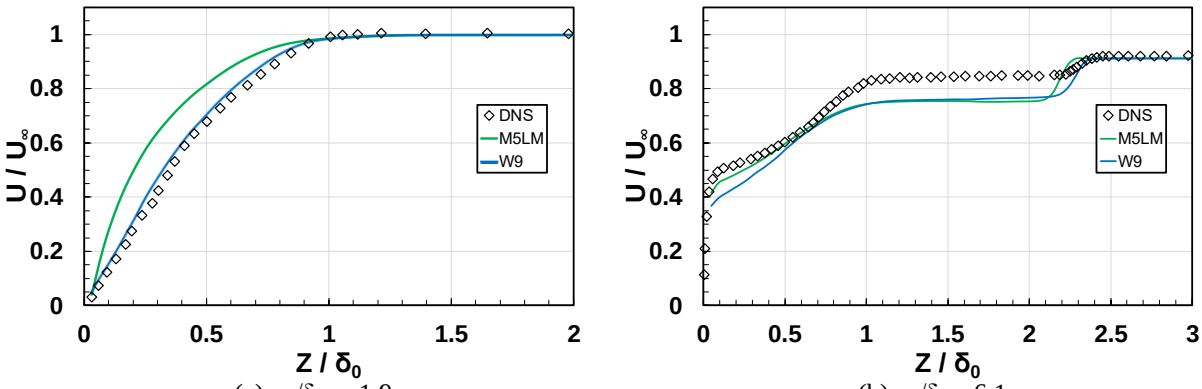
adjust. The DNS study [16] used the rescaling method developed by Xu and Martin [18] in order to generate the turbulent inflow condition, with the recycling station located at $4.5\delta_0$ downstream of the inlet.



(a) M5LM (b) W9

Figure 7: Comparison of inflow streamwise velocity to the DNS and Experiment using the W9 and M5LM schemes on the fine grid.

By the time the flow reaches the intermittent region where the foot of the λ-shock interacts with the separation bubble formed (Figure 5), there is a considerable difference in the mean streamwise boundary layer profile between the two schemes examined (Figure 8a). The W9 scheme shows the best agreement with the DNS.

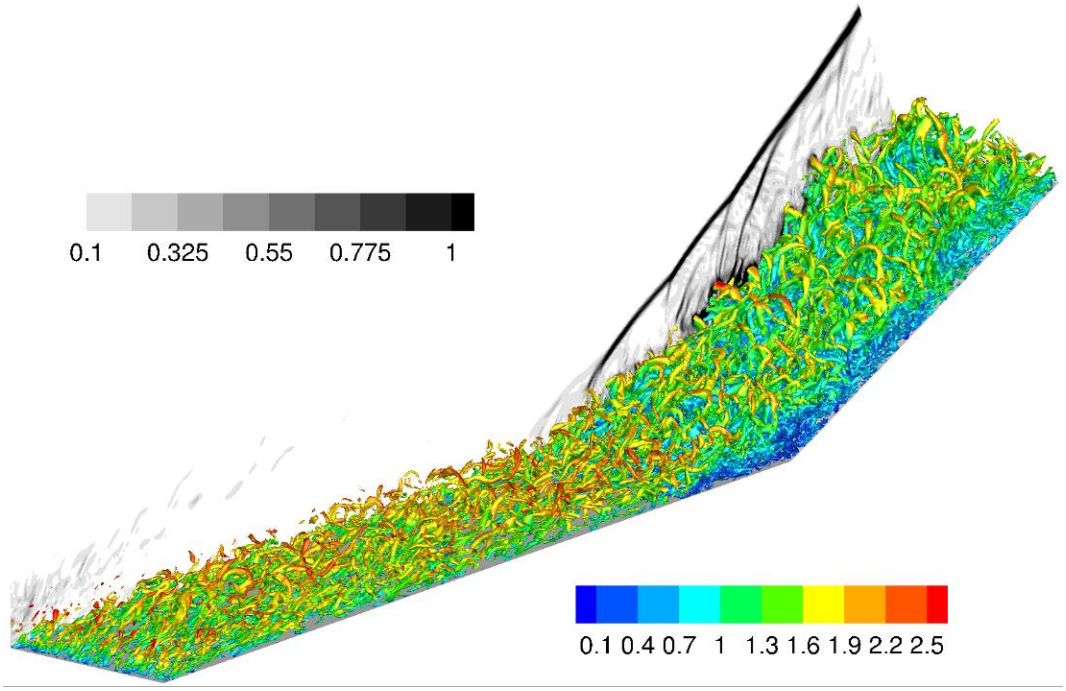


(a) $x/\delta_0 = -1.9$ (b) $x/\delta_0 = 6.1$

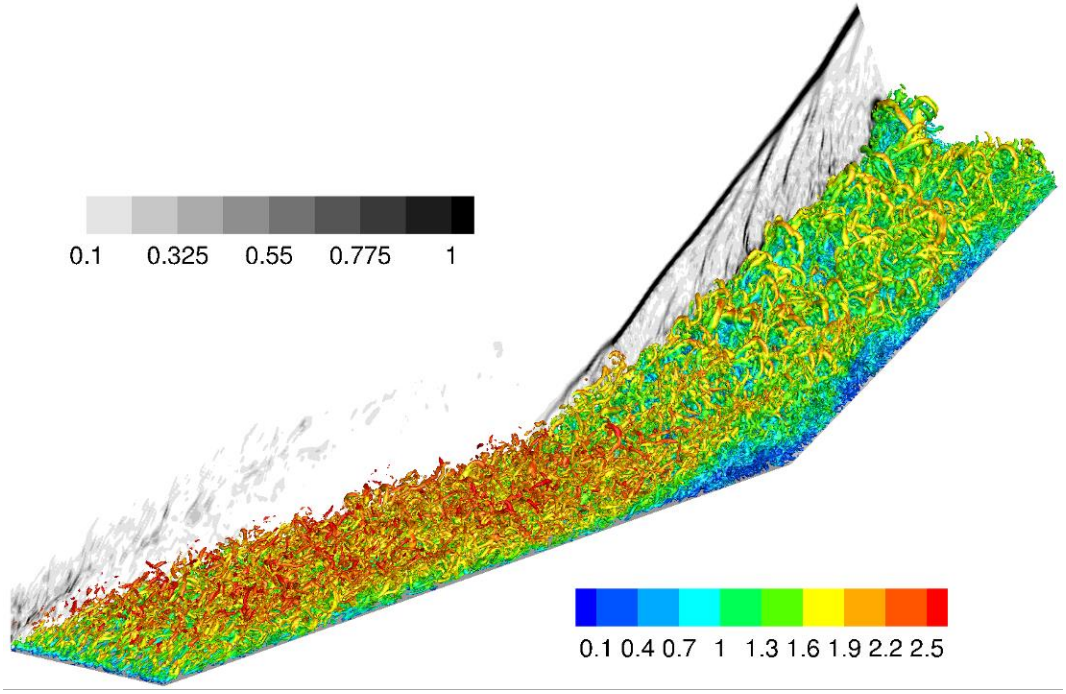
Figure 8: Streamwise velocity profiles.

Near the outlet at $x/\delta_0 = 6.1$ (Figure 8b), the velocity profiles of both M5LM and W9 show a noticeable deviation from the DNS. The freestream velocity tangential to the wall ($z/\delta_0 \geq 2.5$) reaches a value similar to DNS both for M5LM and W9 schemes. However, W9 resolves the separation bubble more accurately and consequently the SWBLI region. This is also reflected in the location where the velocity begins to reach the freestream value ($z/\delta_0 \approx 2.3$). Nonetheless, both schemes fail to accurately capture the DNS profile below this point, possibly due to the rapid coarsening of the streamwise grid resolution closer to the outlet.

Figure 9 shows a qualitative comparison between the two numerical schemes with respect to the resolved turbulent flow by plotting the iso-surfaces of the vortex cores. W9 resolves a greater number of vortices than M5LM upstream of the separation region. The differences between the two schemes becomes less apparent downstream of the STBLI.



(a) M5LM



(b) W9

Figure 9: Iso-surfaces of *Q*-criterion $Q = 2\left(\mathrm{U}_{\infty}/\delta_0\right)^2$ coloured by Mach number; grayscale contours plane of density gradient magnitude $\left|\nabla\rho\right|\left(\delta_0/\rho_{\infty}\right)$.

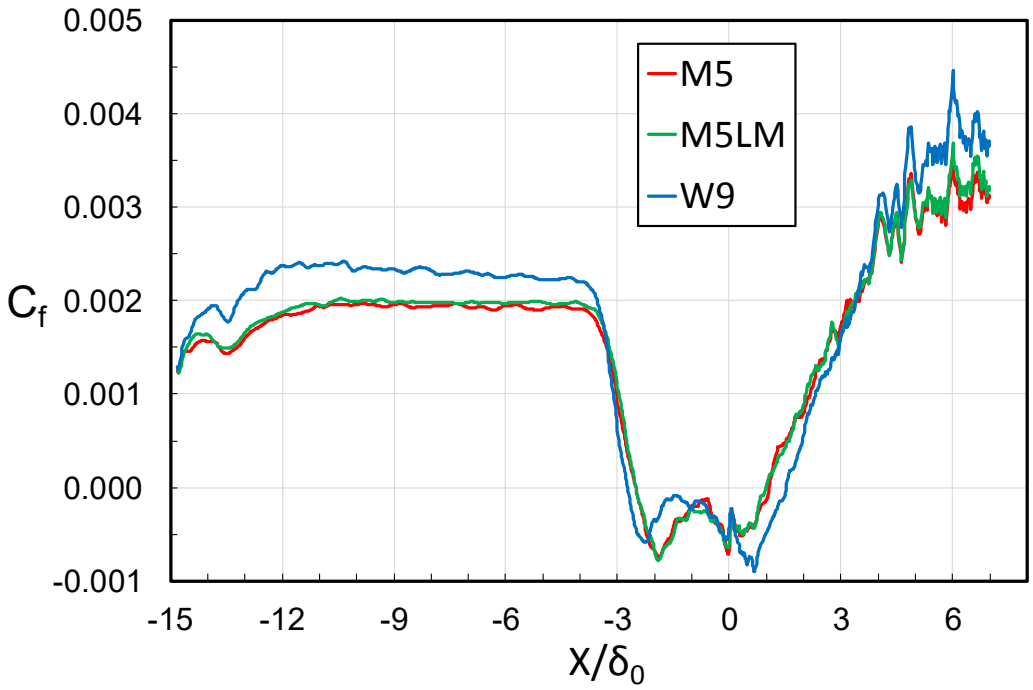


Figure 10: Skin friction coefficient distribution.

Although the flow is supersonic in the freestream, the boundary layer remains subsonic in the first 5 to 10% of the incoming boundary layer height, i.e. approximately 20 computational cells for a relatively fine (LES) grid. The subsonic region of the separation bubble can reach a size of $0.6\delta_0$ and any loss of accuracy here could potentially lead to significant errors. However, the streamwise distribution of the coefficient of friction (C_f) given in Figure 10 suggests that the prediction of separation bubble is primarily influenced by the resolved upstream turbulent boundary layer.

5 CONCLUSIONS

- Low-Mach corrections [5] significantly improve the accuracy of lower order schemes in ILES of subsonic TBL but have little effect on STBLI predictions.
- Low-Mach corrections have a greater effect when they are used in conjunction with lower-order schemes (less or equal to 5th-order).
- Overall, ILES using W9 give very promising results both for subsonic and supersonic turbulent boundary layers.

ACKNOWLEDGEMENTS

Part of this work was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-14-1-0224. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

REFERENCES

[1] Toro, E. F. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction, 3rd Edition*. Springer, 2009.

[2] Van Leer, B., Towards the ultimate conservative difference scheme III. Upstream-centered finite-difference schemes for ideal compressible flow. *Journal of Computational Physics*, **23** (3), 263-275, 1977.

[3] Kim, K.H., Kim, C., Accurate, efficient and monotonic numerical methods for multi-dimensional compressible flows: Part II: Multidimensional limiting process. *Journal of Computational Physics*, **208** (2), 570-615, 2005.

[4] Balsara, D.S., Shu, C.-W., Monotonicity preserving weighted essentially non-oscillatory schemes with increasingly high order of accuracy. *Journal of Computational Physics*, **160** (2), 405-452, 2000.

[5] Thornber, B., Mosedale, A., Drikakis, D., Youngs, D., Williams, R.J.R, An Improved Reconstruction Method for Compressible Flows with Low Mach Number Features. *Journal of Computational Physics*, **227,** 4873-4894, 2008.

[6] Kokkinakis, I.W., Drikakis, D., Implicit Large Eddy Simulation of weakly-compressible turbulent channel flow. *Computer Methods in Applied Mechanics and Engineerin*g, **287,** 229-261, 2015.

[7] Spiteri, R.J. and Ruuth, S.J., A New Class of Optimal High-Order Strong-Stability-Preserving Time Discretization Methods. *SIAM Journal on Numerical Analysis*, **40** (2), 469-491, 2002.

[8] Moser, R. D., Kim, J., Mansour, N. N., Direct numerical simulation of turbulent channel flow up to Re_τ=590. *Physics of Fluids*, **11** (4), 943-945, 1999.

[9] Duan, L., Beekman, I., Martin, M.P., Direct numerical simulation of hypersonic turbulent boundary layers. Part 2. Effect of wall temperature. *Journal of Fluid Mechanics*, **655**, 419-445, 2010.

[10] Coleman, G.N., Kim, J., Moser, R.D., A numerical study of turbulent supersonic isothermal-wall channel flow. *Journal of Fluid Mechanics*, **305**, 159-183, 1995.

[11] Morinishi, Y., Tamano, S., Nakabayashi, K., Direct numerical simulation of compressible turbulent channel flow between adiabatic and isothermal walls. *Journal of Fluid Mechanics*, **502**, 273-308, 2004.

[12] Hunt, J.C.R., Wray, A., Moin, P., Eddies, stream, and convergence zones in turbulent flows. *Center for Turbulence Research*, Report **CTR-S88**, 1988.

[13] Fureby, C., Grinstein, F.F., Large eddy simulation of High-Reynolds-number free and wall-bounded flows. *Journal of Computational Physics*, **181** (1), 68-97, 2002.

[14] Tsoutsanis, P., Kokkinakis, I.W., Konozsy, L., Drikakis, D., Williams, R.J.R., Youngs, D., Comparison of structured and unstructured-grid, compressible and incompressible methods using the vortex pairing problem. *Computer Methods in Applied Mechanics and Engineering*, **293**, 207-231, 2015.

[15] Rana, Z.A., Thornber, B., Drikakis, D., Transverse jet injection into a supersonic turbulent cross-flow. *Physics of Fluids*, **23** (4), 2011.

[16] Wu, M., Martin, M.P., Direct Numerical Simulation of Supersonic Turbulent Boundary Layer over a Compression Ramp. *AIAA Journal*, **45** (4), 879—889, 2007.

[17] Bookey, P. B., Wyckham, C., Smits, A. J., Martin, M. P., New Experimental Data of STBLI at DNS/LES Accessible Reynolds Numbers. *43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, January 10-13, 2005.

[18] Xu, S., Martin, M. P., Assessment of Inflow Boundary Conditions for Compressible Turbulent Boundary Layers. *Physics of Fluids*, **16 (**7), 2623—2639, 2004.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

VALIDATION AND VERIFICATION OF A 2D LATTICE BOLTZMANN SOLVER FOR INCOMPRESSIBLE FLUID FLOW

Tamás István Józsa[1,3], Máté Szőke[2,3], Tom-Robin Teschner[3], László Könözsy[3*], Irene Moulitsas[3]

[1]University of Edinburgh
Institute for Energy Systems
Alrick Building, Max Born Crescent, Edinburgh, Scotland, United Kingdom, EH9 3BF
e-mail: t.jozsa@ed.ac.uk

[2]University of Bristol
Mechanical Engineering
Queen's Building, University Walk, Bristol, United Kingdom, BS8 1TR
e-mail: m.szoke@bristol.ac.uk

[3]Cranfield University
Centre for Fluid Mechanics and Computational Science
Building 52, Cranfield, Bedfordshire, United Kingdom, MK43 0AL
e-mail: {t.teschner, laszlo.konozsy, i.moulitsas}@cranfield.ac.uk

Keywords: Lattice Boltzmann Method, Validation, Verification, Computational Fluid Dynamics, LBM, CFD, CUDA, GPU, UPC, PGAS.

Abstract. *The lattice Boltzmann method (LBM) is becoming increasingly popular in the fluid mechanics society because it provides a relatively easy implementation for an incompressible fluid flow solver. Furthermore the particle based LBM can be applied in microscale flows where the continuum based Navier-Stokes solvers fail. Here we present the validation and verification of a two-dimensional in-house lattice Boltzmann solver with two different collision models, namely the BGKW and the MRT models [1]. Five different cases were studied, namely: (i) a channel flow was investigated, the results were compared to the analytical solution, and the convergence properties of the collision models were determined; (ii) the lid-driven cavity problem was examined [2] and the flow features and the velocity profiles were compared to existing simulation results at three different Reynolds number; (iii) the flow in a backward-facing step geometry was validated against experimental data [3]; (iv) the flow in a sudden expansion geometry was compared to experimental data at two different Reynolds numbers [4]; and finally (v) the flow around a cylinder was studied at higher Reynolds number in the turbulent regime. The first four test cases showed that both the BGKW and the MRT models were capable of giving qualitatively and quantitatively good results for these laminar flow cases. The simulations around a cylinder highlighted that the BGKW model becomes unstable for high Reynolds numbers but the MRT model still remains suitable to capture the turbulent von Kármán vortex street. The in-house LBM code has been developed in C and has also been parallelised for GPU architectures using CUDA [5] and for CPU architectures using the Partitioned Global Address Space model with UPC [6].*

1 Introduction

McNamara and Zanetti introduced the lattice Boltzmann Method (LBM) as a cure to remove the statistical noise encountered in its predecessor, the Lattice Gas Cellular Automata (LGCA) [7]. At the time, the impact of the LBM on research related to fluid dynamics could not have been foreseen. Almost three decades after its introduction, the LBM has enjoyed a wide range of applications including multiphase, porous media, magnetohydrodynamics, acoustics, medical and turbulent flows [8, 9] and of course the automotive industry [10, 11, 12]. The fast grown popularity stems from its ease of implementation and treatment of complex domains, including multi-physics problems which otherwise would require rather sophisticated numerical tools. At its core, it represents a discretised version of the Boltzmann equation and can be shown, via the Chapman–Enskog expansion, to adhere to the incompressible Navier–Stokes equations [13]. Therefore, the LBM exhibits favourable traits of a particle method, such as the ease of implementation for complex systems, while the modified Boltzmann equation represents a single, linear transport equation that only needs 10-20% of the computational time required by a conventional Navier–Stokes solver [14].

In this work we present the results we have obtained from a two-dimensional lattice Boltzmann solver with two different collision models, namely the BGKW and the MRT models [1]. We validate our solver against reference data for (i) a channel flow where the results are compared to the analytical solution, and the convergence properties of the collision models are determined; (ii) the lid-driven cavity problem [2] where the flow features and the velocity profiles are compared to existing simulation results at three different Reynolds number; (iii) the flow in a backward-facing step geometry is validated against experimental data [3]; (iv) the flow in a sudden expansion geometry is compared to experimental data at two different Reynolds numbers [4]; and finally (v) the flow around a cylinder is studied at higher Reynolds number in the turbulent regime.

With the current transistor technology plateauing, focus has shifted towards parallel scientific computing. A number of parallel software and hardware environments have emerged and combinations of these have been proposed in an attempt to tackle large scale problems accurately and efficiently [15]. Our LBM solver has further been parallelised using two different parallel computing environments and paradigms. The first approach relies of graphical processing units (GPUs) and utilises the Compute Unified Device Architecture (CUDA) framework [5] which has recently found interest in fluid dynamics, see for example Kuznik *et al.* [16]. The second approach relies on traditional central processing units and utilises the Unified Parallel C (UPC) framework [6]. UPC is based on the Partitioned Global Address Space (PGAS) programming model that aims to simplify the programming by abstracting the memory address space, hence allowing the programmer to concentrate on the solution of the underlying scientific problem. We further discuss the scalability of our solver for both parallelisation strategies followed.

2 Governing equations and computational methodology

To discretise the Boltzmann equation, particles are first represented by a Density Distribution Function (DDF) which is then restricted to propagate only in certain directions. These directions are determined by the chosen speed model. For the purposes of our study we have employed the D2Q9 speed model, which is a 2–dimensional model with 9 velocity directions. This model, presented in Figure 1, is a common choice for two-dimensional problems. The discretised Boltzmann transport equation is then formulated as

$$\frac{\partial f_i(\mathbf{x},t)}{\partial t} + (\mathbf{c}_i \cdot \nabla) f_i(\mathbf{x},t) = \Omega_i(f_i) \tag{1}$$

where i is the index of the restricted streaming direction and $f_i(\mathbf{x},t)$ is the DDF depending both on space and time. The discrete velocities, denoted by $\mathbf{c}_i$, take binary values and are defined by

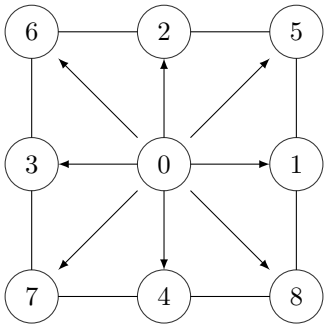


Figure 1: The D2Q9 speed model within a square lattice

$$\mathbf{c}_i = \begin{cases} 0, & i = 0 \\ \left(\cos\left[(i-1)\,\pi/2\right], \sin\left[(i-1)\,\pi/2\right]\right), & i = 1,2,3,4 \\ \sqrt{2}\left(\cos\left[(i-5)\,\pi/2+\pi/4\right], \sin\left[(i-5)\,\pi/2+\pi/4\right]\right). & i = 5,6,7,8 \end{cases} \tag{2}$$

Hence the LBM can be separated into two steps: *collision* and *streaming*. The advection is represented by $(\mathbf{c}_i \cdot \nabla) f_i(\mathbf{x}, t)$ where the DDF is streaming to its neighbouring lattice sites. The collision is represented by $\Omega_i(f_i)$ and it is, unlike in the LGCA where a explicit collision treatment is possible, not straightforward. The collision can, at best, be approximated and different models have been proposed. We will limit ourselves to the Single Relaxation Time (SRT) scheme of Bhatnagar, Gross and Krook [17] (also referred to as the BGK or BGKW scheme due to the contribution of Welander [18]) and the Multi Relaxation Time (MRT) scheme of D'Humieres [19, 20]. Boundary conditions must be formulated for the DDF which are not readily available for incoming directions. We have followed the approach of Zou and He [21] which can be written in a compact form as

$$f_c = \mathrm{opp}(f_c) \pm \frac{2}{3}\varrho u_\perp, \tag{3}$$

$$f_+ = \mathrm{opp}(f_+) + \frac{1}{2}\left(f_{IP-} - f_{IP+}\right) \pm \frac{1}{6}\varrho u_\perp \pm \frac{1}{2}\varrho u_\parallel, \tag{4}$$

$$f_- = \mathrm{opp}(f_-) - \frac{1}{2}\left(f_{IP-} - f_{IP+}\right) \pm \frac{1}{6}\varrho u_\perp \pm \frac{1}{2}\varrho u_\parallel. \tag{5}$$

We have introduced f_c, f_+ and f_- as the unknown distribution functions on the boundary where f_c denotes the central one and $f_\pm$ its right and left neighbours, respectively. The function $\mathrm{opp}()$ returns the DDF pointing in the opposite direction while $f_{IP\pm}$ is the *inplane* DDF, i.e. parallel to the boundary. The boundary's normal and parallel velocity component are denoted by $u_\perp$ and $u_\parallel$ and the sign is positive if its component is pointing in the positive coordinate direction. The density on the boundary is obtained from

$$\varrho = \frac{1}{1 \pm u_\perp}\left[f_0 + \sum_i^{inplane} f_i + 2\left(\sum_j^{outside} f_j\right)\right], \tag{6}$$

where we sum over the DDFs on the boundary (inplane) and outside of the boundary. Dirichlet or Neumann type boundary conditions may be applied to open boundaries to obtain the velocity components required in Eq.(3) through Eq.(5). For solid boundaries, the simple bounce-back scheme suffices and we have incorporated the treatment of Bouzidi [22] for curved boundaries. The macroscopic quantities at the end of each time step are obtained from

$$\varrho = \sum_{i=0}^{8} f_i, \tag{7}$$

$$p = \frac{\varrho}{3}, \tag{8}$$

$$\mathbf{u} = \sum_{i=0}^{8} \frac{f_i \mathbf{c}_i}{\varrho}. \tag{9}$$

We have built on the work of [23] and implemented the above mentioned procedure using the C language for better binding with our parallel approaches. The implementation has been augmented in order to increase its computational performance and lower the memory requirements where possible. The parallel implementations of our solver can run either on GPUs using CUDA or CPUs using UPC. The corresponding validation results along with parallel scalability information are presented in the following section.

3 Results and discussion

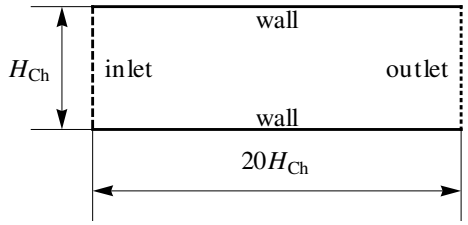
In this section we present the verification and the validation of our LBM solver through a set of five different test cases. The geometries of the investigated domains and the applied boundary conditions are presented in Figure 2. In our discussion we will limit ourselves to only presenting results obtained with our serial LBM solver, upon which we rely when performing the relevant verification and validation studies. Our parallel implementations in CUDA and UPC have been verified by ensuring that they reproduce the same results as the serial solver. Finally we discuss the parallel strategies and present their scaling behaviour.

3.1 Laminar Flow in a Channel

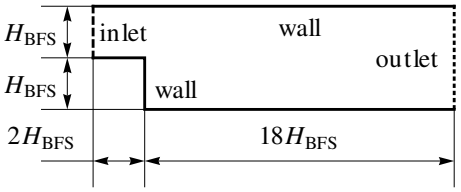
Our first validation case is the widely used laminar channel flow where we can compare our results to an analytic velocity profile. It further lends itself to a detailed grid convergence study as we can make predictions about its exact profile.
Table 1 shows the results obtained for the non-dimensional peak velocity at the centre of the channel for three different grids. We have given the normalised grid spacing in parenthesis, based on the finest grid, and further calculated the extrapolated value of the peak velocity for a grid independent domain. As discussed in detail in Roache [24], we have calculated the Grid Convergence Index (GCI) for the coarse-medium (1,2) and medium-fine (2,3) mesh and calculated the apparent order p_a to judge the accuracy of the solution. We can see from Table 1 that the BGKW collision model initially performs better compared to the MRT scheme, however, the extrapolated value shows that the MRT scheme approaches the exact solution faster. This is also reflected by the step decrease of the GCI when compared to the convergence

Table 1: Grid convergence study of the non-dimensional center (peak) velocity. The normalized grid spacing has been given in parenthesis and a grid independent peak velocity has been extrapolated. The GCI and apparent order is given from the coarse to medium (1,2) and medium to fine (2,3) mesh.

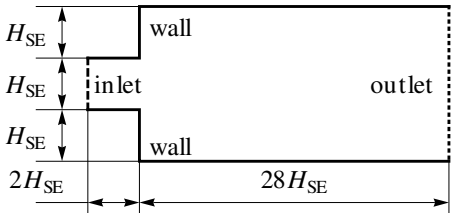
	$u/u_{\max}[-]$				$\mathrm{GCI}_{1,2}$	$\mathrm{GCI}_{2,3}$	p_a
$n_x \times n_y$	400×20 (4)	800×40 (2)	1600×80 (1)	extrapolated			
BGKW	0.9459	0.9761	0.9856	0.9899	1.77	0.55	1.67
MRT	0.7302	0.9191	0.9732	0.9948	10.29	2.78	1.81



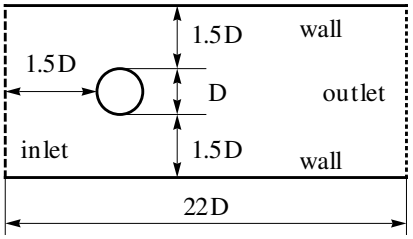
(a) Channel (Ch), 800×40 lattices



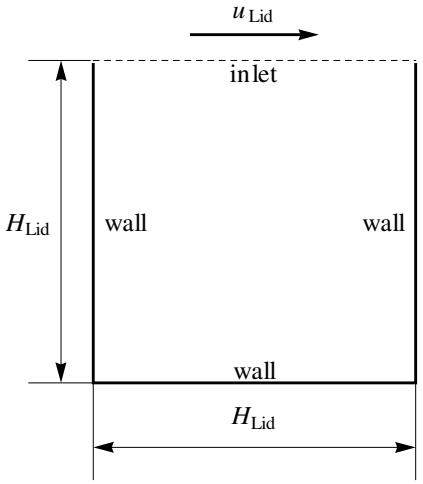
(b) Backward facing step (BFS), 800×80 lattices



(c) Sudden expansion (SE), 800×80 lattices



(d) Cylinder (Cyl), 880×160 lattices



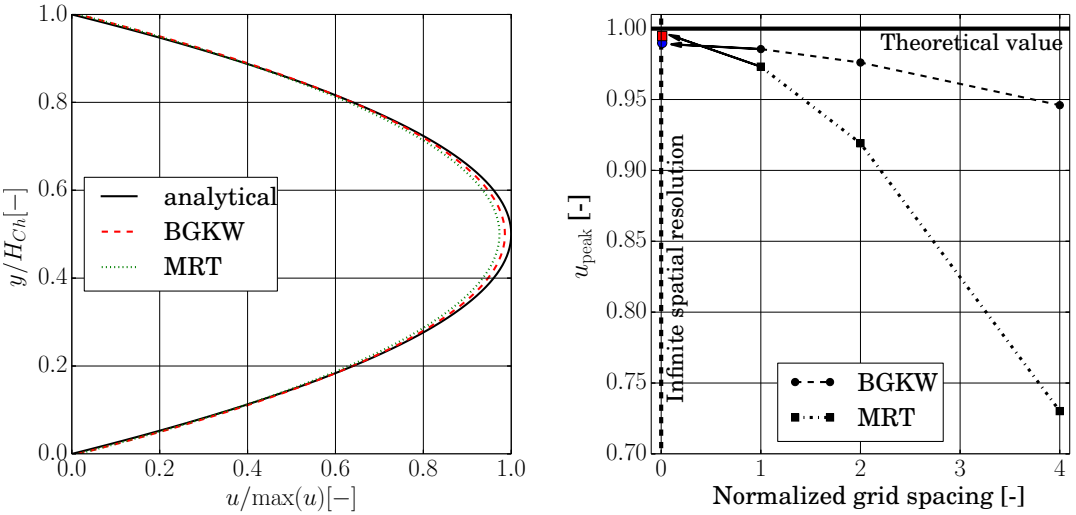
(e) Lid-driven cavity (LC), 400×400 lattices

Figure 2: Computational domains with their dimensions and boundary conditions. Open or moving boundaries are given by dashed lines

behaviour of the BGKW scheme and the apparent order confirms this, which is a measure of how fast the error diminishes.

The obtained velocity profile at the channel's outlet is presented in Figure 3(a) for the fine mesh. Figure 3(b) shows the convergence of the peak velocity over the normalised grid spacing. Figure 3(a) shows that both collision models produce qualitatively the same velocity profile and capture the correct behaviour. In Figure 3(b) we have summarised our findings mentioned above and show the convergence properties of the BGKW and MRT scheme. In the limit of a grid independent result, the MRT scheme asymptotically approaches the theoretical value closer than the BGKW scheme which, however, produces more accurate results on the coarse grid. Although not evident from our presented results, it has to be noted that the MRT scheme is able to simulate lower lattice viscosities than the BGKW model which suffers from numerical instabilities. This is due to the nature of the model itself. While all DDFs are relaxed at the same rate in the BGKW model, the MRT scheme relaxes all of its DDFs in a transformed

momentum space separately. Therefore it can treat each direction (Figure 1) individually and it is less prone to numerical instabilities. The lattice viscosity is used to calculate the Reynolds number and hence the BGKW scheme is not suited for high Reynolds number flows.



(a) Velocity profiles $\mathrm{Re} = 100$ (fine mesh)

(b) Convergence of the collision models

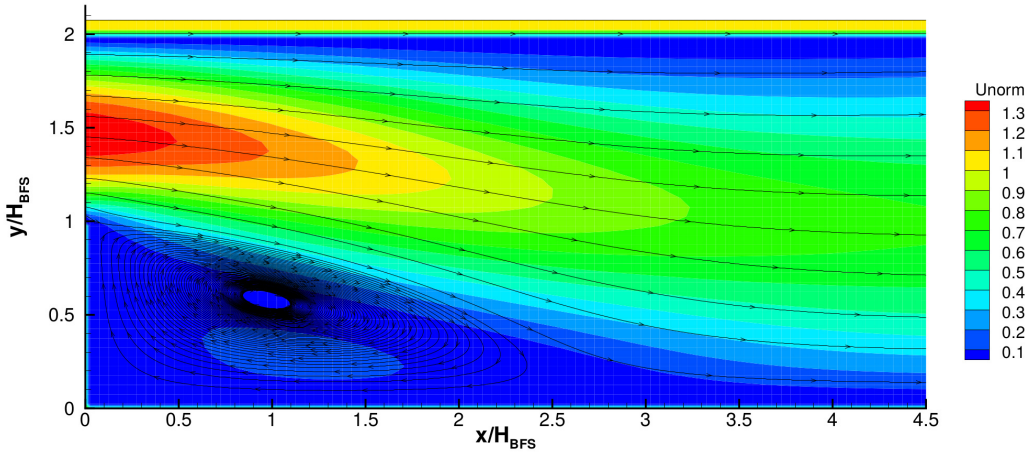
Figure 3: Validation of the solver on the channel flow

3.2 Flow over a backward facing step

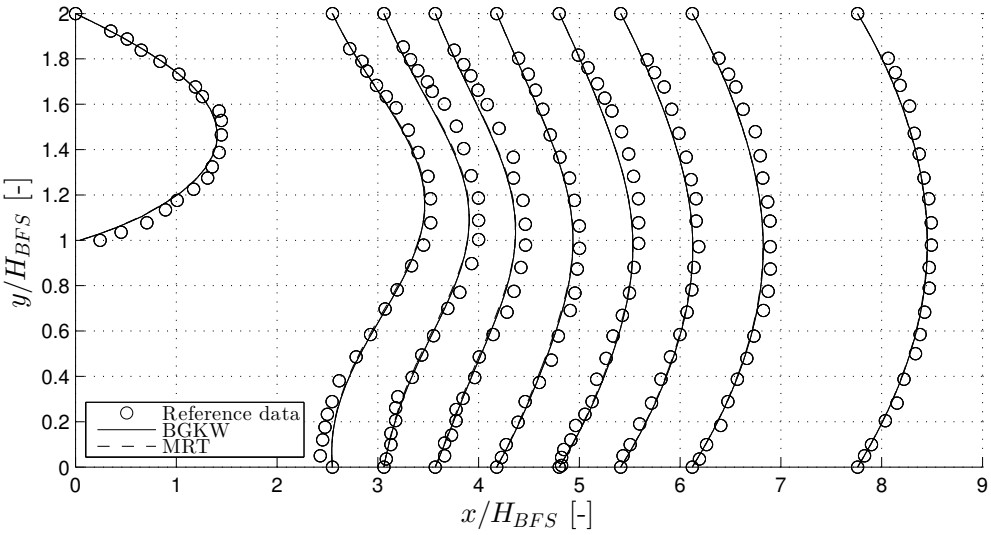
The Backward Facing Step (BFS) is another classical example for which an abundance of numerical and experimental data is available. We have examined the laminar flow at a Reynolds number of $Re = 100$ and we have chosen the experimental results of Armaly *et al.* [3] for our comparison.
Figure 4(a) shows the normalised velocity magnitude, superimposed by the streamlines as a visual guide. We only show the contour plot obtained for the MRT model as the BGKW results were qualitatively indistinguishable. A fully developed channel flow profile is present at the exit location of the smaller channel at $x/H_{BFS} = 0$ which then transitions into the bigger channel. A circulation area is formed downstream which is fed by the upstream channel. We have plotted the velocity profiles at various locations downstream of the channel and compared the results to the experimental data. We can see the fully developed velocity profile at the exit of the smaller channel, as indicated before, and then the development of the downstream profiles. Both the BGKW and MRT model under-predict the velocity profiles in the vicinity of the recirculation area and only approach the experimental data further downstream. Specifically, the magnitude of the velocity of the reverse flow and the jet are both under-estimated. Both collision models performed with little to chose between the two, which is also shown in the reattachment length error with -6.13% for the BGKW model and -5.48% for the MRT model, respectively.
We conclude that our solver closely reproduces the velocity field and predicts the reattachment length within a few percent compared to the experimental data.

3.3 The bifurcation behaviour via a sudden expansion

The Sudden Expansion (SE) is a test case which provides further physical insight into the performance of our LBM solver. While its geometry is symmetrical, as seen in Figure 2(c), two different and distinct flow patterns are encountered for different Reynolds numbers. This is caused by the bifurcation of the fluid and is attributed to the non-linear behaviour of the Navier–Stokes equations. Hence, it is an interesting test case as the governing equation of the lattice Boltzmann method (Eq.(1)) is linear in nature. We



(a) Developed velocity field after the BFS (MRT)

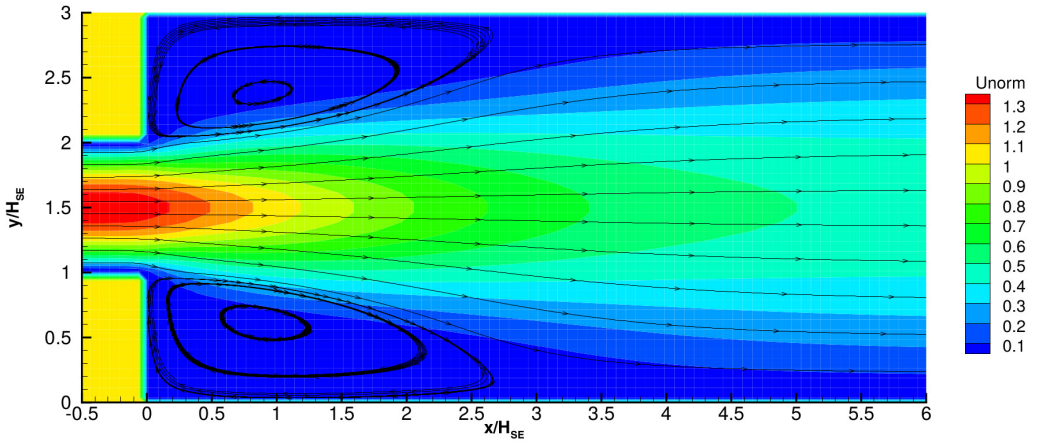


(b) Velocity profiles after the BFS

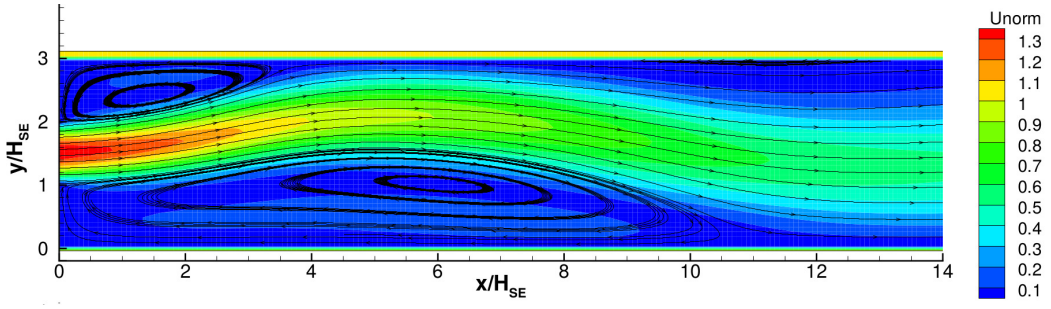
Figure 4: Validation of the solver on the BFS test case at $\mathrm{Re} = 100$

therefore performed simulations at Reynolds numbers of $Re = 25$, for a symmetrical flow field and $Re = 80$, for an anti-symmetrical, and compared our results against experimental data published by Fearn *et al.* [4].

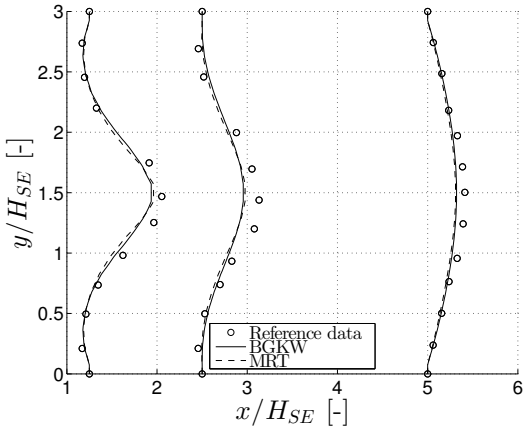
Figures 5(a) and 5(b) show the normalised velocity plots for both the $Re = 25$ and $Re = 80$ cases and we have again shown the streamlines as a visual guide. For the symmetrical flow field we can identify two recirculation zones of the same magnitude, located in the lower and upper corner while the jet is expanding symmetrically downstream. The $Re = 80$ case is showing the expected symmetry breaking effect at higher Reynolds numbers and two distinct recirculation areas are formed. A third one, characteristic for even higher Reynolds numbers, is starting to form at the top before the jet is returning to a symmetric velocity profile. Figures 5(c) and 5(d) show the velocity profile compared to the experimental data at three different downstream locations. Qualitatively the results are matched although we see again



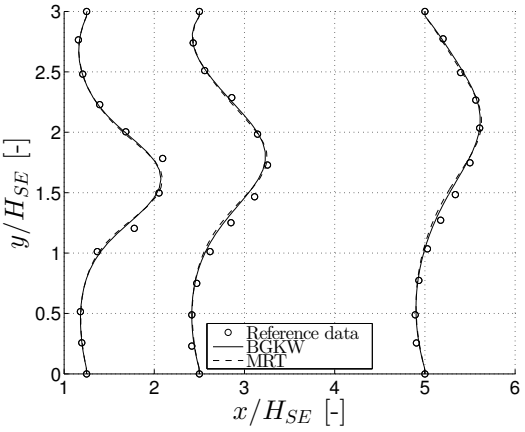
(a) Developed velocity field after the SE at $\mathrm{Re} = 25$ (MRT)



(b) Developed velocity field after the SE at $\mathrm{Re} = 80$ (MRT)



(c) Velocity profiles after the SE at $\mathrm{Re} = 25$



(d) Velocity profiles after the SE at $\mathrm{Re} = 80$

Figure 5: Validation of the solver on the SE test case at $\mathrm{Re} = 25$ and at $\mathrm{Re} = 80$

an under-prediction of the velocity magnitude in the transversal locations. The discrepancies are mainly found in the centre while the flow near the wall is resolved more closely. Interestingly, the $Re = 80$ case with the bifurcation exhibits closer agreement between the experimental and numerical results in terms of velocity profiles. Both the BGKW and MRT collisions models match well the experimental result.
In Table 2 we present the reattachment length L_{re} in non-dimensional units (normalised by H_{SE}, see Figure 2(c)) and its relative deviation to the experimental results. For both cases and for both recirculation zones in the $Re = 80$ case, the reattachment length has been under-predicted which is consistent with the observation of the backward facing step. However, we can see close agreement with the reference data. For the $Re = 25$ case we have less than 4% difference. The $Re = 80$ also shows acceptable

Table 2: Sudden expansion flow for two different Reynolds numbers. The $Re = 25$ is symmetrical and only one reattachment length needs to be calculated while the $Re = 80$ case produces a non symmetrical flow field which has two distinct (i.e. primary and secondary) vortex and hence two distinct reattachment lengths. The relative error corresponds to comparisons with Fearn *et al.* [4].

	$Re = 25$		$Re = 80$ (primary)		$Re = 80$ (secondary)	
	L_{re} [-]	$\varepsilon_{\text{relative}}$ [%]	L_{re} [-]	$\varepsilon_{\text{relative}}$ [%]	L_{re} [-]	$\varepsilon_{\text{relative}}$ [%]
BGKW	3.35	-2.33	3.45	-6.25	10.65	-9.44
MRT	3.30	-3.79	3.56	-3.26	10.60	-9.86

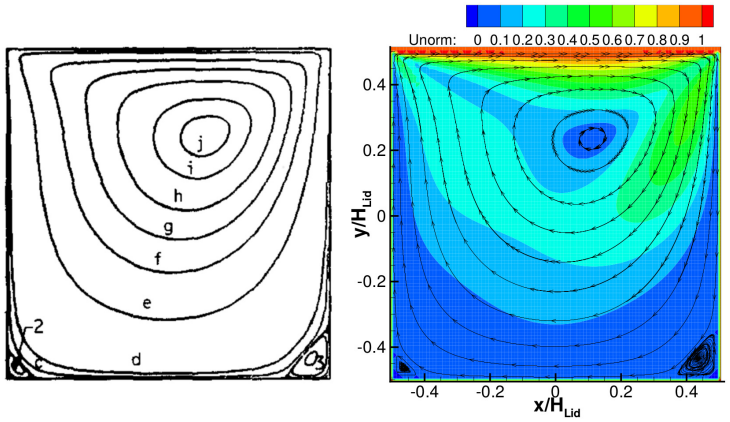
deviations of less than 7% for the primary and 10% for the secondary recirculation area, respectively. Hence the symmetric case was more accurately resolved which correlated well with the reattachment length prediction of the similar flow pattern over the backward facing step. The symmetry breaking at higher Reynolds numbers was captured, which is not surprising as the governing equation reproduces the Navier–Stokes equations via the Chapman–Enskog expansion, as mentioned in the introduction. However, it remains a linear equation and thus the higher difference compared to the symmetric case, in terms of reattachment length, is not surprising.
We can conclude that our solver performed well for the symmetric case and is capable of capturing the bifurcation behaviour of a fluid.

3.4 Flow structures in a lid-driven cavity flow

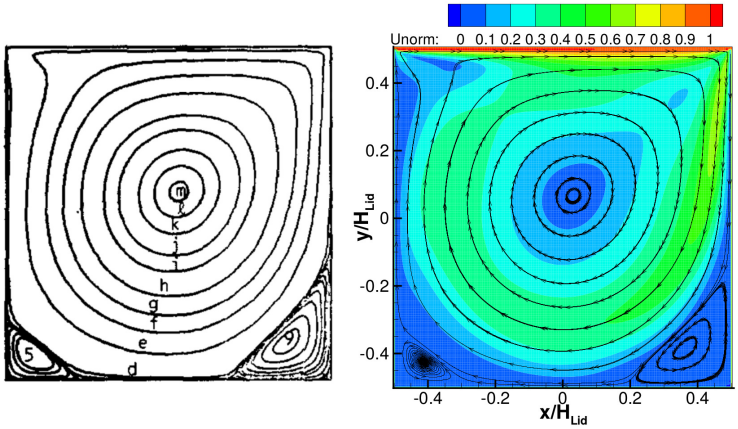
We now investigate the lid driven cavity flow which consists of simple boundary conditions. These can be accurately described and hence we can focus on the physical behaviour of the solver. We compare our results with the numerical data from Ghia *et al.* [2], where the vorticity-stream function approach was chosen to solve the 2D Navier–Stokes equations. It is of interest because it offers 2D data without any turbulence models, hence, we can validate our data against a Navier–Stokes solver which obtained results under the same conditions.
Figures 6(a)–6(c) show the streamlines obtained from the reference solutions (left) and our simulations (right) at $Re = 100$, $Re = 1000$ and $Re = 3200$, respectively. We see that the primary and secondary vortices in the lower corners are well-resolved for the $Re = 100$ and $Re = 1000$ case while the movement of the main vortex towards the centre is captured. For the high Reynolds number flow at $Re = 3200$, a tertiary vortex is formed near the lid which is consistent with the reference solution. Our simulations compare well with the reference data by pure visual means.
We further investigate the velocity profiles at the centerlines of the box, i.e. horizontally and vertically, and present our results in Figure 7(a)–7(c). For all investigated cases we see a good agreement between the velocity profiles of the reference data and both the BGKW and MRT scheme. Minor differences exist which can be either due to the numerical scheme used in the reference solution or the accuracy in the boundary condition imposition. The order of accuracy at the wall depends on the placement of the lattice, i.e. either coinciding with the DDFs or in-between them. See Succi [1] for a discussion.
We conclude that the two-dimensional flow features are accurately resolved and good agreement could be obtained with respect to the reference data.

3.5 Turbulent Flow around a cylinder

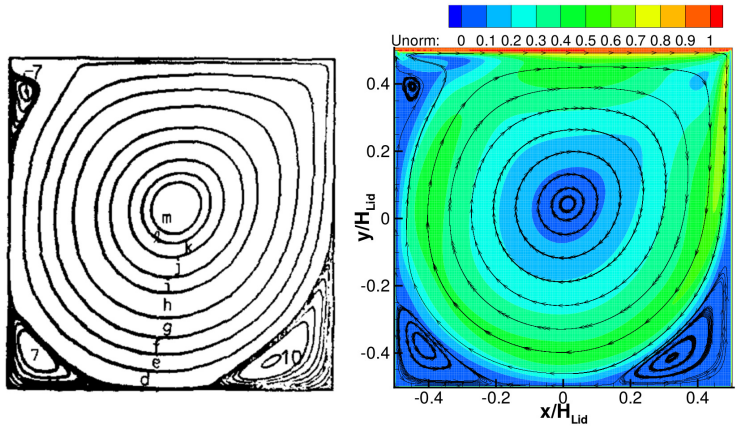
The solver presented in [23] has been extensively tested for the laminar flow around the channel. The von Kármán vortex street has been properly resolved and thus the sinusoidal lift and drag behaviour recovered. Here we present initial results for our LBM solver applied to the same cylinder flow at $Re = 10000$ to demonstrate the capabilities of the solver to resolve turbulent structures. The cylinder



(a) Re = 100



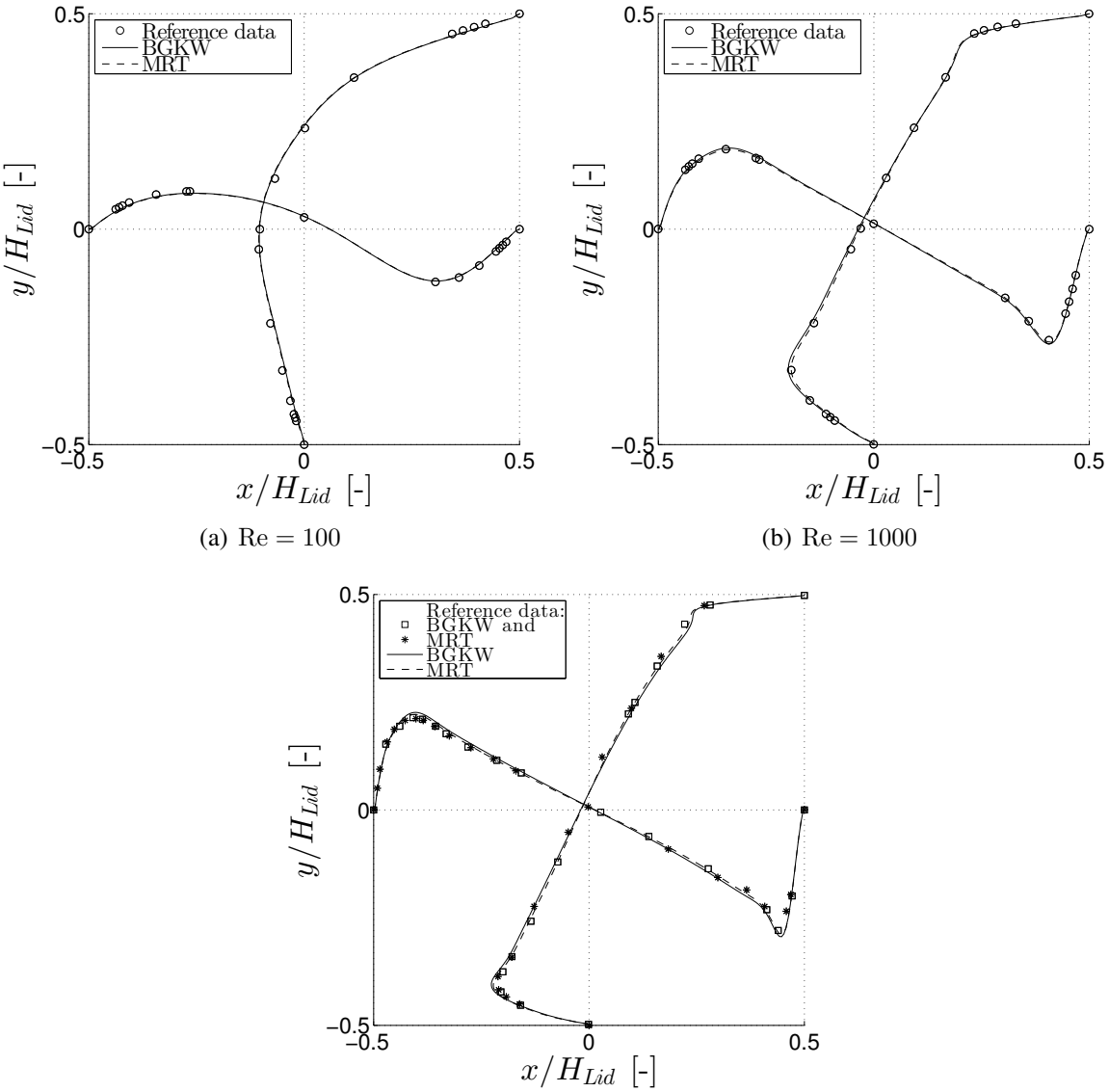
(b) Re = 1000



(c) Re = 3200

Figure 6: Velocity field comparison between Ghia *et al.* [2] (left) and the results obtained with the MRT collision model (right)

itself is an interesting test case as it consists of curved boundaries. Despite the stair-stepping profile around the cylinder, the treatment of Bouzidi *et al.* [22] accounts for the curvature of the geometry which is seen by the smooth vorticity distribution around the cylinder.
The presented results were obtained by the MRT collision model as the BGKW, as mentioned in section 3.1, is not suitable for high Reynolds number flows due to its inherent numerical instabilities. Fig-



(a) Re = 100

(b) Re = 1000

(c) Re = 3200

Figure 7: Velocity profiles at the horizontal and vertical centerline of the lid driven cavity at $Re = 100$, $Re = 1000$ and $Re = 3200$, respectively

ure 8 shows the contours of instantaneous, normalised vorticity magnitude in the downstream cylinder location. We can see that vortices are forming just downstream of the cylinder which then detach and are convected into the wake region. We show positive rotation in red and negative in blue so as to show the asymmetric vortex formation at the cylinder and its pairwise existence close to it, before turbulent mixing diffuses the overall vorticity.

In this proof of concept study we wanted to show the applicability of our solver to curved boundaries and high Reynolds number flows. Before we can properly validate our solver for turbulent flows, the solver needs to be extended to three dimensions.

3.6 Scaling behaviour of the solver

We conclude our discussion on validation and verification with the scaling behaviour of our parallel LBM solvers. We start by introducing the notions of parallel speed-up P_{SU} and parallel efficiency P_E which

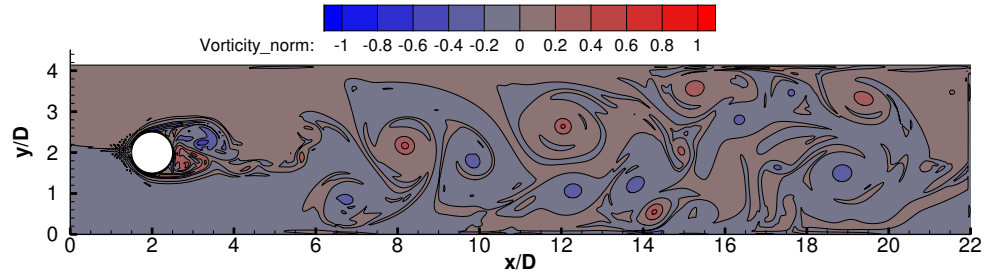


Figure 8: Flow past a circular cylinder at $Re = 10000$ showing the normalised vorticity magnitude

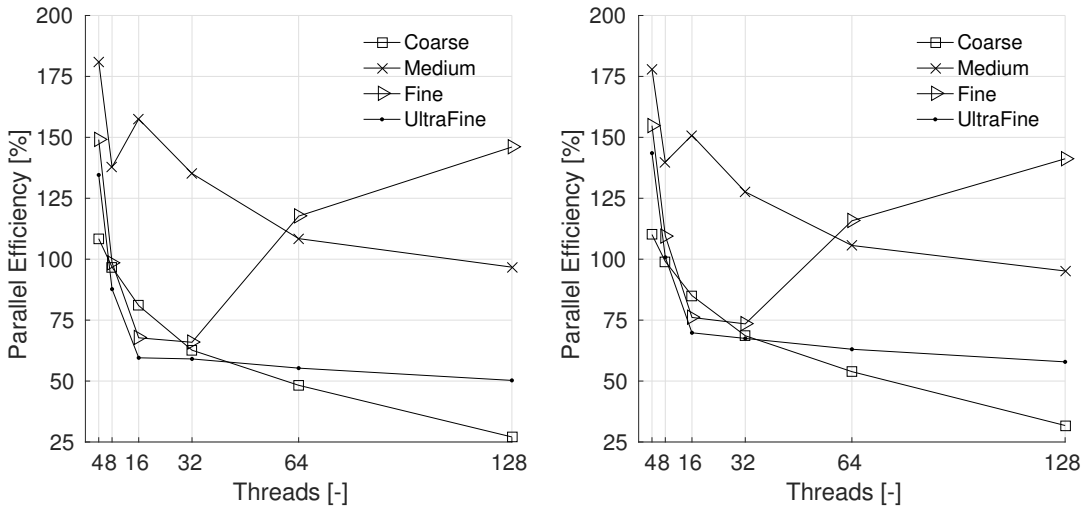
are defined respectively as

$$P_{SU} = t_{\text{serial}}/t_{\text{parallel}} \tag{10}$$

and

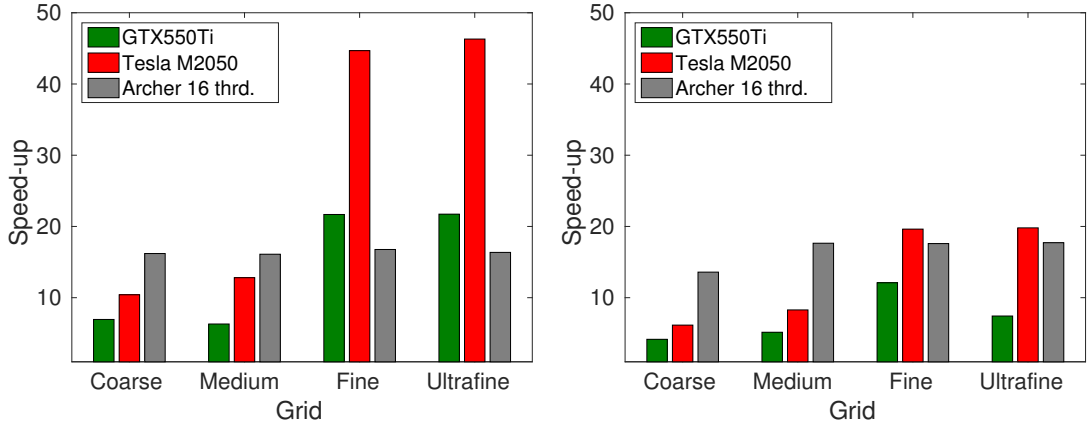
$$P_E = (P_{SU}/n_{\text{cores}})100\% = t_{\text{serial}}/(t_{\text{parallel}} \cdot n_{\text{cores}})100\% \tag{11}$$

By t_{serial} we denote the time required for a simulation to complete serially, by t_{parallel} we denote the time required for a parallel simulation to complete and by n_{cores} we denote the number of cores employed during a parallel simulation. We have tested our solver on the lid driven cavity flow for which we have created four different meshes with increasing number of points: the spatial resolutions 128^2, 256^2, 512^2, and 1024^2 correspond to the coarse, medium, fine and ultrafine grids, respectively. The results for the UPC and CUDA strategy are summarised in Figure 9 and Figure 10 respectively.
The UPC implementation shows that the parallel efficiency can be greater than 100%. Opposed to the serial computation, tasks can be overlapped in the UPC environment which enables the solver to achieve such high rates of efficiency. However, we also see that it drops for a larger amount of threads and all, except the fine mesh, experience degeneration of the efficiency due to increasing overheads. The fact that the fine mesh does not follow this trend may have several roots. Although PGAS gives a coherent view on memory, physically it is still separated. The way the partitioning and communications are done in the background may explain the trend of the fine grid in Figure 9 but also highlights the need for code profiling and compiler optimisation. Results obtained with the BGKW and the MRT models show similar



(a) BGKW model (b) MRT model

Figure 9: Parallel performance of the parallel UPC code using double precision arithmetic



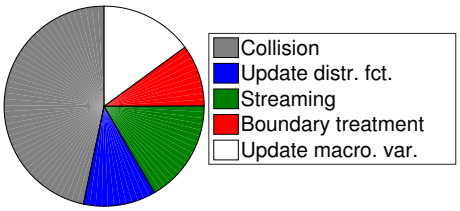
(a) BGKW model (b) MRT model

Figure 10: Speed-up evaluation of the parallel codes using single precision arithmetic on two different GPUs

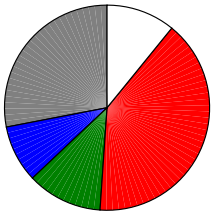
parallel efficiency trends as seen in Figure 9(a) and 9(b). It is worth to mention that the performance, measured with single and double precision, showed only marginal differences.
Figure 10 presents the speed-up of the CUDA implementation. We have tested the code on two different GPUs; the GTX550Ti which is targeted as a consumer graphics card and the Tesla M2050 which is purely designed for scientific computations. We have given speed-ups obtained with UPC for 16 threads on all four grid levels as a reference. The Tesla M2050 consistently yielded higher speed-ups compared to the GTX550Ti which were about twice as much. While the UPC approach yielded a speed-up approximately equal to its number of cores, the CUDA implementation, due to its different architecture, was able to achieve speed-ups of up to 45 compared to the serial code, see Figure 10(a). In the case of the CUDA implementation we experienced a significant drop for double precision computations which only yielded about one third of the single precision speed-ups. We measured a relatively high speed-up and strong scalability on the GPUs with the BGKW model, however, a considerable performance drop was experienced with the MRT collision model when we compare Figures 10(a) and 10(b).
The profiling results, presented in Figure 11, help to understand the speed-up behaviour of the solver. In Figure 11(c) we can see that the collision is the most computationally expensive operation in the case of the MRT collision model. As a matter of fact in the case of the CUDA implementation it takes approximately 60% of every time step, seen in Figure 11(d). While the performance difference between the BGKW and the MRT model seems to be negligible based on the serial profiling results of Figure 11(a) and Figure 11(c), the MRT collision step is the bottleneck of the time marching on the GPUs as can bee seen in Figure 11(b) and Figure 11(d). The distribution functions are stored in physically different locations for the MRT collision model. Hence, during the summation of all nine discrete velocity directions, data needs to be transferred via the global memory which has a relatively low bandwidth. Therefore, the MRT scheme exhibits a slower computational time, compared to the BGKW model, which could be circumvented by using shared memory on the GPU side.

4 Conclusions

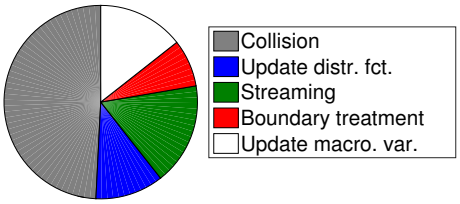
In this study, we have validated our 2-dimensional lattice Boltzmann solver against the classical benchmark problems of a channel, backward facing step, sudden expansion, lid driven cavity and turbulent cylinder flow. We have shown that our solver is capable of treating complex geometries with curved boundaries at high Reynolds number flows and have demonstrated the accuracy of our solver to be within a small percentage of reference data. We extended our solver with high performance capabilities and included CUDA as a GPU-based, and UPC, as a CPU-based parallelsiation strategy, respectively. High efficiencies were obtained for the UPC approach for moderate numbers of CPUs while overheads



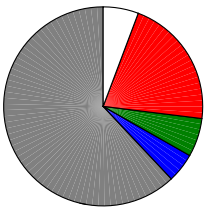
(a)



(b) CUDA, BGKW



(c) Serial code, MRT



(d) CUDA, MRT

Figure 11: Profiling results on the medium mesh using single precision arithmetic

were introduced at higher numbers. CUDA was able to give speed-ups as high as 45 compared to the serial solver which decreased drastically for double precision computations. The extension of solver to three dimensions is left to the future which will allow for turbulent flow applications.

5 Acknowledgement

This work used the ARCHER UK National Supercomputing Service (`http://www.archer.ac.uk`), and the Eddie2 local cluster at the University of Edinburgh for the parallel performance evaluation; we are grateful for the support of their staff. Special thanks to Gennaro Abbruzzese for the original C++ version of the code and the mesh generator which was used for our simulations.

REFERENCES

[1] S. Succi, *The lattice Boltzmann equation: for fluid dynamics and beyond*. Oxford University Press, Oxford, 2001.

[2] U. Ghia, K. N. Ghia, and C. T. Shin, High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method. *Journal of Computational Physics*, **48**, 387–411, 1982.

[3] B. F. Armaly, F. Durst, J. C. F. Pereira, and B. Schnung, Experimental and theoretical investigation of backward-facing step flow. *Journal of Fluid Mechanics*, **127**, 473–496, 1983.

[4] R. M. Fearn, T. Mullin, and K. A. Cliffe, Nonlinear flow phenomena in a symmetric sudden expansion. *Journal of Fluid Mechanics*, **211**, 595–608, 1990.

[5] J. Sanders and E. Kandrot, *CUDA by example: an introduction to general-purpose GPU programming*. Addison-Wesley Professional, 2010.

[6] T. El-Ghazawi, W. Carlson, T. Sterling, K. Yelick, *UPC: Distributed Shared Memory Programming*. Wiley, 2005.

[7] G. R. McNamara and G. Zanetti, Use of the Boltzmann equation to simulate lattice-gas automata. *Physical Review Letters*, **61**, 2332, 1988.

[8] S. Chen, G. D. Doolen, Lattice-Boltzmann method for fluid flows. *Annual Review of Fluid Mechanics*, **30**, 329–364, 1998.

[9] C. K. Aidun, J. R. Clausen, Lattice-Boltzmann method for complex flows. *Annual Review of Fluid Mechanics*, **42**, 439–472, 2010.

[10] D.M. Holman, R.M Brionnaud, Z. Abiza, Solution to industry benchmark problems with the lattice-Boltzmann code XFlow. *6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012)*, Vienna, Austria, September 10–14, 2012.

[11] C.L.A. Lafferty, A.C. Alajbegovic, K.B. Horrigan, Under-hood Thermal Simulation of a Class 8 Truck. *SAE Technical Paper*, 2007.

[12] G. Balasubramanian, L.A. Raghu Mutnuri, Z. Sugiyama, S. Senthooran, D. Freed, A computational process for early stage assessment of automotive buffeting and wind noise. *SAE International Journal of Passenger Cars - Mechanical Systems*, **6**, 2, 1231–1238, 2013.

[13] X. He, L.-S. Luo, Lattice Boltzmann Model for the Incompressible Navier–Stokes Equation. *Journal of Statistical Physics*, **88**, 3, 927–944, 1997.

[14] M.R. Salimi, M. Taeibi-Rahni, F. Jam, Heat transfer analysis of a porously covered heated square cylinder, using a hybrid Navier–Stokes-lattice Boltzmann numerical method. *International Journal of Thermal Sciences*, **91**, 59–75, 2015.

[15] J.A. Diaz, C.B. Muñoz-Caro, A.B. Niño, A survey of parallel programming models and tools in the multi and many-core era. *IEEE Transactions on Parallel and Distributed Systems*, **23**, 8, 1369–1386, 2012.

[16] F. Kuznik, C. Obrecht, G. Rusaouen, J.-J. Roux, LBM based flow simulation using GPU computing processor. *Computers and Mathematics with Applications*, **59**, 7, 2380–2392, 2010.

[17] P. L. Bhatnagar, E. P. Gross, and M. Krook, A model for collision processes in gases I: Small amplitude processes in charged and neutral one-component systems. *Physical Review*, **94**, 511, 1954.

[18] P. Welander, On the temperature jump in a rarefied gas. *Arkiv Fysik*, **7**, 1954

[19] D. d'Humières, Generalized lattice-Boltzmann equations. In *Rarefied gas dynamics: Theory and simulations* (ed. B. D. Shizgal & D. P. Weaver), 1992.

[20] D. d'Humières, Multiplerelaxationtime lattice Boltzmann models in three dimensions. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **360**, 437–451, 2002.

[21] Q. Zou and X. He, On pressure and velocity boundary conditions for the lattice Boltzmann BGK model. *Physics of Fluids*, **9**, 1591–1598, 1997.

[22] M. A. Bouzidi, M. B. Firdaouss, P. A. Lallemand, Momentum transfer of a Boltzmann-lattice fluid with boundaries. *Physics of Fluids*, **13**, 11, 3452–3459, 2001.

[23] G. Abbruzzese, Development of a 2D lattice Boltzmann code. *Master's thesis, Cranfield University*, 2013.

[24] P. J. Roache, *Verification and Validation in Computational Science and Engineering*. Hermosa Publishers, Albuquerque, New Mexico, 1998.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

VALIDATION OF A MAGNETO- AND FERRO-HYDRODYNAMIC MODEL FOR NON-ISOTHERMAL FLOWS IN CONJUNCTION WITH NEWTONIAN AND NON-NEWTONIAN FLUIDS

L. Könözsy[1], P. Scienza[1], and D. Drikakis[2]

[1]Centre for Fluid Mechanics and Computational Science
Cranfield University, Cranfield, Bedfordshire, MK43 0AL, United Kingdom
e-mail: laszlo.konozsy@cranfield.ac.uk, pietro.scienza@icloud.com

[2] Faculty of Engineering
University of Strathclyde, Glasgow, G1 1XW, United Kingdom
e-mail: dimitris.drikakis@strath.ac.uk

Keywords: magnetohydrodynamics, ferro-hydrodynamics, non-isothermal Hartmann flows, Newtonian fluids, non-Newtonian fluids, non-isothermal blood flows.

Abstract. *This work focuses on the validation of a magnetohydrodynamic (MHD) and ferro-hydrodynamic (FHD) model for non-isothermal flows in conjunction with Newtonian and non-Newtonian fluids. The importance of this research field is to gain insight into the interaction of non-linear viscous behaviour of blood flow in the presence of MHD and FHD effects, because its biomedical application such as magneto resonance imaging (MRI) is in the centre of research interest. For incompressible flows coupled with MHD and FHD models, the Lorentz force and a Joule heating term appear due to the MHD effects and the magnetization and magnetocaloric terms appear due to the FHD effects in the non-linear momentum and temperature equations, respectively. Tzirtzilakis and Loukopoulos [1] investigated the effects of MHD and FHD for incompressible non-isothermal flows in conjunction with Newtonian fluids in a small rectangular channel. Their model excluded the non-linear viscous behaviour of blood flows considering blood as a Newtonian biofluid. Tzirakis et al. [2, 3] modelled the effects of MHD and FHD for incompressible isothermal flows in a circular duct and through a stenosis in conjunction with both Newtonian and non-Newtonian fluids, although their approach neglects the non-isothermal magnetocaloric FHD effects. Due to the fact that there is a lack of experimental data available for non-isothermal and non-Newtonian blood flows in the presence of MHD and FHD effects, therefore the objective of this study is to establish adequate validation test cases in order to assess the reliability of the implemented non-isothermal and non-Newtonian MHD-FHD models. The non-isothermal Hartmann flow has been chosen as a benchmark physical problem to study velocity and temperature distributions for Newtonian fluids and non-Newtonian blood flows in a planar microfluidic channel. In addition to this, the numerical behaviour of an incompressible and non-isothermal non-Newtonian blood flow has been investigated from computational aspects when a dipole-like rotational magnetic field generated by infinite conducting wires. The numerical results are compared to available computational data taken from literature [2].*

1 INTRODUCTION

For modelling incompressible flows coupled with magnetohydrodynamic (MHD) and ferrohydrodynamic (FHD) approaches, the Lorentz force and a Joule heating term appear due to the MHD effects and the magnetization and magnetocaloric terms appear due to the FHD effects in the non-linear momentum and temperature equations, respectively. This work focuses on the validation of an implemented MHD-FHD model for non-isothermal flows in conjunction with Newtonian and non-Newtonian fluids. The importance of this research work is to investigate the interaction of non-linear viscous behaviour of blood flow with MHD and FHD effects, because its biomedical applications such as magneto resonance imaging (MRI) is in the centre of interest. Due to the fact that there is a lack of experimental data available for non-isothermal and non-Newtonian blood flows in the presence of MHD and FHD field effects, the objective of this study is to establish adequate computational test cases for validation purposes in order to assess the reliability of the non-isothermal and non-Newtonian MHD-FHD models.

A number of theoretical and numerical studies were carried out in the literature to investigate the blood flow behaviour under the action of a magnetic field with particular interest to flows in rectangular ducts and stenosed tubes [1, 2, 3]. Tzirtzilakis et al. [4] proposed a numerical approach to study a blood flow in a three-dimensional rectangular duct where an 8 $Tesla$ magnetic field was generated by an electrically conducing wire on the lower wall of the duct. The blood was considered as a Newtonian electrically non-conducting biofluid and a FHD model was implemented. In their model, the magnetization was assumed linearly dependent on the magnetic field intensity and the magnetic field completely saturated the biofluid. They concluded that two recirculation regions generated slowing down the flow when the magnetic field intensity was increased, and the strong magnetic field also affected the wall stresses near to the area where it was located. In addition to this, a secondary flow was found to be present and strengthen due to the increase of the magnetic field intensity. Loukopoulos and Tzirtzilakis [5] investigated a laminar, viscous incompressible flow of a Newtonian biomagnetic fluid in a two-dimensional channel under the action of spatially varying magnetic field at a low Reynolds number (Re = 250). In this numerical study, the magnetic field was located on the lower wall of the planar channel by taking into account FHD effects with a magnetization linearly dependent on the temperature and magnetic field intensity. They considered realistic blood parameters, different temperatures at the lower and upper walls and a fully developed velocity profile was prescribed as a boundary condition at the inlet. It was observed that a recirculation region developed in proximity of the magnetic field and extended when the magnetic field intensity increased. The flow separated close to the magnetic source and reattached downstream. The velocity profile became fully developed while temperature perturbations were observed at the outlet.

Another numerical study was carried out by Tzirtzilakis and Loukopoulos [1] to investigate the behaviour of blood flows in a two-dimensional small planar channel under the action of a uniform external magnetic field of 8 $Tesla$ generated from two finite disks. The blood was considered as a Newtonian fluid with constant electrical conductivity. The combination of both MHD and FHD models were taken into account in this numerical study and the formation of two recirculation regions was observed where the magnetic field started and ended. In addition to this, two secondary eddies were generated downstream close to the second recirculation region on the lower and upper plates. Due to the rotation of these two vortices, the fluid is pushed on the upper wall within the region of constant magnetic field and the flow slowed down near to the first recirculation region. Furthermore, the velocity profile became fully developed downstream behind the applied uniform magnetic field. It is important to note that the temperature increased

dramatically closer to the lower plate in proximity of the source of the magnetic field and further increased slowly in the downstream flow regime. In contrast with the velocity field behaviour, the temperature distribution showed a perturbed profile even at the outlet. A variable trend of the wall shear stress inside the magnetized region was also observed in this study.

It has been emphasized in [6] that the effects of temperature field on the numerical solution in conjunction with MHD-FHD models have to be further investigated for biomedical and microfluidic applications. This is due to the fact that there is a lack of experimental data available on the interaction between the temperature and MHD-FHD fields for non-Newtonian blood flows. Therefore this interaction is further investigated and the existing MHD-FHD models are further validated by considering computational aspects in this paper.

Tzirakis et al. [2] conducted a remarkable numerical investigation on the effect of three different magnetic fields for blood flows in four different geometrical configurations. They investigated an incompressible and isothermal flow considering blood as a Newtonian magnetizable and electrically conducting fluid. In their numerical study, a constant magnetic field was applied to plane Couette and Hartmann flows for validating their model implementation. A spatially variable magnetic field generated by an external wire was used for investigating the plane Couette flow and a three-component rotational magnetic field were applied to the flow through a stenosed tube. Their results showed that an irrotational magnetic field does not affect the velocity field, however a rotational magnetic field brakes the flow symmetry inducing separations and altering the velocity distribution. The work of Tzirakis et al. [2] represented an in-depth analysis of magnetic field effects and its investigation of blood flow behaviour, therefore their benchmark test case has also been further validated for non-isothermal flows in this paper.

As a matter of fact, there is a lack of experimental data available for non-isothermal non-Newtonian blood flows in the presence of MHD and FHD effects, therefore the interaction between the temperature field and MHD-FHD models for non-Newtonian blood flows has been investigated in this paper. A two-dimensional, incompressible, laminar and non-isothermal Hartmann flow has been chosen as a benchmark physical problem to study velocity and temperature distributions for Newtonian fluids and non-Newtonian blood flows in a planar microfluidic channel. In addition to this, the numerical behaviour of an incompressible and non-isothermal non-Newtonian blood flow has been investigated from computational aspects when a dipole-like rotational magnetic field generated by infinite conducting wires. The numerical results are compared to available computational data taken from literature [2].

2 MATHEMATICAL FORMULATION AND SOLUTION METHODOLOGY

2.1 Governing equations in conjunction with combined MHD-FHD models

In this study, for modelling incompressible non-isothermal flows in conjunction with MHD and FHD models, we adopt the mathematical model equations proposed by Tzirtzilakis [7]. The system of governing equations consists of the continuity, momentum and energy equations, thus the mass conservation for incompressible flows can be written in a vector form as

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

which is the divergence-free (incompressibility) constraint for the velocity field $\mathbf{u}$. For steady-state incompressible flows, the momentum equation in conjunction with the source terms of MHD and FHD models can be expressed by

$$\rho \left(\mathbf{u} \cdot \nabla \right) \mathbf{u} = \rho \mathbf{g} - \nabla p + \nabla \cdot \underline{\underline{\tau}} + \mu_0 \left(\mathbf{M} \cdot \nabla \right) \mathbf{H} + \mathbf{J} \times \mathbf{B}, \tag{2}$$

where ρ is the fluid density, $\mathbf{g}$ represents the gravity field, p is the hydrodynamic pressure, $\underline{\underline{\tau}}$ is the viscous stress tensor which can be defined for both Newtonian and non-Newtonian fluids, μ_0 is the magnetic permeability of vacuum, $\mathbf{M}$ is the magnetization vector, $\mathbf{H}$ is the magnetic field intensity, $\mathbf{J}$ is the current density vector and $\mathbf{B}$ is the magnetic flux density, respectively. For modelling non-isothermal flows of Newtonian and non-Newtonian fluids with various rheological behaviours, the energy equation for the temperature field can be written as

$$\rho c_p \left(\mathbf{u}\cdot\nabla\right) T = \lambda\nabla^2 T - \mu_0 T \frac{\partial \mathbf{M}}{\partial T}\left(\mathbf{u}\cdot\nabla\right)\mathbf{H} + \frac{\mathbf{J^2}}{\sigma} + \underline{\underline{\tau}}\cdot\cdot\left(\nabla\otimes\mathbf{u}\right), \tag{3}$$

where c_p is the specific heat at constant pressure, T is the temperature, and σ is the electric conductivity of the fluid. The first term on the right hand side of the temperature equation represents heat transfer through conduction, the second term represents the magnetocaloric effect of FHD, the third term represents the effect of MHD, and the fourth term $\varphi_D = \underline{\underline{\tau}}\cdot\cdot\left(\nabla\otimes\mathbf{u}\right)$ is the viscous dissipation function which is the double dot (inner) product of the viscous stress tensor and the velocity gradient tensor. In the set of governing equations of incompressible non-isothermal flows in conjunction with MHD and FHD models, the current density vector can be expressed by the rotation (curl) of the magnetic field intensity vector $\mathbf{H}$ as

$$\mathbf{J} = \nabla\times\mathbf{H} = \sigma\left(\mathbf{u}\times\mathbf{B}\right), \tag{4}$$

where the electric field effect has been neglected in the present case. We assume that the magnetic flux density field $\mathbf{B}$ is divergence-free as

$$\nabla\cdot\mathbf{B} = 0. \tag{5}$$

For modelling MHD-FHD flows for Newtonian and non-Newtonian fluids, the magnetization force term appears as an external force on the right hand side of the non-linear momentum equation (2) due to the effect of FHD as

$$\mathbf{f_M} = \mathbf{f_{FHD}} = \mu_0\left(\mathbf{M}\cdot\nabla\right)\mathbf{H}, \tag{6}$$

and the Lorentz-force appears due to the effect of MHD which can be written as

$$\mathbf{f_L} = \mathbf{f_{MHD}} = \mathbf{J}\times\mathbf{B} = \sigma\left(\mathbf{u}\times\mathbf{B}\right)\times\mathbf{B}. \tag{7}$$

When the fluid is directly sensitive to a magnetic field, its behaviour can be described by FHD model equations and by a magnetic state through the magnetization equation $\mathbf{M} = \mathbf{M}\left(\mathbf{H}, T\right)$, which is usually dependent on the magnetic field intensity and the temperature often through too complex or very simple non-realistic relations. However, for some weak magnetic conductors such as blood, a simplified equation can be employed as

$$\mathbf{M} = \chi\mathbf{H}, \tag{8}$$

which expresses the relationship between the magnetization field $\mathbf{M}$ and the magnetic field intensity $\mathbf{H}$ through the magnetic susceptibility of the fluid χ, which can be defined by

$$\mathbf{H} = \frac{1}{\mu_0}\frac{1}{\left(1+\chi\right)}\mathbf{B}, \tag{9}$$

thus the momentum equation (2) for incompressible, stationary, non-isothermal flows for Newtonian and non-Newtonian fluids can also be written as

$$\rho\left(\mathbf{u}\cdot\nabla\right)\mathbf{u}=-\nabla p+\nabla\cdot\underline{\underline{\tau}}+\alpha\left(\mathbf{B}\cdot\nabla\right)\mathbf{B}+\sigma\left(\mathbf{u}\times\mathbf{B}\right)\times\mathbf{B}, \tag{10}$$

and the temperature equation (3) becomes

$$\rho c_p\left(\mathbf{u}\cdot\nabla\right)T=\lambda\nabla^2 T+\frac{\mathbf{J^2}}{\sigma}+\underline{\underline{\tau}}\cdot\cdot\left(\nabla\otimes\mathbf{u}\right), \tag{11}$$

and the parameter α can be defined by

$$\alpha=\frac{1}{\mu_0}\frac{\chi}{(1+\chi)^2}, \tag{12}$$

where χ is the magnetic susceptibility which expresses the material reactivity to an external magnetic field and can change according to the oxygenation of blood in the present case. For oxygenated blood, the magnetic susceptibility can be chosen as $\chi_{ox}=-6.6\cdot10^{-7}$ which is related to the diamagnetic behaviour of blood, and $\chi_{deox}=3.5\cdot10^{-6}$ for de-oxygenated blood according to [2] which characterizes a paramagnetic behaviour. In the present numerical study, the magnetic susceptibility of blood has been considered as de-oxygenated.

The magnetization force appears in Eq. (2) due to the FHD effects which can be written as

$$\mathbf{f_M}=\mathbf{f_{FHD}}=\alpha\left(\mathbf{B}\cdot\nabla\right)\mathbf{B}=f_{Mx}\mathbf{e}_x+f_{My}\mathbf{e}_y, \tag{13}$$

where the scalar components of the magnetization vector $\mathbf{f_M}$ can be defined as

$$f_{Mx}=\alpha\left(B_x\frac{\partial B_x}{\partial x}+B_y\frac{\partial B_x}{\partial y}\right), \tag{14}$$

$$f_{My}=\alpha\left(B_x\frac{\partial B_y}{\partial x}+B_y\frac{\partial B_y}{\partial y}\right), \tag{15}$$

which terms appear as additional source terms in the scalar momentum equations for x and y spatial directions for two-dimensional flows. The Lorentz-force appears due to the MHD effects in the momentum equation (2) which can be expressed as

$$\mathbf{f_L}=\mathbf{f_{MHD}}=\mathbf{J}\times\mathbf{B}=\sigma\left(\mathbf{u}\times\mathbf{B}\right)\times\mathbf{B}, \tag{16}$$

where the scalar components of the Lorentz force vector $\mathbf{f_L}$ can be written as

$$f_{Lx}=\sigma\left(-uB_y^2+vB_xB_y\right), \tag{17}$$

$$f_{Ly}=\sigma\left(uB_xB_y-vB_x^2\right), \tag{18}$$

which terms again appear as additional source terms in the scalar momentum equations for x and y spatial directions. The magnetic field can be considered as a dipole-like rotational field which was also used by Tzirakis et al. [2] as

$$\mathbf{B}=\mathbf{B}\left(x,y\right)=B_x\mathbf{e}_x+B_y\mathbf{e}_y, \tag{19}$$

where the scalar components of the magnetic flux density vector $\mathbf{B}$ can be defined by

$$B_x = -C\left[\frac{2\left(x-x_i\right)^2 - r^2}{r^6}\right], \tag{20}$$

$$B_y = -C\left[\frac{2\left(x-x_i\right)\cdot\left(y-y_i\right)}{r^6}\right], \tag{21}$$

where the dipole radius is defined for a two-dimensional problem as

$$r = \sqrt{\left(x-x_i\right)^2 + \left(y-y_i\right)^2}, \tag{22}$$

where x_i and y_i are the spatial coordinates of field application points, and C is an appropriately chosen constant ($C = 1.72 \cdot 10^{-10}\, Tm^4$ in the present numerical study).

2.2 Non-Newtonian Power-Law Fluid (PLF) Model

The momentum equation (10) is valid for either Newtonian or non-Newtonian fluids depending on the definition of the viscous stress tensor $\underline{\underline{\tau}}$. Blood is known to be a non-Newtonian fluid which exhibits a shear thinning attitude in particuler to those flows when the Reynolds number is low and the viscous term is dominant. In consequence of this, the non-linear relationship between the viscous stress tensor and the rate of strain has to be taken into account in order to model the non-Newtonain fluid flow behaviour. For non-Newtonian fluids, the dynamic viscosity coefficient which is also called as apparent viscosity is a function of the shear rate as

$$\mu_a = \mu_a\left(\left|\underline{\underline{\dot{\gamma}}}\right|\right), \tag{23}$$

therefore the relationship between the viscous stress and shear rate tensors is non-linear as

$$\underline{\underline{\tau}} = \mu_a\left(\left|\underline{\underline{\dot{\gamma}}}\right|\right)\underline{\underline{\gamma}}, \tag{24}$$

where the shear rate tensor is defined as two times of the rate of strain (deformation) tensor as

$$\underline{\underline{\gamma}} = 2\underline{\underline{\mathrm{S}}} = \nabla \otimes \mathbf{u} + \left(\nabla \otimes \mathbf{u}\right)^T, \tag{25}$$

and the magnitude of the shear rate tensor can be defined by

$$\dot{\gamma} = \left|\underline{\underline{\gamma}}\right| = \sqrt{\frac{1}{2}\underline{\underline{\gamma}} \cdot \cdot \underline{\underline{\gamma}}}. \tag{26}$$

For modelling non-Newtonian blood flows, different blood rheological approaches are available in the literature such as Power-Law Fluid (PLF), Quemada and Casson models to make an attempt to describe the realistic rheological behaviour of blood [8]. In the present study, we choose the aforementioned PLF model due to its simplicity to model the non-Newtonian non-linear viscous behaviour of blood. In the PLF blood rheological model, the shear stress is a function of an n-power coefficient of the shear strain rate as

$$\tau = k_0 \dot{\gamma}^n, \tag{27}$$

where k_0 and n are appropriately chosen coefficients of the PLF model, and these parameters are suggested to be selected as $k_0 = 14.67 \cdot 10^{-7} Pa \cdot s^n$ and $n = 0.7755$ [8]. Relying on the non-Newtonial PLF model, the viscous stress tensor can be written as

$$\underline{\underline{\tau}} = k_0 \dot{\gamma}^{n-1} \underline{\underline{\gamma}}, \tag{28}$$

and the Reynolds number definition for the non-Newtonian viscous model can be written [8] as

$$Re_{PLF} = \frac{\rho h^n}{k_0 U_\infty^{n-2}}. \tag{29}$$

2.3 The Hartmann Flow

The Hartmann flow in its initial configuration is a planar flow between two parallel flat plates in the presence of a constant magnetic field $\mathbf{B} = B_0 \mathbf{j}$ normal to the plates. In this case, the Lorentz force component f_{Lx} appears in the momentum equation as

$$f_{Lx} = -\sigma u B_0^2. \tag{30}$$

Due to the fact that the magnetic field is constant and parallel to the y-axis, the flow is affected along the x-direction only, which represents the main advantage of the Hartmann flow for validation purposes. When the Hartmann flow is considered to be a one-dimensional flow, it is possible to derive an analytical solution for the outlet velocity profile which can be compared to the numerical implementation of an MHD model. The outlet velocity profile of a two-dimensional laminar Hartmann flow for a Newtonian fluid can be written [2] as

$$u\left(y\right) = \frac{1}{\sigma B_0^2} \left(\frac{dp}{dx}\right) \left[1 - \frac{\cosh\left(\frac{Hy}{h}\right)}{\cosh\left(H\right)}\right], \tag{31}$$

where the dimensionless Hartmann number can be defined as

$$H_a = h B_0 \sqrt{\frac{\sigma}{\mu}}. \tag{32}$$

The effect of the magnetic field acting on the non-isotermal flow field is studied also in relation to the temperature perturbations. The temperature-velocity coupling is made clear in the energy equation where an additional source term appears as

$$e_{MHD} = B_x^2 u^2 + B_y^2 v^2. \tag{33}$$

2.4 Implementation of MHD-FHD model equations

The combination of MHD and FHD models [2, 7] described in this Section is not available in the ANSYS-FLUENT commercial software package by default. Therefore, the MHD-FHD model for the governing momentum (10) and energy (11) equations in conjunction with the magnetization (13) and Lorentz force (16) source terms has been implemented in the ANSYS-FLUENT environment with User-Defined Functions (UDFs) in the C programming language. The scalar components of the magnetization force (14) and (15), and scalar components of the Lorentz force (17) and (18) has been added to the ANSYS-FLUENT Navier-Stokes solver by employing the SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) pressure-correction algorithm [10] with second-order discretization scheme for the convective flux terms.

The magnetic field intensity with a single-wire and the rotational magnetic field have been initialized through the UDF implementing its components along the spatial directions. It should be mentioned in terms of the UDF C code implementation that the source terms have to be defined in a "generation-rate/volume" unit. The magnetic field components (x_i, y_i) have been stored at the cell-centre on the computational domain and the appropriately chosen C constant has been defined by the actual $4\,Tesla$ field area of action which can be easily adapted and modified according to the requirements. The reason for choosing the magnetic field strength as $4\,Tesla$ is in fact that it was difficult to obtain converged solution for the governing equations (10) and (11) by taking higher values above this value. In order to establish reliable validation test cases for MHD-FHD applications, the convergence of the solution has been considered as a requirement within a small threshold value (smaller than 10^{-7} or 10^{-8}). The simulation parameters, geometrical data of a two-dimensional microfluidic benchmark channel including realistic material properties for blood have been summarized in Table 1.

Symbol	Value	Description	Units
h	0.01	Height of the channel	(m)
L	0.12	Length of the channel	(m)
μ_0	$1.256637 \cdot 10^{-6}$	Magnetic permeability of vacuum	(H/m)
χ_{deox}	$3.5 \cdot 10^{-6}$	Magnetic susceptibility of de-oxygenated blood	$(-)$
B	4	Magnetic field intensity	(T)
x_0	0.03	Magnetic field location in direction-x	(m)
y_0	$7.42 \cdot 10^{-3}$	Magnetic field location in direction-y	(m)
C	$1.72 \cdot 10^{-10}$	Rotational magnetic field constant	$(T \cdot m^4)$
ρ	1050	Density of blood	(kg/m^3)
μ_{Newt}	0.0035	Dynamic viscosity of the Newtonian fluid	$(Pa \cdot s)$
C_p	3617	Specific heat of blood	$(J/Kg \cdot K)$
k	0.52	Thermal conductivity of blood	$(W/m \cdot K)$
σ	0.8	Electrical conductivity of blood	(S/m)
T	300	Temperature of blood	(K)
T_{lw}	283.65	Lower wall temperature	(K)
T_{uw}	315.15	Upper wall temperature	(K)
Re	100	Reynolds number	$(-)$
Re_{PLF}	120.025	Reynolds number of the Power-Law Fluid	$(-)$
k_0	$14.67 \cdot 10^{-3}$	Power-Law viscosity coefficient	$(-)$
n	0.7755	Power-Law exponent	$(-)$
U_∞	0.1	Inlet flow velocity	(m/s)

Table 1: Simulation parameters, geometrical data and material properties for blood.

For the temperature distribution at the inlet section of the microfluidic channel, a linear temperature profile has been prescribed as a boundary condition as

$$T_{inlet}\,(y) = y(T_{uw} - T_{lw}) + T_{lw}, \tag{34}$$

and for the velocity field, a uniform inlet velocity distribution has been taken with an average velocity magnitude of $0.1\,m/s$ (see Table 1).

3 RESULTS AND DISCUSSION

3.1 Non-isothermal Hartmann flow for Newtonian and non-Newtonian fluids

In this subsection, we investigate an incompressible, laminar and non-isothermal Hartmann flow in conjunction with Newtonian and non-Newtonian rheological approaches for blood in a two-dimensional microfluidic channel. The computational data has been compared to the numerical data taken from Tzirakis et al. [1] where the fluid has been considered as Newtonian. For this simple preliminary analysis, differently from [1] where a computational mesh consists of 2000 quadrilateral elements was used, a grid of approximately 14000 elements has been found [9] to be more effective in the present study providing high-accuracy and agreement with the results of Tzirakis et al. [1]. A uniform inlet velocity profile has been considered relying on [1] as well as the magnetic field strength has been taken as $B_0 = 25\sqrt{7}\,Tesla$. The Hartmann number defined by Eq. (32) has been taken as equal to 5. It is important to remark from a practical point-of-view that this value of magnetic field strength is considerably high and unsuitable for biomedical applications. However, the two-dimensional Hartmann flow as a first preliminary case is considered for the validation of the implemented UDFs in the ANSYS-FLUENT environment, because the flow physics is relatively simple and the size of the computational domain is small. The fully developed outlet velocity profile of the Hartmann flow compared to the computational data of Tzirakis et al. [1] is shown in Figure 1.

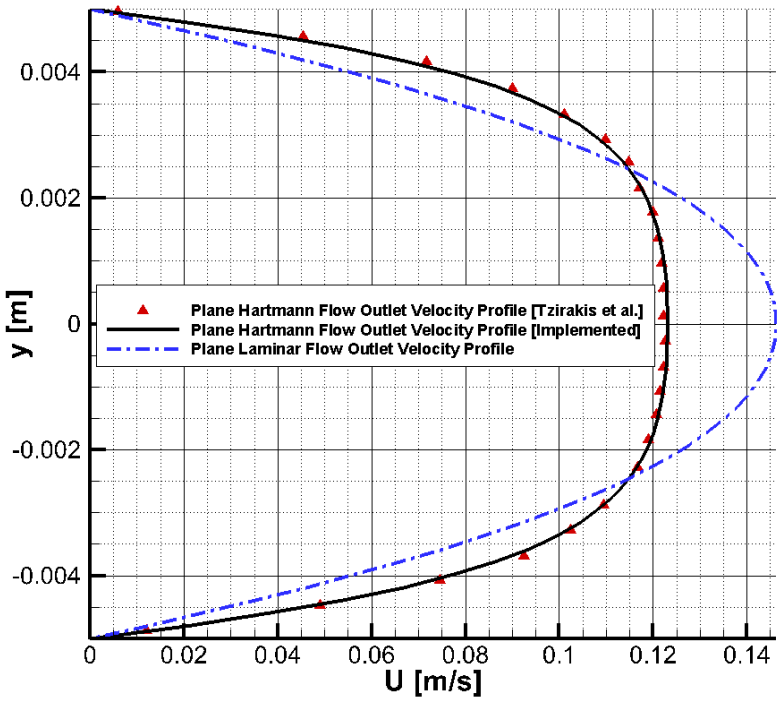


Figure 1: Comparison of outlet velocity profiles for a Hartmann flow in a microfluidic channel against computational data of Tzirakis et al. [2].

It can be seen in Figure 1 that the outlet velocity profile of a fully developed Hartmann flow is different from the fully developed velocity distribution of an incompressible laminar flow

without the presence of a magnetic field as it is expected. The effect of the magnetic force can be observed as the flow slows down under the action of a constant magnetic field. In addition to this, comparing the Hartmann flow and the laminar flow between parallel flat plates, it can be inferred that the first one develops quicker with a flatter velocity profile at the outlet compared to the second one (see the contour plots in Figures 2 and 3). This behaviour is caused by the magnetic field which slows down the blood flow (see Figures 3 and 5). The contour plot of the temperature field in the presence of a constant magnetic field is shown in Figures 4 and 6.

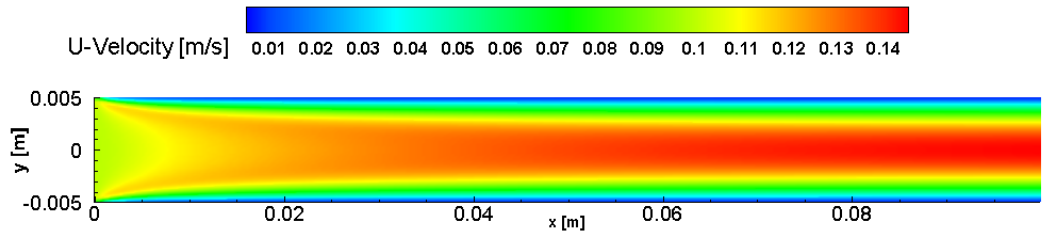


Figure 2: Contours of the velocity component U for a non-isothermal Hartmann flow in a two-dimensional microfluidic channel in conjunction with a Newtonian fluid with a uniform inlet velocity distribution.

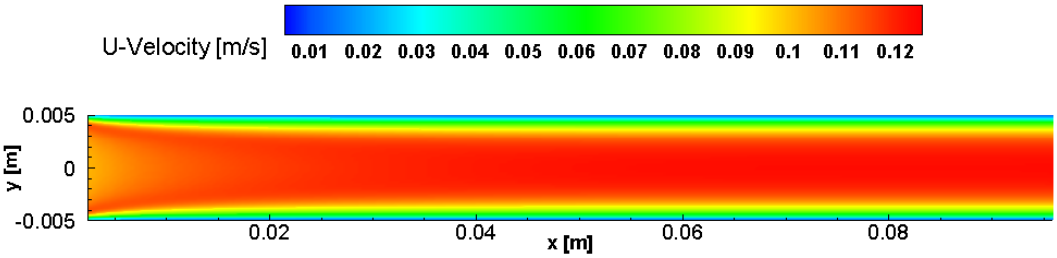


Figure 3: Contours of the velocity component U for a non-isothermal Hartmann flow in a two-dimensional microfluidic channel in conjunction with a non-Newtonian fluid (blood) with a uniform inlet velocity distribution.

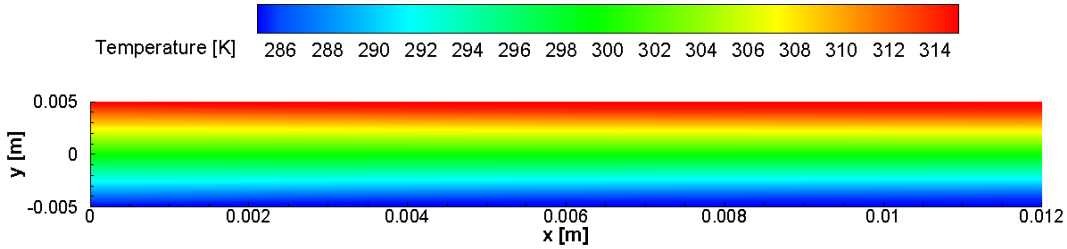


Figure 4: Contours of the temperature field for a non-isothermal Hartmann flow in a two-dimensional microfluidic channel in conjunction with a non-Newtonian fluid (blood) with a linear inlet temperature distribution.

The velocity profiles in a two-dimensional microfluidic channel at different cross-sections have been shown in Figure 5. It can be seen that there are minor differences between the cross-sectional velocity profiles for considering blood as a Newtonian and a non-Newtonian fluid

corresponding to the assumtion that blood can be considered as a weak non-Newtonian fluid. The effect of a constant magnetic field on the incompressible and laminar flow can also be observed in Figures 2 and 3, because the cross-sectional velocity profiles appear in a round-like shape compared to the fully developed laminar velocity profile without the presence of a constant magnetic field (see the blue curve in Figure 1).

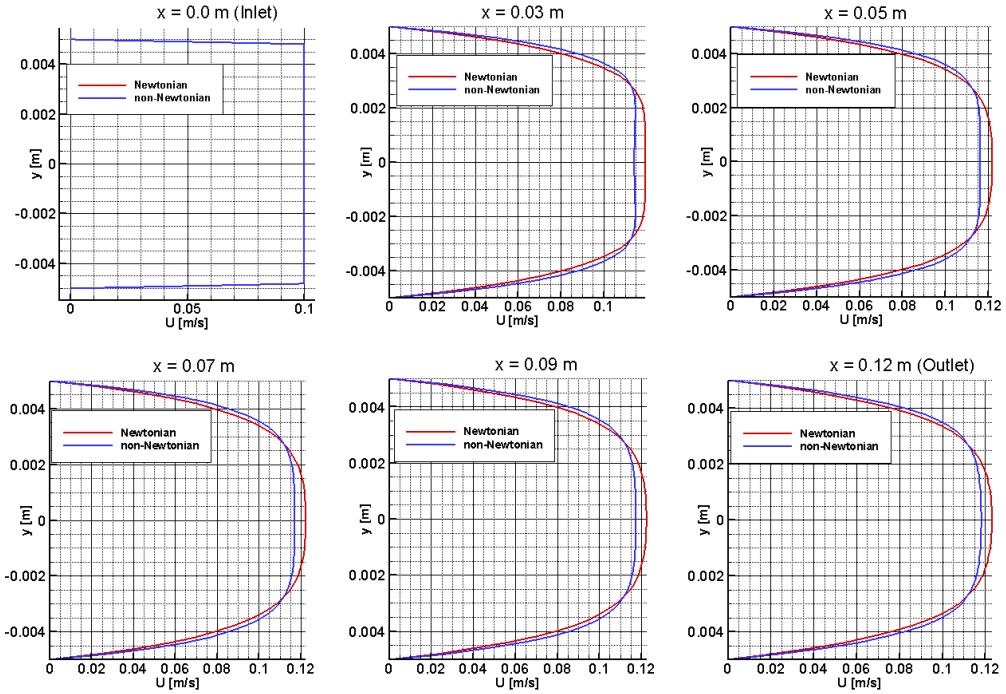


Figure 5: Comparison of velocity profiles for Newtonian and non-Newtonian blood rheological approaches in conjunction with a non-isothermal Hartmann flow in a microfluidic channel at different cross-sections.

For non-isothermal Hartmann flows in conjunction with Newtonian and non-Newtonian rheological models for blood, the temperature field is affected by the presence of the Lorentz force due to the effect of MHD. In this case, the source terms appearing due to the theory of FHD are neglected in the momentum and energy equations. The upper and lower wall temperatures can be found in Table 1, which are $T_{lw} = 283.65\,K$ and $T_{uw} = 315.15\,K$, respectively. The total temperature difference between the lower and upper walls is $31.5\,K$ which is a relatively small variation, however considering the small size of the microchannel, temperature profiles can also be affected by the presence of a constant magnetic field. It is valid in particular for those problems when a high magnetic field strength is applied perpendicular to the microfluidic channel. However, for the present Hartmann flow case, the effect of the MHD source term on the right hand side of the temperature equation is small due to the fact that the applied magnetic field strength is also small. The inlet temperature profile has been considered as a linear temperature distribution which can be seen in Figure 6. Since the temperature difference between the lower and upper walls as well as the applied constant magnetic field strength is also small, the temperature distribution at different cross-sections of the microfluidic channel exhibits minor differences between Newtonian and non-Newtonian rheological approaches for blood. The presence of the applied constant magnetic field affects slightly the temperature distribution compared to the assumed inlet linear profile. One can see in the following subsections

that the effect of the applied magnetic field becomes significant when a single wire or two wires are located at different axial coordinates in the microfluidic channel. This fact was also observed by Tzirtzilakis et al. [1] when a strong uniform magnetic field was applied.

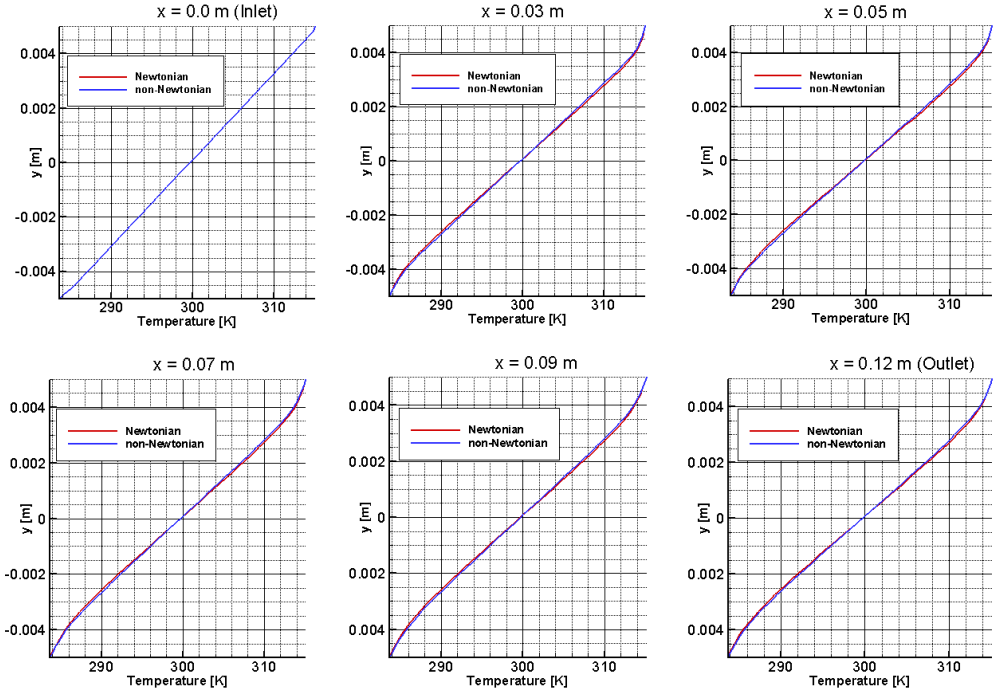


Figure 6: Comparison of temperature profiles for Newtonian and non-Newtonian blood rheological approaches in conjunction with a non-isothermal Hartmann flow in a microfluidic channel at different cross-sections.

The convergence history of the numerical solution for non-isothermal Hartmann flow in conjunction with Newtonian and non-Newtonian blood rheological models can be seen in Figure 7. For solving a steady-state fluid flow problem, the convergence criterion has been prescribed within a small threshold value (smaller than 10^{-8}). The maximum number of iterations for Newtonian and non-Newtonian fluids was 145 to obtain convergence to the approximate solution of the governing equations. The numerical results show that the ANSYS-FLUENT environment by using UDFs in C programming language is capable of solving MHD non-isothermal Newtonian and non-Newtonian fluid flows within a relatively small number of iterations on a fine computational mesh consists of approximately 14000 control volume elements. It has been observed that the magnetic field strength has to be also small in particular to non-Newtonian fluids, otherwise the approximate solution of the system of equations can lead to oscillatory and non-physical solutions. For taking a high value of the applied constant magnetic field strength, the source terms in the momentum (10) and temperature equations (11) might become very stiff, thus a careful numerical treatment and even advanced method development could be required within the framework of an in-house code implementation to obtain physically reasonable solutions for the system of governing equations. The stiffness of the MHD source terms in the set of model equations could lead to a numerical solution when the convergence of the numerical procedure could be difficult to ensure. Therefore, the applied magnetic field strength is not higher than $4\,Tesla$ in the following subsections for the investigated validation test cases to ensure numerical solution which converges to a small threshold value.

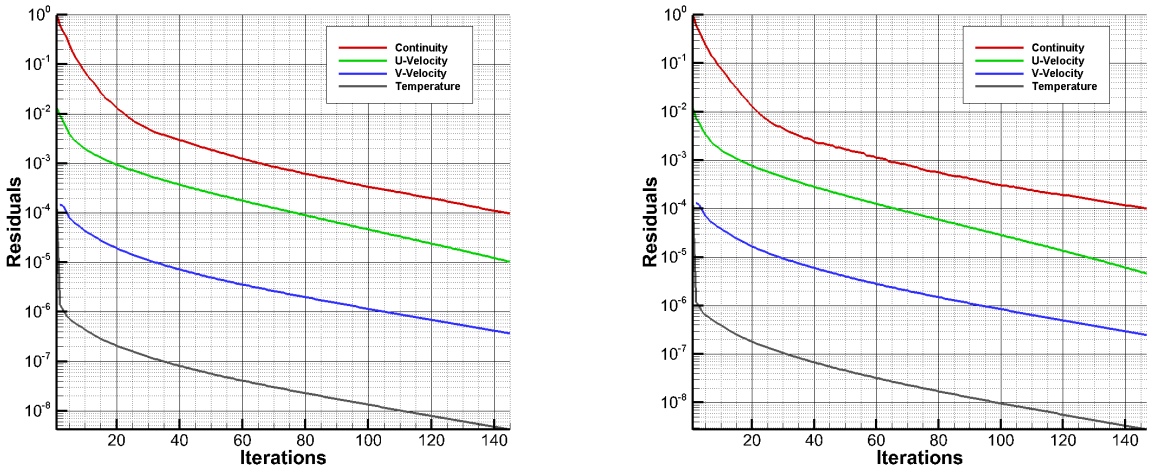


Figure 7: Residual history of non-isothermal Hartmann flow simulations: Newtonian fluid flow (left) and non-Newtonian blood flow (right).

3.2 Single wire problem

In this subsection, we present a benchmark problem for solving incompressible, laminar and non-isothermal flows in conjunction with MHD and FHD models and Newtonian and non-Newtonian blood rheological approaches. A single wire has been located in a two-dimensional microfluidic channel at the spatial coordinate $x_1 = 0.03\,m$ where the magnetic field strength has been taken as $4\,Tesla$. The local arrangement of the single wire in the two-dimensional microfluidic channel has been shown in Figure 8.

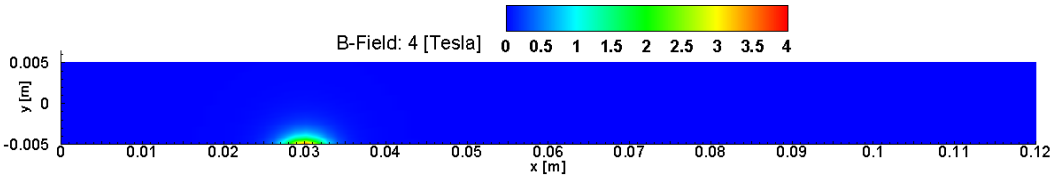


Figure 8: Contours of a single wire induced magnetic field of $4\,Tesla$ on the lower wall at $x_1 = 0.03\,m$.

This benchmark problem is a hybridization of the single wire model used by Tzirakis et al. [2] and the mathematical model for blood flows and magnetic fields proposed by Tzitzilakis [7]. The magnetization force due to the effect of FHD and the Lorentz force due to the effect of MHD have been taken into account in the momentum equation (10). We assume that the magnetic state through the magnetization equation $\mathbf{M} = \mathbf{M}(\mathbf{H})$ depends only on the magnetic field intensity $\mathbf{H}$, however the magnetization depends on the temperature T in reality. Relying on this assumption which can be used for weak magnetic conductors such as blood [2], the effects of MHD and FHD is taken into account in the momentum equation (10), however, only the effect of MHD appears in the temperature equation (11). This is due to the fact that the temperature derivative of the magnetization vector $\mathbf{M}$ vanishes from the energy equation (3), because it is assumed that the magnetization is function of the magnetic field intensity, but not

function of the temperature. In other words, it also means that the magneticaloric effect of FHD has been neglected in the MHD-FHD set of model equations in conjunction with Newtonian and non-Newtonian fluids, because the blood can be considered as a weak magnetic conductor.

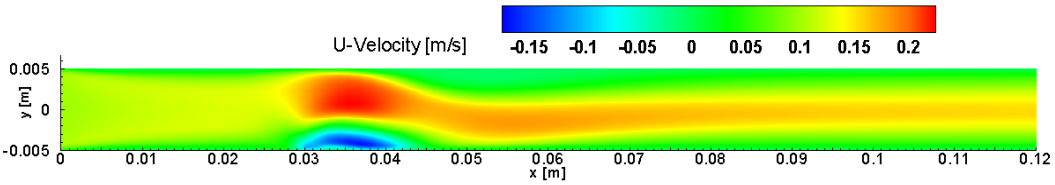


Figure 9: Contours of the velocity component U for a non-isothermal non-Newtonian blood flow in a microfluidic channel with a uniform inlet velocity distribution when the applied magnetic field is induced by a single wire.

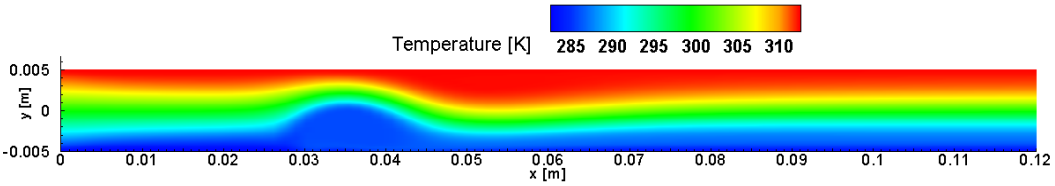


Figure 10: Contours of the temperature field for a non-isothermal non-Newtonian blood flow in a microfluidic channel with a linear inlet temperature distribution when the applied magnetic field is induced by a single wire.

The reason for investigating incompressible and non-isothermal MHD-FHD Newtonian and non-Newtonian fluid flows with the hybridization of Tzirakis et al. [2] and Tzitzilakis [7] MHD-FHD models is to study further the interaction between the temperature and MHD-FHD fields for non-Newtonian blood flows, because there is a lack of experimental data available in the literature. In addition to this, there is a lack of well-established and reliable numerical benchmark test case available for verification and validation purposes, thus we also focus on the computational fluid dynamics (CFD) aspects of this physical problem. For steady-state flows, the governing equations (10)-(11) with MHD and FHD models in conjunction with Newtonian and non-Newtonian fluids can be implemented in an in-house code or in a commercial software environment by using User-Defined Functions. The governing equations of the combination of MHD and FHD models [2, 7] described in Section 2 is not available in the ANSYS-FLUENT software package by default, therefore the implementation of a complex multiphysics problem related to MHD and FHD models gives an opportunity to study the strength and limitations of these models from computational aspects. Therefore, it is important to mention from a numerical modelling point-of-view that it has been observed that the presence of the magnetocaloric term of FHD on the right hand side of the temperature equation (3) could lead to numerically unstable and non-physical results when the magnetization vector $\mathbf{M}$ is depending on the temperature. This can be explained by the fact that the temperature dependence of the magnetization can be expressed through either too complex mathematical relationships or may be described with very simple and non-realistic functions. As an example, our numerical experience shows that the fully coupled MHD-FHD model for non-isothermal Newtonian flows under the action of a uniform localized magnetic field proposed by Tzirtzilakis and Loukopoulos [1] led to unstable numerical solution by using the ANSYS-FLUENT environment at high dimensionless

magnetic numbers. In other words, the computational benchmark problem of Tzirtzilakis and Loukopoulos [1] might be difficult to investigate and further validate with a commercial software package even though their research work contains a relevant study on the the interaction between the temperature and MHD-FHD fields by considering blood as a Newtonian fluid. As it has been mentioned above that there is a lack of experimental data available on investigation of the interaction between the temperature and MHD-FHD fields for non-Newtonian blood flows, the computational benchmark presented in this subsection could be used in both in-house and commercial software package environment to validate the implementation of an MHD-FHD model for non-isothermal non-Newtonian blood flows.

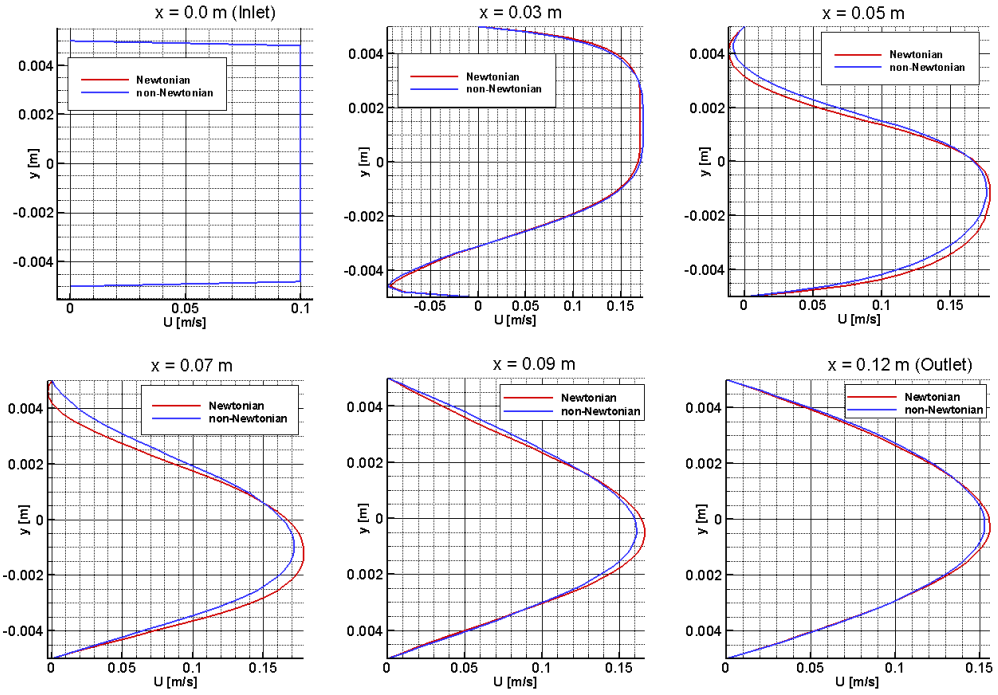


Figure 11: Comparison of velocity profiles for a non-isothermal blood flow in a microfluidic channel at different cross-sections when the applied magnetic field is induced by a single wire.

In the present benchmark problem, the magnetic field is applied through a single wire located in a two-dimensional microfluidic channel and when the temperature dependence of the magnetization is neglected, the effect of MHD and FHD appears in the momentum equation (10) and the temperature equation (11) is only affected by the source term due to MHD. In order to avoid observed numerical oscillations and possible non-physical solution for the benchmark problem presented in this subsection, the MHD-FHD model equations have been set up to cancel out the magnetocaloric effect of FHD on the right hand side of the energy equation (3) which would be a numerically sensitive term in the temperature equation if the magnetization would depend on the temperature field as well. It can be seen in Figure 13 that when the incompressible and non-isothermal flow was influenced by an applied single wire magnetic field along with the temperature independent magnetization, the numerical convergence of the approximate solution of the fully coupled governing equations was ensured for both Newtonian and non-Newtonian blood rheological approaches. In other words, the combination of existing MHD and FHD models [2, 7] proposed in this paper leads to an appropriate computational benchmark test case to

ensure convergence and accurate solution of MHD-FHD model equations with a convergence criterion smaller than 10^{-7} (see Figure 13) up to the magnitude of the applied magnetic field strength of $4\,Tesla$ (see Figures 8 to 12).

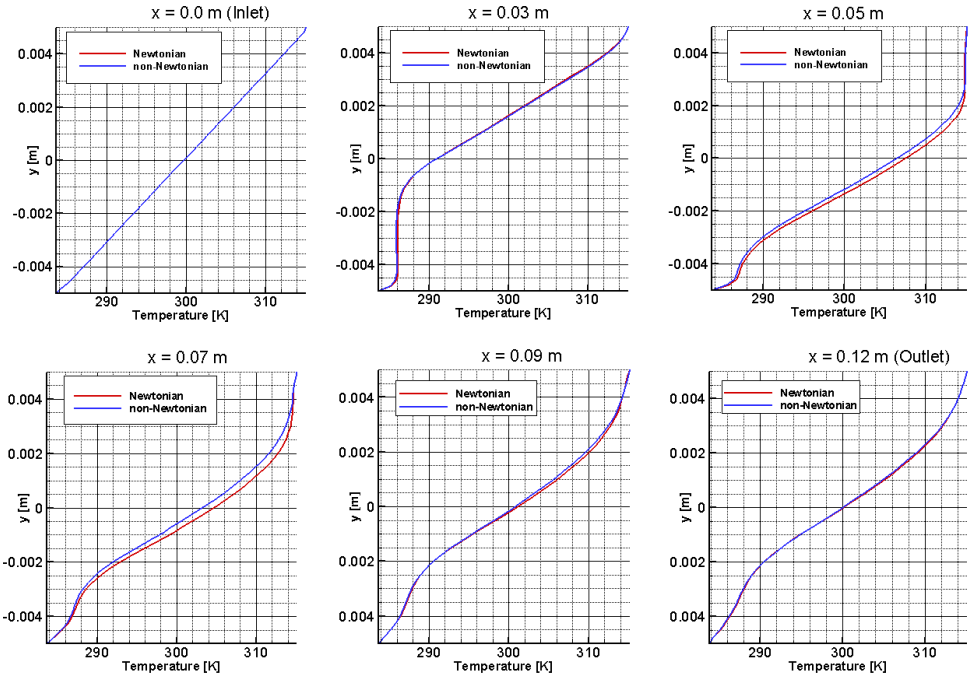


Figure 12: Comparison of temperature profiles for a non-isothermal blood flow in a microfluidic channel at different cross-sections when the applied magnetic field is induced by a single wire.

For incompressible and non-isothermal flow influenced by an applied single wire magnetic field for Newtonian and non-Newtonian fluids, the velocity and temperature profiles at different locations of the two-dimensional microfluidic channel have been shown in Figures 11 and 12, respectively. The velocity and temperature fields are strongly affected by the location area of the applied magnetic field through a single wire as well as the downstream location behind the wire (see Figures from 9 to 12). The velocity field of the incompressible, laminar and non-isothermal flow becomes fully developed far from the applied magnetic field area. It can be seen in Figure 12 that the outlet velocity profile is not affected by the magnetic field strength, thus the profile is similar to the fully developed laminar velocity distribution for both Newtonian and non-Newtonian fluids. The temperature field behaviour is different compared to the non-isothermal Hartmann flow due to the fact that the effects of MHD and FHD terms in the momentum equation (10) has impact on the temperature distribution through convection even if the magnetocaloric effect of the FHD term does not appear in the temperature equation (11). The outlet temperature profile (see Figure 12) exhibits a non-linear behaviour compared to the outlet temperature distribution of the non-isothermal Hartmann flow whereas the profile becomes quasi-linear at the outlet section of the microfluidic channel (see Figure 6). For steady-state Newtonian and non-Newtonian fluid flows, the numerical solution converged to a small threshold value of 10^{-7} by performing 450 total number of iterations. For this numerical test case, approximately 3 times more total number of iterations were required compared to the non-isothermal Hartmann flow simulation, because the magnetization force due the effect of FHD has been taken into account on the right hand side of the momentum equation (10).

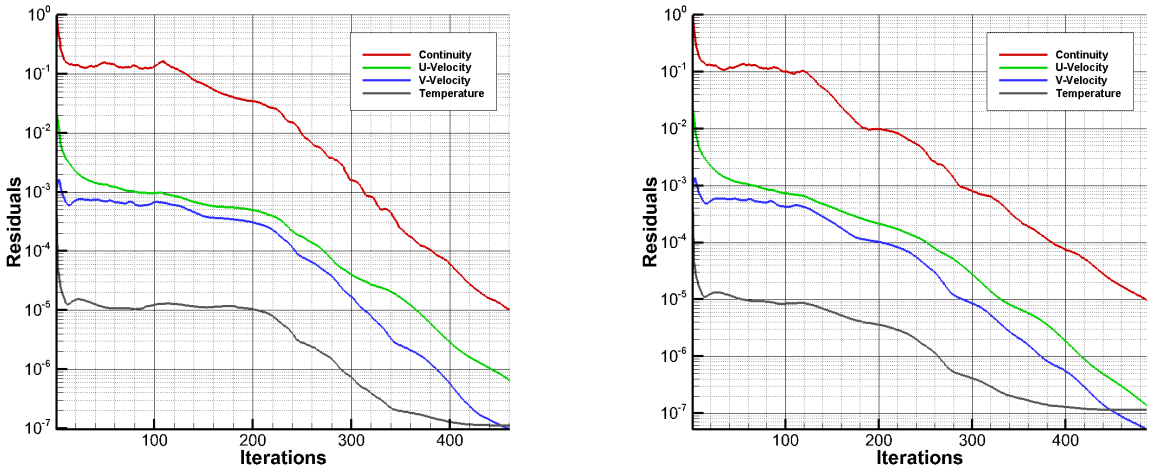


Figure 13: Residual history of non-isothermal blood flow simulations for a single wire problem: Newtonian fluid flow (left) and non-Newtonian blood flow (right).

The computational benchmark problem presented in this subsection could be considered as a reliable verification and validation test case for non-isothermal Newtonain and non-Newtonian MHD-FHD fluid flows when microfluidic applications are concerned, because the numerical solution was converged to a very small threshold value of 10^{-7} within a relatively small total number of iterations. Furthermore, the obtained results are well-explainable comparing them to the non-isothermal Hartmann flow. It is important to mention again the limitation of the current implementation, because it has been observed relying on numerical experience that when the applied magnetic field strength is higher than $4\,Tesla$, it was difficult to ensure convergence of the numerical solution for both Newtonian and non-Newtonian fluids. In consequence of this, it is recommended to employ higher than second-order treatment of the convective and/or MHD-FHD terms in the system of governing equation in order to overcome the oscillatory numerical behaviour of the stiff MHD and FHD source terms at high dimensionless magnetic numbers.

3.3 Double wire problem

In this subsection, a benchmark problem has been presented for an incompressible, laminar and non-isothermal MHD-FHD blood flow in conjunction with Newtonian and non-Newtonian rheological models when the fluid flow is influenced by a double wire arrangement of the applied magnetic field strength. Similar to the numerical example presented in the previous subsection, a wire is located at the spatial coordinate $x_1 = 0.03\,m$ on the lower wall in a two-dimensional microfluidic channel where the magnetic field strength is taken as $4\,Tesla$. In addition to this, another wire is located at the spatial coordinate $x_2 = 0.07\,m$ on the upper wall where the magnetic field strength is equal to $1.2\,Tesla$. The set of governing equations are relying on the mathematical model equations proposed by Tzitzilakis [7], however, we define the viscous stress tensor $\underline{\underline{\tau}}$ according to the constitutive equation related to a viscous Newtonian and a power-law non-Newtonian fluid model for non-isothermal blood flows. This benchmark test case represents a more complicated multiphysics problem in conjunction with MHD-FHD flows which triggers off vortical structures in a two-dimensional microfluidic channel similar to the benchmark problem investigated by Tzirtzilakis and Loukopoulos [1] in the presence of

a localized uniform magnetic field. For this multiphysics problem, we assume again that the magnetic state through the magnetization equation $\mathbf{M} = \mathbf{M}(\mathbf{H})$ depends only on the magnetic field intensity $\mathbf{H}$ and it is independent from the temperature field T. It means that the effects of MHD and FHD have been taken into account in the momentum equation (10) and the FHD effect have been neglected in the temperature equation (11). For an incompressible, laminar and non-isothermal non-Newtonian blood flow, the velocity and temperature field contours have been shown in Figures 14 and 15, respectively.

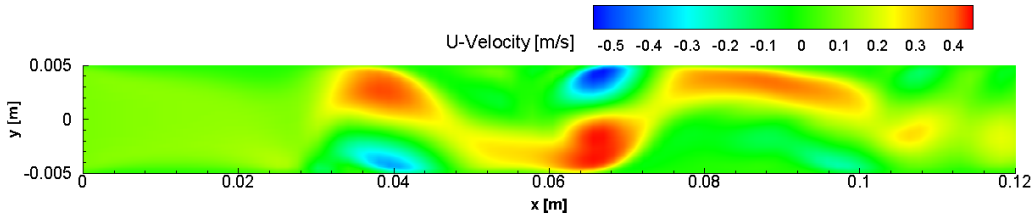


Figure 14: Contours of the velocity component U for a non-isothermal non-Newtonian blood flow in a microfluidic channel with a uniform inlet velocity distribution when the applied magnetic field is induced by a double wire.

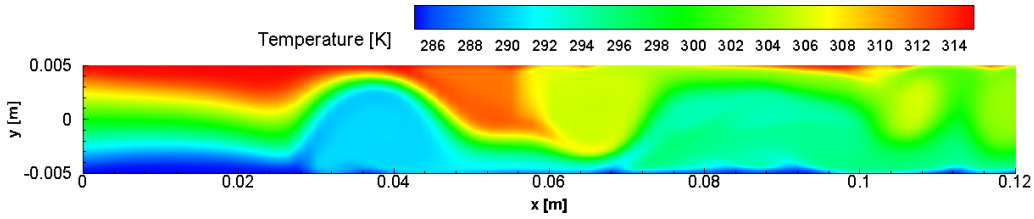


Figure 15: Contours of the temperature field for a non-isothermal non-Newtonian blood flow in a microfluidic channel with a linear inlet temperature distribution when the applied magnetic field is induced by a double wire.

Tzirtzilakis and Loukopoulos [1] investigated a non-isothermal Newtonian flow under the action of a uniform localized magnetic field of $8\,Tesla$ which was two times more than the magnitude of the applied magnetic field through a wire for the benchmark problem presented in this subsection. Furthermore, they took into account the magnetocaloric effect of FHD in the energy equation which could be a stiff source term in the set of governing equations. The streamfunction-vorticity formulation was used by Tzirtzilakis and Loukopoulos [1] for solving the MHD-FHD model equations along with a line by line implicit method. Their numerical solution was reduced to solve the discretized system of equations by using the Thomas algorithm which also means that this numerical approach was basically a one-dimensional solution of the discretized governing equations along the axial direction, and the effect of the other direction was taken into account on the right hand side of the system of linear equations. It might be important to highlight that the numerical solution of a fully coupled MHD-FHD governing equations by using an approximate one-dimensional line by line approach is not equivalent to a fully two-dimensional explicit or implicit numerical solution to the set of MHD-FHD model equations. Therefore the convergence properties in conjunction with the presence of stiff

MHD-FHD source terms could lead to different numerical behaviours when various methods are used for the same computational benchmark problem. According to our numerical experience related to a fully two-dimensional computational approach, the relatively stiff MHD-FHD source terms, implemented in the ANSYS-FLUENT environment, caused numerical instabilities when the magnitude of the applied uniform localized magnetic field was $8\,Tesla$. In other words, the fully two-dimensional numerical solution of the benchmark problem of Tzirtzilakis and Loukopoulos [1] in the ANSYS-FLUENT environment was not converged to the threshold value of 10^{-5} as it was prescribed in [1] for an approximate line by line solution method.

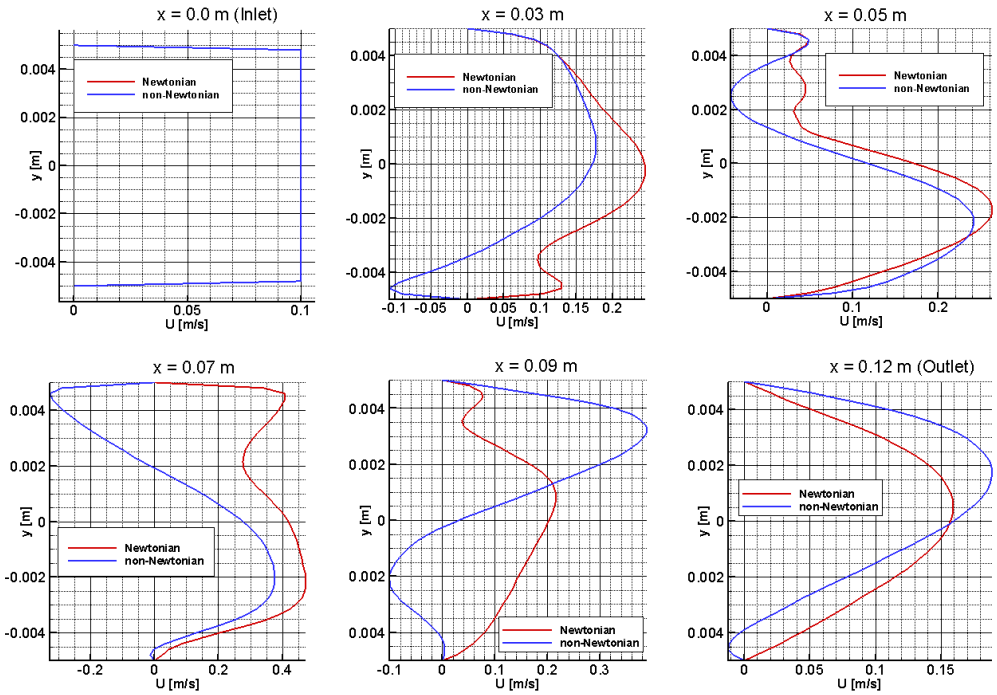


Figure 16: Comparison of velocity profiles for a non-isothermal blood flow in a microfluidic channel at different cross-sections when the applied magnetic field is induced by wires located on the lower and upper walls.

For an incompressible, laminar and non-isothermal MHD-FHD flow influenced by an applied double wire magnetic field for Newtonian and non-Newtonian fluids, the velocity and temperature profiles at different locations of the microfluidic channel have been shown in Figures 16 and 17, respectively. The reason for the numerical investigation presented in this subsection is that the double wire magnetic field arrangement has to trigger off a counter-clockwise and another clockwise rotating vortex at the location of the wires which leads to a similar vortical flow pattern to the benchmark problem of Tzirtzilakis and Loukopoulos [1] even if the fluid has to flow through a non-uniform applied magnetic field. The velocity and temperature fields are strongly affected at the locations of the applied magnetic field along with the flow upstream and downstream as well. A strong circulation regime can be observed closed to the magnetic field sources and the perturbation of the velocity field is significant about the wire located on the upper wall of the channel. When the fluid moves upstream, the flow influenced the fluid motion close to the wire located on the lower wall. The fluid flow downstream exhibits strong mixing due to the effect of the wire located on the upper wall. The difference between the Newtonian and non-Newtonian blood rheology approaches can be seen in Figures 16 and 17.

In both cases, the velocity profile does not become fully-developed at the outlet section of the channel which is also valid for the temperature field. It also means that the flow supposed to be become fully-develop far downstream from the second wire located on the upper wall which could occur in a longer microfluidic channel. The difference of velocity and temperature distributions between Newtonian and non-Newtonian fluids becomes significant at the downsteam flow regime and at the outlet section of the channel. This phenomenon can be explained by supposing that the non-linear viscous effect of non-Newtonian blood flows turns out to be dominant when the magnetic field intensity is increased which comes from the material behaviour of blood. The difference could also arise purely due to the unstable numerical behaviour of the non-Newtonian blood flow in the presence of a strong magnetic field, because the numerical solution was not converging to a small threshold value for the steady-state non-Newtonian fluid flow in this case, thus the residual history appears to be strongly oscillatory (see Figure 18).

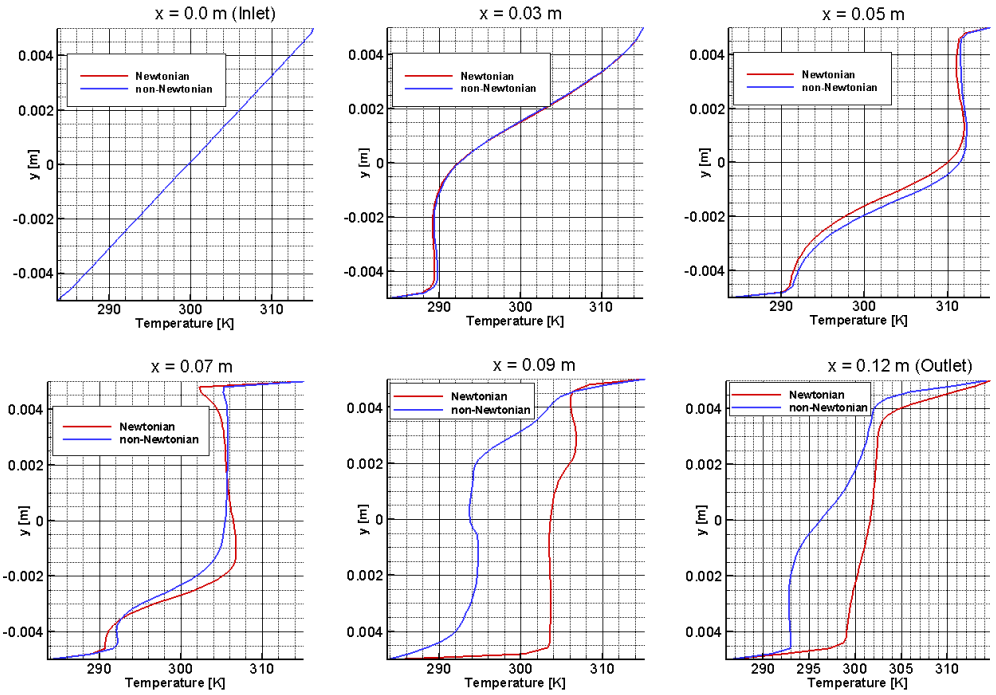


Figure 17: Comparison of temperature profiles for a non-isothermal blood flow in a microfluidic channel at different cross-sections when the applied magnetic field is induced by wires located on the lower and upper walls.

For the investigated steady-state Newtonian fluid flow in a microfluidic channel, the numerical solution of continuity and momentum equations was converging to a small threshold value of 10^{-14} while the solution of the temperature equation was converging up to the threshold value of 10^{-4} by performing 4000 total number of iterations. For non-Newtonian blood flow, the numerical solution exhibited strongly oscillatory behaviour (see Figure 18). Furthermore, this benchmark problem required the highest total number of iterations compared to the previous two numerical examples presented in this paper. It can also be seen relying on the residual histories (see Figure 18) that this benchmark could be used for MHD-FHD code verification and model validation for Newtonian fluids, however it might be considered as a non-conclusive example for non-Newtonian blood flows when the power-law fluid model is employed.

Relying on the obtained results, it can be seen that the simulation of non-isothermal MHD-

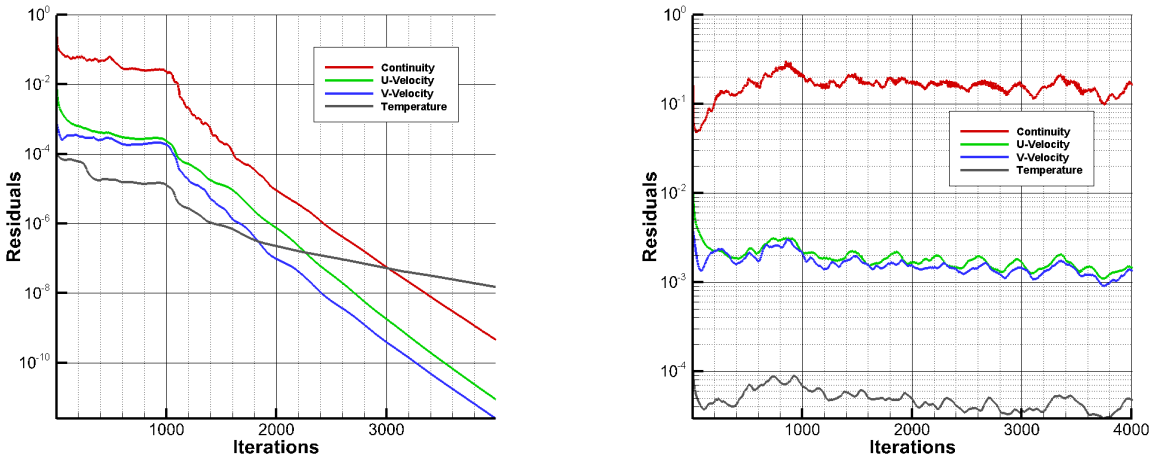


Figure 18: Residual history of non-isothermal blood flow simulations for a double wire problem: Newtonian fluid flow (left) and non-Newtonian blood flow (right).

FHD flows in conjunction with Newtonian and non-Newtonain blood rheological approaches for microfluidic and biomedical applications could be challenging. This is due to the fact that the stiff MHD and FHD source terms in the governing equations including the non-linear viscous behaviour of non-Newtonain blood flows could lead to oscillatory numerical solutions when the applied magnetic field strength is a high value. Therefore the behaviour of other numerical solution methods in conjunction with non-isothermal non-Newtonian blood flows has to be further investigated which could be a subject of another study.

4 CONCLUSIONS

In this paper, we proposed an incompressible, laminar and non-isothermal MHD-FHD model for Newtonian and non-Newtonian rheological models of blood relying on the combination of existing MHD-FHD approaches of Tzirakis et al. [2] and Tzirtzilakis [7]. This paper focused on the implementation of the proposed MHD and FHD models in the ANSYS-FLUENT environment for a) non-isothermal Newtonian and non-Newtonian Hartmann flows in the presence of an applied constant magnetic field and b) non-isothermal non-Newtonian blood flows relying on the power-law fluid model when the flow field was influenced by a single and double wire arrangement of an applied magnetic field in a microfluidic planar channel. The objective of this work was to provide reliable computational benchmark problems for CFD code verification and MHD-FHD model validation purposes. The following conclusions can be drawn as

- For solving an incompressible steady-state non-isothermal Hartmann flow problem in conjunction with Newtonian and non-Newtonian blood rheological models, the numerical solution of the MHD governing equations was converging to a small threshold value of 10^{-8} within 145 number of iterations. The obtained results have been compared to the computational data of Tzirakis et al. [2] taken from the literature. This benchmark problem can be considered as a reliable benchmark test case for code verification and MHD model validation purposes for non-isothermal non-Newtonian blood flows.

- When a single wire induced magnetic field was located on the lower wall of the microflu-

idic channel for solving an incompressible steady-state non-isothermal MHD-FHD flow problem in conjunction with Newtonian and non-Newtonian blood rheological models, the numerical solution of the MHD-FHD governing equations was converging to a small threshold value of 10^{-7} within 450 total number of iterations. This validation test case can also be considered as a reliable computational benchmark to investigate the interaction between the temperature and MHD-FHD fields for non-Newtonian blood flows, because there is a lack of experimental data available in the literature.

- When a double wire arrangement of the applied magnetic field was located on the lower and upper walls of the microfluidic channel for solving an incompressible steady-state non-isothermal MHD-FHD flow problem related to Newtonian and non-Newtonian blood rheological models, the numerical solution of the MHD-FHD governing equations exhibited strongly oscillatory behaviour for non-Newtonian blood flows. Relying on these results, this benchmark case could only be used for MHD-FHD code verification and model validation for Newtonian fluids, however, it might be considered as a non-conclusive example for non-Newtonian blood flows when the power-law fluid model is employed.

- For solving incompressible non-isothermal MHD-FHD problems, it has been observed that the limitation of the current implementation is that when the applied magnetic field strength was higher than $4\,Tesla$, it was difficult to ensure convergence of the numerical solution for both Newtonian and non-Newtonian fluids. This limitation can be explained by the fact that the MHD-FHD source terms are stiff source terms in the momentum and energy equations, therefore the numerical behaviour of higher-order methods [11] should be further investigated for MHD-FHD problems at high dimensionless magnetic numbers.

REFERENCES

[1] Tzirtzilakis, E. E., Loukopoulos, V. C., Biofluid flow in a channel under the action of a uniform localized magnetic field. *Computational Mechanics, Springer*, **36**(5), 360–374, 2005.

[2] Tzirakis, K., Papaharilaou, Y., Giordano, D., Ekaterinaris, J., Numerical investigation of biomagnetic fluids in circular ducts. *International Journal for Numerical Methods in Biomedical Engineering*, **30**(3), 297–317, 2014.

[3] Tzirakis, K., Botti, L., Giordano, D., Papaharilaou, Y., Numerical modelling of non-Newtonian polarizable and magnetizable fluid flow. *8th European Symposium on Aerothermodynamics for Space Vehicles, 2-6 March 2015, IST Congress Centre*, Lisbon, Portugal, p. 152 (Abstract), Full Paper Published Online (http://www.congrexprojects.com), 2015.

[4] Tzirtzilakis, E. E., Sakalis, V. D., Kafoussias, N. G., Hatzikonstantinou, P. M., Biomagnetic fluid flow in a 3D rectangular duct, *International Journal for Numerical Methods in Fluids*, **44**(12), 1279–1298, 2004.

[5] Loukopoulos, V. C., Tzirtzilakis, E. E., Biomagnetic channel flow in spatially varying magnetic field. *International Journal of Engineering Science*, **42**, 571–590, 2004.

[6] Könözsy, L., Multiphysics CFD Modelling of Incompressible Flows at Low and Moderate Reynolds Numbers, *Ph.D. Thesis*, *Centre for Fluid Mechanics and Computational Science*, *Cranfield University*, Cranfield, Bedfordshire, MK43 0AL, UK, 2012.

[7] Tzirtzilakis, E. E., A mathematical model for blood flow in magnetic field. *Physics of Fluids*, **17**, 077103, 2005.

[8] Neofytou, P., Drikakis, D., Non-Newtonian flow instability in a channel with a sudden expansion. *Journal of Non-Newtonian Fluid Mechanics*, **111**, 127–150, 2003.

[9] Scienza, P., Numerical Investigation of Non-Newtonian Biomagnetic Fluid Flows in Presence of Localized Magnetic Field, *M.Sc. Thesis*, *Centre for Fluid Mechanics and Computational Science*, *Cranfield University*, Cranfield, Bedfordshire, MK43 0AL, UK, 2015.

[10] ANSYS Inc., ANSYS FLUENT Theory Guide. *ANSYS*, Release 14.0, November, 2011.

[11] Könözsy, L., Drikakis, D., A Unified Fractional-Step, Artificial Compressibility and Pressure-Projection Formulation for Solving the Incompressible Navier-Stokes Equations. *Communications in Computational Physics*, **16**(5), 1135–1180, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL STUDY OF MAGNETIC PARTICLES CONCENTRATION IN BIOFLUID (BLOOD) UNDER THE INFLUENCE OF HIGH GRADIENT MAGNETIC FIELD IN MICROCHANNEL

V. C. Loukopoulos[1], G. C. Bourantas[2], D. Labropoulos[3], V. M. G. Nikiforidis[3] S. P. A. Bordas[2,4,5] and G. C. Nikiforidis[3]

[1] Department of Physics, University of Patras
Patras, GR 26500, Rion, Greece
e-mail: vxloukop@physics.upatras.gr

[2] Faculty of Science, Technology and Communication, University of Luxembourg
Campus Kirchberg, 6 rue Richard Coudenhove-Kalergi L-1359, Luxembourg
{georgios.bourantas,stephane.bordas}@ uni.lu

[3] Department of Medical Physics, School of Medicine, University of Patras
GR 26504,Rion, Greece
dimlab@upatras.gr, martin-nikif@hotmail.com,gnikif@med.upatras.gr

[4] School of Engineering, Cardiff University
The Parade, CF24 3AA Cardiff, United Kingdom

[5] Intelligent Systems for Medicine Laboratory, School of Mechanical and Chemical Engineering, The University of Western Australia
Stirling Highway, Crawley/Perth, WA 6009, Australia

Keywords: Biomagnetic Separation, Drug Delivery, Hyperthermia Treatment, Non Newtonian, Two-phase Flow, Magnetophoretic Mass Transfer; Molecular Diffusion, Meshless.

Abstract. *A meshless numerical scheme is developed in order to simulate the magnetically mediated separation of biological mixture used in lab-on-chip devices as solid carriers for capturing, transporting and detecting biological magnetic labeled entities, as well as for drug delivering, magnetic hyperthermia treatment, magnetic resonance imaging, magnetofection, etc. Due to the imposed magnetic field stagnation regions are developed, leading to the accumulation of the magnetic labeled species and finally to their collection from the heterogeneous mixture. The role of (i) the intensity of magnetic field and its gradient, (ii) the position of magnetic field, (iii) the magnetic susceptibility of magnetic particles, (iv) the volume concentration of magnetic particles and their size, (v) the flow velocity in the magnetic fluidic interactions and interplay between the magnetophoretic mass transfer and molecular diffusion are thoroughly investigated. Both Newtonian and non-Newtonian blood flow models are considered.*

1 INTRODUCTION

Biomagnetic fluid dynamics (BFD) investigates the dynamics of biological fluids under the presence of magnetic field. BFD accounts in numerous bioengineering applications [1-3] and there are a number of mathematical models proposed regarding biomagnetic fluid flow under the action of an applied magnetic field.

The governing equations for incompressible fluid flow are similar to those used in ferrohydrodynamics (FHD). In [3, 4] the viscous, steady, two-dimensional, incompressible, laminar biomagnetic fluid (blood) flow between two parallel flat plates (channel) was studied, with the fluid considered as homogeneous. Moreover, magnetic fluids are used in medicine, especially in high-gradient magnetic separation (HGMS), magnetic drug targeting (MDT) and ferrofluid sealing [5]. The flow model in [2-4], assumes that suspended particles and base fluid constitute a stable colloidal suspension with infinitely strong coupling, in which the particles are not amenable to re-distribution or drifting relative to the base fluid under the effect of an external magnetic force. Thus, the fluid-particles mixture interacts with the magnetic force as a whole fluid. On the other hand, the model employed in [5-9], accommodates magnetophoretic drifting relative to the base fluid and therefore allows particles re-distribution when an advection-diffusion particle concentration is coupled with the Navier-Stokes equations. In comparison, the traditional BFD and FHD studies treat the volumetric magnetic body force as simply that of a single-phase flow, while the [5-9] studies treat it as of a two-phase flow where each phase can exhibit its distinctive magnetic response. One-way particle-fluid coupling approach (Eulerian) was adopted in [5-8] and two-way particle-fluid coupling approach (Lagrangian-Eulerian) in [9].

In the present study we investigate a two phase flow of a non-Newtonian biomagnetic fluid between two horizontal parallel plates, under the action of a non uniform magnetic field and we examine the role of (*i*) the intensity of magnetic field and its gradient, (*ii*) the position of magnetic field, (*iii*) the magnetic susceptibility of magnetic particles, (*iv*) the volume concentration of magnetic particles (nanoparticles) and their size, and (*v*) the flow velocity.

2 EQUATIONS AND MATHEMATICAL MODEL

A non-uniform magnetic field is applied to microchannel flow containing magnetically tagged microspheres. A two-dimensional rectangle microchannel is considered with width (W) and length (L) equal to 1 mm and 20 mm respectively. Assumptions as in [6, 7] are adopted in the present study.

The governing dimensional equations of the isothermal fluid flow in x- and y- direction are as follows:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0\,, \tag{1}$$

$$\rho\left(u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}\right)=-\frac{\partial p}{\partial x}+\mu\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)+F_x\,, \tag{2}$$

$$\rho\left(u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}\right)=-\frac{\partial p}{\partial y}+\mu\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+F_y\,, \tag{3}$$

where F_x and F_y are the volume forces in the x- and y- directions, respectively and represent the effect of magnetic particles concentration, while ρ and μ are the density and the viscosity

of the biofluid (blood). Magnetic particles are considered as super-paramagnetic material without permanent magnetization at room temperature. Velocity components in the *x*- and *y*-direction are defined by *u* and *v*, respectively. The volume forces (F_x and F_y) are equal to the magnetic forces applied on a single particle ($F_{mag,x}$ and $F_{mag,y}$), multiplied by the number of particles per unit volume ($F_x = n_p\left(x,y\right)F_{mag,x}$ and $F_y = n_p\left(x,y\right)F_{mag,y}$). They are defined as [5-8]:

$$\boldsymbol{F}_{mag} = \frac{1}{2}V_p\mu_0\chi\nabla H^2, \tag{4}$$

with *H*, μ_0 and χ being the external magnetic field intensity, the magnetic permeability of vacuum and the magnetic susceptibility of the particles, respectively (p subscript refers to particles). The dimensionless concentration ($C\left(x,y\right) = n_p\left(x,y\right)/n_0$) is defined with respect to the reference particle density η_0 at the flow entrance. In addition, the non-dimensional volume concentration (*C*) is written as [5-8]

$$\nabla\left(Cv_p\right) = \nabla\left(D\nabla C\right). \tag{5}$$

In Eq. (5), *D* is the diffusion coefficient and the velocity of particles $\boldsymbol{v}_p$ is calculated by balancing the hydrodynamic and magnetic forces [6, 7]

$$\boldsymbol{v}_p = \boldsymbol{v}_f + \boldsymbol{v}_{mag}, \tag{6}$$

where $\boldsymbol{v}_{mag} = \boldsymbol{F}_{mag}/6\pi\mu r_p$ is given by Stokes drag law, $\boldsymbol{v}_f = u\boldsymbol{i} + v\boldsymbol{j}$ is the velocity of the fluid computed by Eqs. (1) and (2) and, r_p is the particle radius.

The boundary conditions are defined as:

- for the velocity components: at inlet fully developed flow is considered ($u=u(y)$, $v=0$), no-slip boundary conditions are imposed on the upper and lower wall and fully developed flow at the outlet ($u_{,x}=0$, $v=0$).
- for the dimensionless concentration: concentration $C_{in}=1$ at the inlet, insulated on the upper and lower walls and, convection at the outlet.

Since the viscosity of a non-Newtonian fluid (blood) is a function of shear rate, a relation between the viscosity and the shear rate is required, (shear-thinning viscosity of the blood). The relation between the shear stress τ and the shear rate $\dot{\gamma}$ is as $\tau = \mu[\dot{\gamma}]$. The viscosity is given by the power law model $\mu = m\dot{\gamma}^{n-1}$ [10, 11].

3 RESULTS AND DISCUSSION

3.1 Verification of the numerical results and grid independent test

A Meshless Point Collocation (MPCM) scheme along with the Moving Least Squares (MLS) approximation method is applied for the numerical solution of the governing equations. An iterative scheme, developed and presented in [12], was used to numerically solve the velocity-vorticity formulation of the flow equations. Mass conservation (continuity equation) was ensured enforcing a velocity-correction scheme. The accuracy of the proposed scheme was verified against the numerical results obtained by a Meshless Point Collocation method (MPCM) along with the IPOT (Implicit Potential) and DCP SE (Discretization Correction Particle Strength Exchange) approximation for the field variables for the numerical solution

of the model equations in the velocity-pressure formulation [13], Fig. 1. The magnetic field intensity is $1x10^6\,A\,m^{-1}$. The base fluid is blood with density $800\,kg\,m^{-3}$ [8] and viscosity given by the power law (m=0.035 and n=0.6 [10, 11]), in Poise unit. We assume that the target biological entity and the attached magnetic particles constitute a single microspherical particle with $2\mu m$ radius [6-8]. The effective magnetic susceptibility χ (relative to that of magnetic particles) is equal to 0.004, volume fraction φ is equal to 0.03, the mean inlet velocity is $0.001\,m\,s^{-1}$ (Plug flow) and the Particle diffusivity D is $1\times10^{-9}\,m^2s^{-1}$.

An extensive mesh-refinement procedure has been conducted, in order to ensure a grid-independent numerical solution. The numerical results presented are obtained using a 31x20x31 uniform nodal distribution.

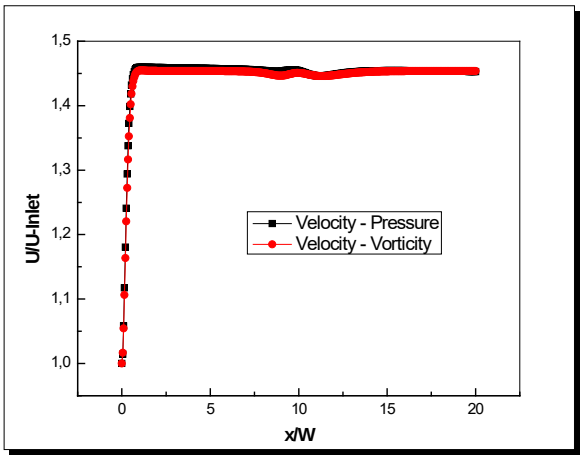


Figure 1: Verification.

3.2 Discussion

Figure 2 depicts velocity magnitude of the Newtonian and non-Newtonian fluid (blood mixture) along the channel central line, considering mean velocity inlet, magnetic particles diameter and magnetic susceptibility equal to $0.001\,m\,s^{-1}$, $2\mu m$ and 0.004, respectively.

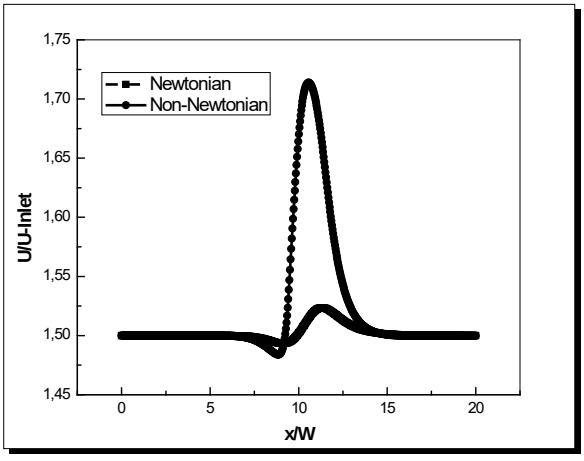


Figure 2: Non-dimensional velocity profile at the central line of the channel for Newtonian and non-Newtonian behavior of the magnetic fluid (blood) mixture.

Figure 3 shows the stream function and the particles concentration contours for mean inlet velocity equal to $0.001\,m\,s^{-1}$, magnetic particles diameter $2\mu m$ and magnetic susceptibility 0.004, considering two different positions for the magnetic source, that is, b= 4 and 3 mm. We can observe that a secondary vortex close to the upper plate is formed when b=4 mm, while a secondary vortex close to the lower plate is formed when b=3 mm.

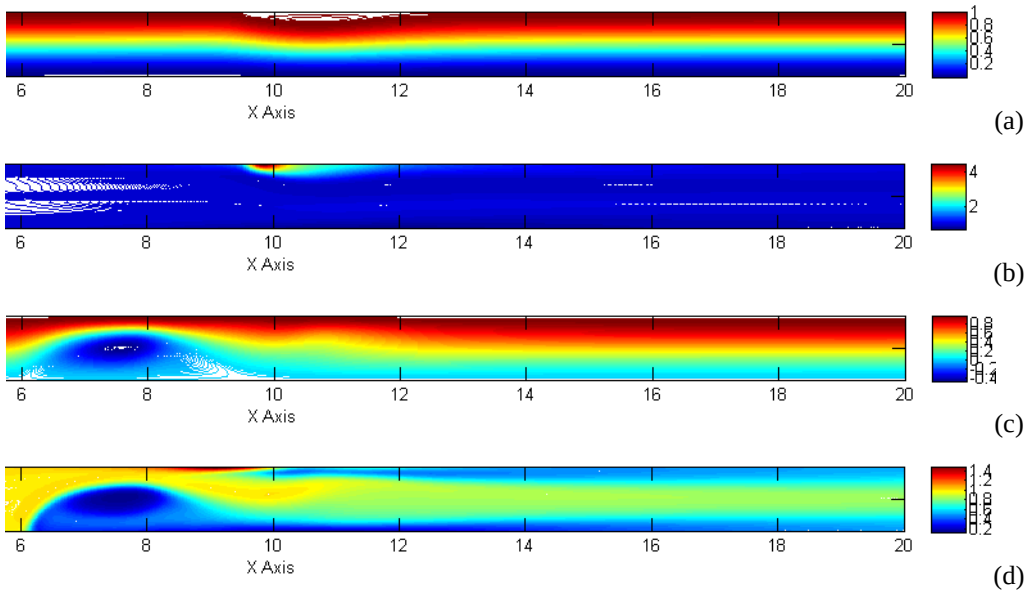
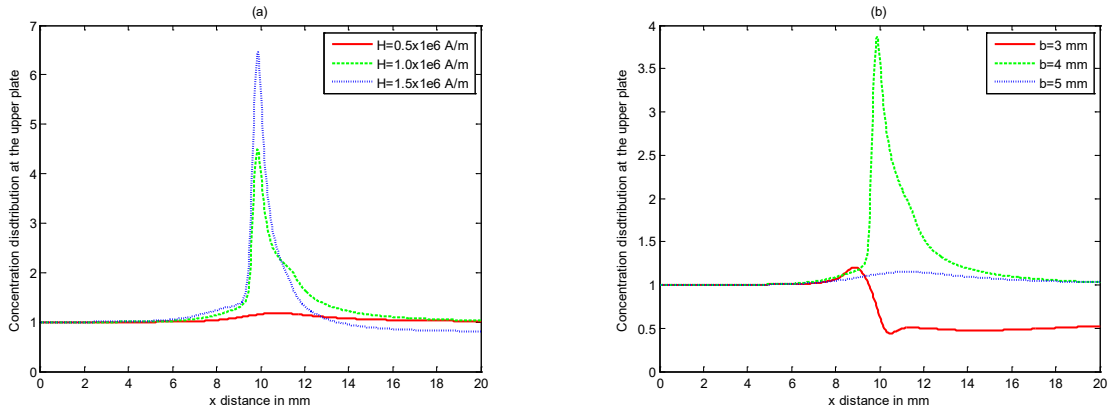


Figure 3: Stream function contours and concentration contours for mean velocity inlet, magnetic particles diameter and magnetic susceptibility equal to $0.001\,m\,s^{-1}$, $2\mu m$ and 0.004 when b=4 (a, b) and 3 mm (c, d).

In order to investigate the performance of the two-phase fluid flow (blood and solid magnetic particles), computations have been carried out for different parameter values. Blood viscosity values were given by the power law.

Figure 4 shows the variation of the dimensionless particle concentration at y=0.97 mm (close to the upper plate) along x- direction (0 mm, 20 mm) according to (*i*) the intensity of magnetic field and its gradient, (*ii*) the position of magnetic field, (*iii*) the magnetic susceptibility of magnetic particles, (*iv*) the volume concentration of magnetic particles (nanoparticles) and their size, (*v*) the flow velocity in the magnetic fluidic interactions, while the interplay between the magnetophoretic mass transfer and molecular diffusion is also investigated.



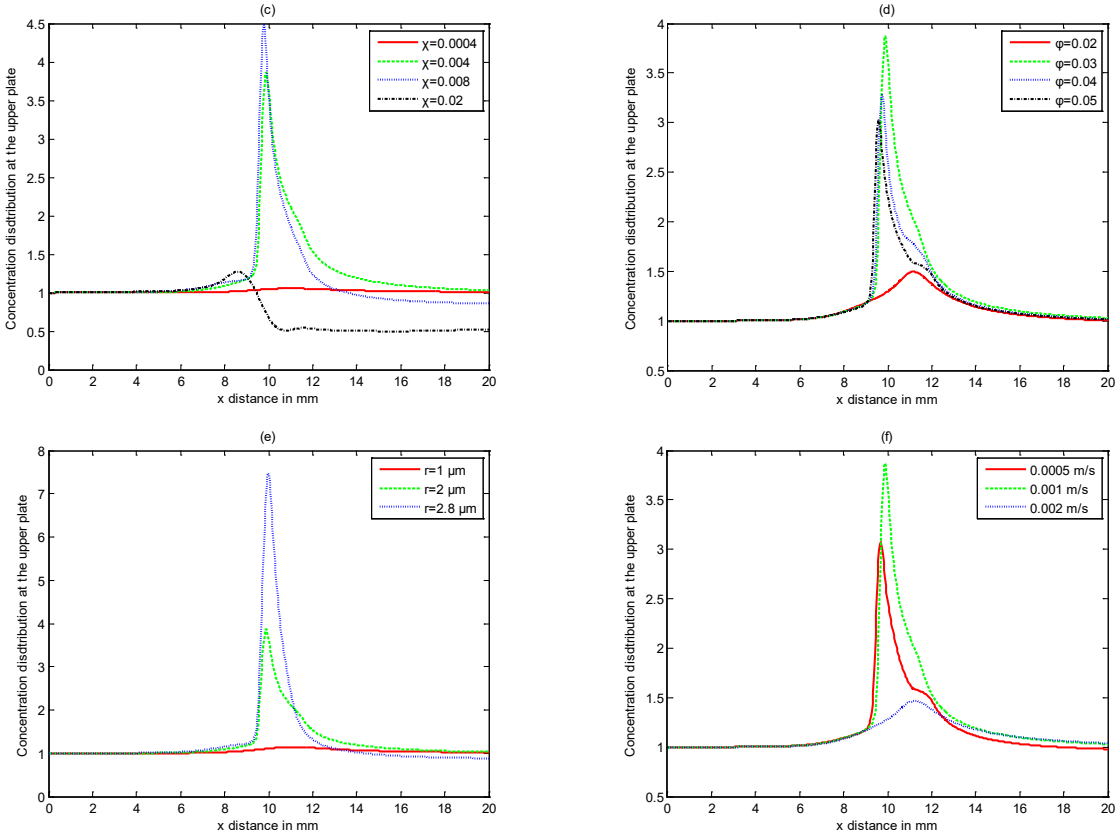


Figure 4: The variation of the dimensionless particle concentration at y= 0.97 mm (close to the upper plate) along x direction (0 mm, 20 mm) according to (a) the intensity of magnetic field and its gradient, (b) the position of magnetic field, (c) the magnetic susceptibility of magnetic particles, (d) the volume concentration of magnetic particles (nanoparticles), (e) the size of magnetic particles, (f) the flow velocity.

Figure 4a presents the variation of the dimensionless particle concentration according to the intensity of magnetic field and its gradient. The base fluid is assumed to be blood with density $800\,\mathrm{kg\,m^{-3}}$ [8] and viscosity computed from the power law (m=0.035 and n=0.6 [11]), in Poise. The target biological entity and the attached magnetic particles consist a single micro-spherical particle with 2μm radius [8]. The effective magnetic susceptibility χ (relative to that of magnetic particles) is equal to 0.004, the volume fraction φ is equal to 0.03 and the mean inlet velocity is equal to $0.001\,\mathrm{m\,s^{-1}}$. The center of the permanent magnet is located at the midpoint of the plate ($a = 10\,\mathrm{mm}$), 4 mm far from the lower plate ($b = 4\,\mathrm{mm}$). Particle diffusivity D is assumed $1\times10^{-9}\,\mathrm{m^2 s^{-1}}$. For the calculations presented in Fig. 4, density, viscosity and particle diffusivity are constants. As the magnetic field intensity increases ($0.5\mathrm{x}10^{6}\,\mathrm{A\,m^{-1}}$, $1\mathrm{x}10^{6}\,\mathrm{A\,m^{-1}}$, $2\mathrm{x}10^{6}\,\mathrm{A\,m^{-1}}$), the concentration increases near the point (10 mm, 1 mm) with the maximum value of the magnetic field for the flow field domain concentration. This means that increasing the intensity of the magnetic field the attractive magnetic force is dominant (magnetophoresis) over its counter tendency to diffuse upward into regions with less concentration, which eventually results to particle redistribution on the close vicinity of the point with highest magnetic strength.

In order to understand the flow behavior and the concentration contribution with varying position of the magnetic field, the virtual source is placed at 3 mm, 4 mm and 5 mm above the midpoint of the lower plate, as shown in Fig. 4b. It can be seen that the concentration increas-

es as the magnetic source is moving towards the microchannel, until the point where the flow profile is changed. Following, the vortex formed close to the upper plate disappears and a new vortex forms close to the lower plate, as in Fig. 3c. For the computations the magnetic field intensity is $1x10^{6}\,A\,m^{-1}$.

Figure 4c presents the variation of the dimensionless particle concentration according to the magnetic susceptibility of magnetic particles. The source is positioned at (10 mm, 4 mm) and the magnetic field intensity is $1x10^{6}\,A\,m^{-1}$. As the magnetic susceptibility increases (0.0004, 0.004, 0.008, 0.02), the concentration increases near the place (10 mm, 1 mm) where the magnetic field gets its maximum value, up to the point where the flow profile is changing and the vortex, formed close to the upper plate, disappears and a new vortex forms closer to the lower plate, as in Fig. 3c.

Figure 4d presents the variation of the dimensionless particle concentration according to the volume concentration of magnetic particles (nanoparticles) for volume fraction 0.02, 0.03, 0.04 and 0.05. Actually, it alters the number of particles per unit volume as $6x10^{5}\,mm^{-3}$, $9x10^{5}\,mm^{-3}$, $12x10^{5}\,mm^{-3}$ and $15x10^{5}\,mm^{-3}$. The concentration increases as the volume concentration of magnetic particles is increased, up to a point beyond which while the vortex formed close to the upper plate is yet appeared, the concentration decreases as the volume concentration of magnetic particles is more increased.

Figure 4e presents the variation of the dimensionless particle concentration according to the size of magnetic particles. As the size of magnetic particles increases $1\mu m$, $2\mu m$ and $2.8\mu m$, the concentration increases around the position (10 mm, 1 mm) where there is the maximum value of the magnetic field for the flow field domain.

Figure 4f presents the variation of the dimensionless particle concentration according to the flow velocity ($0.0005\,m\,s^{-1}$, $0.001\,m\,s^{-1}$, $0.002\,m\,s^{-1}$). The concentration increases as the flow velocity is increased until the point where the flow profile is changed and the vortex which is formed close to the upper plate is disappeared.

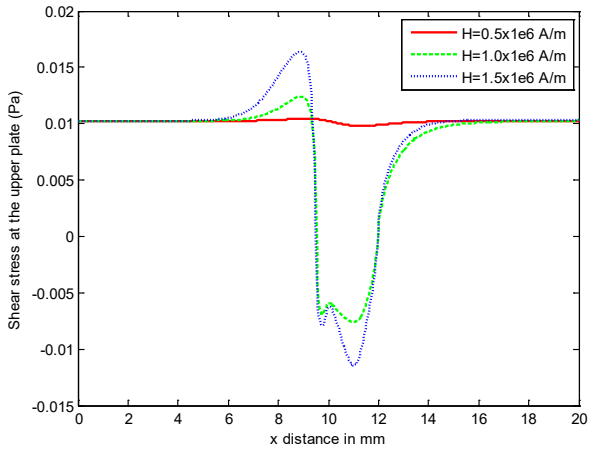


Figure 5: The variation of the dimensionless particle concentration at *y*= 1 mm along *x* direction (0 mm, 20 mm).

Figure 5 shows the variation of the shear stress magnitude along the upper plate in the micro-channel. It can be seen that high levels of shear stress are restricted to the narrow regions where the external magnetic field is applied. Peak values increase by increasing the intensity

of magnetic field. As it mentioned in [6, 7], the fact that peak values remain far below the lytic limit implies a safe transport mechanism for the separation of biological cells.

4 CONCLUSIONS

- Particle size and magnetic field intensity are the two factors that mostly affect particle concentration in the vicinity of the magnetic source.
- For the position of magnetic field, the magnetic susceptibility of magnetic particles, the volume concentration of magnetic particles (nanoparticles) and the flow velocity, there is a limit, beyond which the particle concentration decreases.
- High levels of shear stress are restricted to narrow regions near the upper plate where the external magnetic field is applied. Peak values of shear stress increase by increasing the intensity of magnetic field.

AKNOWLEDGMENTS

This research has been financed by the Research Committee - University of Patras - Research Funding Program: K. Karatheodoris 2013.

REFERENCES

[1] J.M.R. Carlton, C.A. Yowell, K.A. Sturrock, J.B. Dame, Biomagnetic separation of contaminating host leukocytes from plasmodium-infected erythrocytes. *Experimental Parasitology*, **97**, 111–114, 2001.

[2] Y. Haik, V. Pai, C.J. Chen, Development of magnetic device for cell separation. *Journal of Magnetism and Magnetic Materials*, **194**, 254–261, 1999.

[3] V. Loukopoulos, E. Tzirtzilakis, Biomagnetic channel flow in spatially varying magnetic field. *International Journal of Engineering Science*, **42**, 571–590, 2004.

[4] E.E. Tzirtzilakis, V.C. Loukopoulos, Biofluid flow in a channel under the action of a uniform localized magnetic field. Computational Mechanics, **36**, 360–374, 2005.

[5] X.L. Li, K.L. Yao, Z.L. Liu, CFD study on the magnetic fluid delivering in the vessel in high-gradient magnetic field. *Journal of Magnetism and Magnetic Materials*, **320**, 1753–1758, 2008.

[6] S.A. Khashan, E. Elnajjar, Y. Haik, CFD simulation of the magnetophoretic separation in a microchannel. *Journal of Magnetism and Magnetic Materials*, **323**, 2960–2967, 2011.

[7] S.A. Khashan, E. Elnajjar, Y. Haik, Numerical simulation of the continuous biomagnetic separation in a two-dimensional channel. *International Journal of Multiphase Flow*, **37**, 947-955, 2011.

[8] M. R. Habibi, M. Ghasemi, Numerical study of magnetic nanoparticles concentration in biofluid (blood) under influence of high gradient magnetic field. *Journal of Magnetism and Magnetic Materials*, **323**, 32–38, 2011.

[9] S.A. Khashan, E.P. Furlani, Coupled Particle-Fluid Transport and Magnetic Separation in Microfluidic Systems with Passive Magnetic Functionality. *Journal of Physics D Applied Physics*, **46**, 125002 (15pp), 2013.

[10] C. Kleinstreure, *Two-phase Flow Theory and Application*, Taylor & Francis, New York, 2003.

[11] G.C. Bourantas, E.D. Skouras, V.C. Loukopoulos, G.C. Nikiforidis, Meshfree Point Collocation Schemes for 2D Steady State Incompressible Navier-Stokes Equations in Velocity-Vorticity Formulation for High Values of Reynolds Number. *Computer Modeling in Engineering & Sciences*, **59**, 31-63, 2010.

[12] B.M. Johnston, P.R. Johnson, S. Corney, D. Kilpatrick, Non-Newtonian blood flow in human right coronary arteries: steady state simulations. *Journal of Biomechanics*, **37**, 709–720, 2004.

[13] G.C. Bourantas, V.C. Loukopoulos, E.D. Skouras, V.N. Burganos, G.C. Nikiforidis, "An IPOT Meshless Method Using DC PSE Approximation for Fluid Flow Equations in 2D and 3D Geometries", 13th *International Conference of Numerical Analysis and Applied Mathematics 2015 (ICNAAM 2015)*, Rodos, Greece, September 23-29, AIP Conference Proceedings, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NEW ALGORITHM FOR NUMERICAL SIMULATION OF SURFACE WAVES WITHIN THE FRAMEWORK OF THE FULL NONLINEAR DISPERSIVE MODEL

Zinaida I. Fedotova[1], Oleg I. Gusev[2], and Gayaz S. Khakimzyanov[2]

[1]Institute of Computational Technologies SB RAS
Lavrentiev Avenue 6, 630090 Novosibirsk, Russia
e-mail: zf@ict.nsc.ru

[2] Institute of Computational Technologies SB RAS
Lavrentiev Avenue 6, 630090 Novosibirsk, Russia
e-mail: gusev_oleg_igor@mail.ru, khak@ict.nsc.ru

Keywords: Surface Waves, Nonlinear Dispersive Equations on a Sphere, Finite Difference Method, Numerical Algorithm, Stability, Dispersion.

Abstract. *The paper describes a new numerical method for modeling long waves with dispersion on a rotating attracting sphere. The algorithm is presented in the form of a finite-difference scheme of predictor-corrector, so that at each step the inhomogeneous hyperbolic system similar to classical shallow water equations with a right-hand side, and the uniformly elliptic equation are solved alternately. Necessary conditions for the stability of the method are obtained and the dispersion properties are investigated.*

1 INTRODUCTION

Approximate hydrodynamic models describing long-wave processes in a liquid have been actively developed over several decades. In recent years they enjoy popularity both among applied researchers and theoreticians. The interest of engineers is caused by the growing demands of practice in regard to the protection and development of sea coasts, whereas theoreticians have found a lot of gaps in the justification of both the models and numerical methods for solving the respective differential equations. Among the recently published papers devoted to the development of nonlinear dispersive (NLD) hydrodynamic models, we note studies [1, 2, 3] which take into account Earth's sphericity and its rotation. Such complication of the NLD-models is necessary for adequate description of the surface waves behaviour in the ocean and along the coastal zone.

In [3, 4] with using the unified approach to a construction of long-wave approximations, the hierarchy of hydrodynamic models, which possess the physically meaningful properties inherited, has been built. The hierarchical chains of shallow water models of first- and second-order long-wave approximation enclosed in each other are constructed on a rotating sphere and on a plane. With regard to numerical methods for the NLD-models, in contrast to the classical case of nonlinear shallow water (NLSW) equations, we have to state their lagging behind the development of the models themselves.

On the one hand, today there is urgent necessity in the numerical simulation based on practice needs, which should take into account sharp changes in the bottom surface, the non-linearity, vorticity, and other effects, on the other hand, there are many gaps in the study of fundamental properties of numerical methods, even at the simplest NLD-equations.

The peculiarity of our investigation is a methodical analysis which was initiated in [5]. The studies showed that the hierarchical approach helps to reveal not only the properties inheritance of the chain differential equations, but the same for the corresponding numerical algorithms, as well as describes the additional properties of difference schemes, acquired during the transition from the classical shallow water equations to NLD-equations. In particular, for Boussinesq type equations the improved stability conditions of some popular difference schemes of first- and second-order approximations are obtained. These conditions are characterized by presence of a new parameter defining the grid fineness compared to the characteristic depth, and in the limit of a grid refinement the conditions can be written in the form of the limitation on the time step only. It is also shown that in the domain of stability of some schemes there are the values of Courant numbers for which the influence of "scheme dispersion" is minimal.

The purpose of this paper is to present new numerical method for solving the NLD-equations on a rotating sphere [1]. The main idea lies in special splitting of the problem allowing the successive solution of two less difficult tasks at each time step. The first task is to calculate a dispersive component of the pressure using the uniformly elliptic equation. The second task is to solve the system of hyperbolic type consisting of the equations of continuity and motion in the dispersion-free shallow water model. The decomposition of the original system on two subproblems allows us to solve numerically the NLD-equations using the well known finite difference predictor-corrector type scheme for classical shallow water model modified so that the pressure is calculated additionally at each time step [6]. The corrected stability conditions and new knowledge on the dispersion properties of the modified predictor-corrector scheme are obtained.

2 NONLINEAR DISPERSIVE HYDRODYNAMIC MODELS

The full nonlinear dispersion (FNLD-) model on a rotating sphere based on the depth-averaged velocity was obtained in 2010 [1]. The paper [7] shows that the FNLD-model of [1] can be derived without any assumptions about the potentiality of the original 3D-flow, and that its defining equations, unlike [2], can be written in the quasi conservative form of mass and momentum balances. Additionally, this model has the balance equation of total energy agreed with a similar equation of 3D-model, that confirms not only the physical consistency of the FNLD-model, but also allows an additional control in the calculations. The equations of our FNLD-model on a rotating attracting sphere have the following form [1]:

$$(HR\sin\theta)_t + (Hu)_\lambda + (Hv\sin\theta)_\theta = 0, \tag{1}$$

$$\begin{aligned}(HuR\sin\theta)_t + (Hu^2 + g\frac{H^2}{2})_\lambda + (Huv\sin\theta)_\theta = gHh_\lambda - Huv\cos\theta - \\ -fvHR\sin\theta + \varphi_\lambda - \psi h_\lambda,\end{aligned} \tag{2}$$

$$\begin{aligned}(HvR\sin\theta)_t + (Huv)_\lambda + \Big[(Hv^2 + g\frac{H^2}{2})\sin\theta\Big]_\theta = gHh_\theta\sin\theta + g\frac{H^2}{2}\cos\theta + \\ +Hu^2\cos\theta + fuHR\sin\theta + (\varphi_\theta - \psi h_\theta)\sin\theta + \Omega^2 HR^2\sin^2\theta\cos\theta.\end{aligned} \tag{3}$$

Here R is a radius of a sphere rotating with a constant velocity Ω around the axis Oz of the fixed coordinate system $Oxyz$, with coordinate plane Oxy coinciding with the equatorial plane of the sphere.

To describe the flow of water a rotating together with sphere coordinate system $O\lambda\theta r$ is used, the beginning of which is located in the center of the sphere. Here λ is the longitude counted in the direction of a rotation from a certain meridian ($0 \le \lambda < 2\pi$), $\theta = \pi/2 - \varphi$ is the addition to the latitude ($-\pi/2 < \varphi < \pi/2$), r is the radial coordinate measured from the center of the sphere.

The Newtonian attractive force $\mathbf{g}$, acting on a liquid particle of unit mass, is directed to the center of the Earth. The thickness of the water layer $H = \eta + h > 0$ is assumed to be small compared with the radius of the sphere, so the values of $g = |\mathbf{g}|$, and of water density ρ are assumed to be constant throughout the liquid layer bounded below by the impermeable moving bottom and above by a free surface:

$$r = R - h(\lambda,\theta,t), \quad r = R + \eta(\lambda,\theta,t). \tag{4}$$

The symbols u and v denote the physical components of the velocity vector ($u = Rc^1\sin\theta$, $v = Rc^2$, $c^1 = \dot\lambda$, $c^2 = \dot\theta$), $f = 2\Omega\cos\theta$ is the Coriolis parameter, expressed in terms of latitude's addition θ, here it is assumed that

$$\theta_0 \le \theta \le \pi - \theta_0, \tag{5}$$

where $\theta_0 = \text{const} > 0$ is a small angle (poles are excluded from the consideration).

The values of φ and ψ, included in the right sides of the equations of motion (2), (3) of the FNLD-model, are the dispersive components, respectively, of depth-integrated pressure p and the pressure p_0 at the bottom:

$$p = \frac{gH^2}{2} - \varphi, \quad p_0 = gH - \psi. \tag{6}$$

Dispersive additives are expressed by the following formulas [8]:

$$\varphi = \frac{H^3}{3}Q_1 + \frac{H^2}{2}Q_2, \quad \psi = \frac{H^2}{2}Q_1 + HQ_2, \tag{7}$$

where $Q_1 = D(\nabla \cdot \mathbf{c}) - (\nabla \cdot \mathbf{c})^2$, $Q_2 = D^2 h$, $\mathbf{c} = (c^1, c^2)$,

$$D = \frac{\partial}{\partial t} + \mathbf{c} \cdot \nabla, \quad \nabla = \Big(\frac{\partial}{\partial \lambda}, \frac{\partial}{\partial \theta}\Big), \quad \mathbf{c} \cdot \nabla = c^1 \frac{\partial}{\partial \lambda} + c^2 \frac{\partial}{\partial \theta},$$

$$\nabla \cdot \mathbf{c} = c^1_\lambda + \frac{1}{J}\left(Jc^2\right)_\theta, \quad J = -R^2 \sin\theta \, .$$

More detailed formulas for Q_1 and Q_2 have the form

$$Q_1 = (\nabla \cdot \mathbf{c})_t + \frac{1}{R\sin\theta}\Big(u\,(\nabla \cdot \mathbf{c})_\lambda + v\,(\nabla \cdot \mathbf{c})_\theta \sin\theta\Big) - (\nabla \cdot \mathbf{c})^2,$$

$$Q_2 = (Dh)_t + \frac{1}{R\sin\theta}\Big(u\,(Dh)_\lambda + v\,(Dh)_\theta \sin\theta\Big),$$

where

$$\nabla \cdot \mathbf{c} = \frac{1}{R\sin\theta}\Big(u_\lambda + (v\sin\theta)_\theta\Big), \quad Dh = h_t + \frac{1}{R\sin\theta}\Big(uh_\lambda + vh_\theta \sin\theta\Big).$$

The FNLD-model (1)–(3) with φ and ψ as in (7) is called "full" [8] for the reason that it is derived without the assumption on the smallness of wave amplitudes and in the equations are stored all the nonlinear terms associated with dispersion. So, it can be used to calculate the surface waves propagation over the uneven bottom both in deep water and in the coastal zone. Additionally, the accounting for the waves dispersion can give more accurate results than dispersionless shallow water model. The FNLD-model described above can also simulate the waves generated by the long-time shifts of bottom fragments, that extends the range of problems that can be solved within the framework of the famous [9, 10, 11] weakly nonlinear dispersive (WNLD-) models of shallow water on the sphere.

Moreover, the structure of the FNLD-equations and the presence of the important properties such as conservatism and existence of the total energy balance equation agreed with 3D-model, is stored during the special simplification of the dispersion component and under transition to a flat geometry. Thus, the equations of the weakly nonlinear dispersion model obtained in [7] and the classical (nondispersive) shallow water model on a sphere have the same form as the equations of the FNLD-model on a sphere. The differences appear in the expression for the kinetic energy and in the calculations of the pressure terms (6). We provide the single form of the equations of the entire hierarchical chain of long-wave models in spherical and planar geometries [3], that made it possible to construct the uniform numerical algorithm for them.

3 FINITE-DIFFERENCE METHODS FOR SHALLOW WATER EQUATIONS WITH DISPERSION

Since NLD-equations and finite-difference approximations can't be studied analytically in the total statement, it is reasonably to consider the simplified statement (cases of plane geometry, a flat bottom $h \equiv h_0 = \text{const}$, linear analogues of equations and difference schemes). This makes it possible to investigate the properties in particular cases (for example, the dispersion properties of linear difference schemes) and leads to a set of necessary restrictions on the use

of the total algorithm (for example, a necessary condition for stability, preservation by the algorithm of some properties of the original differential problem, etc.). Here we consider two finite-difference algorithms for solving a system of linear equations of one-dimensional flows

$$\eta_t + h_0 u_x = 0, \quad u_t + g\eta_x = \nu u_{xxt}, \qquad \nu = \frac{h_0^2}{3}, \tag{8}$$

into which the NLD-models from the papers [12, 13, 14] and some other NLD-models can be transformed after linearization. Finite-difference approximations of these linear equations are used to obtain necessary stability conditions of nonlinear difference schemes, and also to research their dissipative and dispersive properties.

NLD-equations together with its linear variant (8) contain third order mixed derivatives with respect to time and space, that requires specific approaches to their numerical solution. The main technique to ensure inheritance of existing computational methods and development of new effective numerical algorithms for NLD-models, is in follows. By introducing new variables NLD-equations can be splitted in such way that one component is either a system of ordinary differential equations (ODE's) or inhomogeneous system of hyperbolic equations, and the other component does not include derivatives with respect to time and in the simplest case looks like as an elliptical equation. This opens up opportunities for using methods of ODE solutions and finite difference/finite-volume schemes, and the finite element method, as well as various combinations of these and other suitable methods.

The numerical algorithms based on splitting of NLD-equations on the system of ODE's and the elliptic equation, was first proposed in [15] for the unidirectional scalar equation, and then was extended to systems of WNLD- and FNLD-equations [16, 17, 18, 19, 20]. The feature of solving systems of NLD-equations is that the group of terms describing the dispersion contains the time derivative of the velocity u (in one-dimensional case). The easiest way to solve this task lies in the possibility to represent all terms with the time derivative as an expression Q_t with some Q, that allows to bring the implementation of the corresponding numerical algorithm as a step-by-step solving ODE's and the elliptic equation. With regard to the system of equations (8) this method leads to the following system:

$$\eta_t + h_0 u_x = 0, \quad Q_t = g\eta_x, \quad \nu u_{xx} - u = Q.$$

One of the most common ways of constructing a numerical algorithm is as follows. η^{n+1} is first calculated from the continuity equation by the two-layer explicit scheme (or implicit with respect to η). Then Q^{n+1} is calculated from ODE using η^{n+1} obtained. After that the elliptical (in one dimension – ordinary differential) equation for u^{n+1} is solved (in one dimension – by a scalar sweep).

The disadvantage of this approach is the non-divergent form of the system in the nonlinear case. It should also be noted that in two-dimensional case the system of two elliptic equations for the components of the velocity vector is got, which complicates the computational algorithm. But the most significant drawback is missing the opportunity of using methods for solving hyperbolic equations, well developed for NLSW-shallow water model of the first hydrodynamic approximation.

In this paper we propose the different way of splitting. The initial system of NLD-equations is replaced on the extended system consisting of the system of hyperbolic equations similar to NLSW-equations and differing from them by the right side, and of the scalar equation of elliptic type for the dispersive pressure component. This approach to the construction of numerical algorithms can be used for all FNLD- and WNLD-models described above.

Here we look at the essence of this approach on the example of the linear system (8). In this case, the extended system takes the form

$$\eta_t + h_0 u_x = 0, \quad u_t + g\eta_x = \frac{\varphi_x}{h_0}, \quad \varphi_{xx} - \frac{\varphi}{\nu} = c_0^2 \eta_{xx}, \qquad c_0 = \sqrt{g h_0}. \tag{9}$$

It contains the new equation for the dependent variable $\varphi = \nu u_{xt}$, which is a linear analog of the dispersion component of the pressure φ from (7). Let Δx and τ be steps of the uniform grid in the plane of variables x and t, $(x_j,\, t^n)$ are grid nodes, $x_j = j\Delta x$, $t^n = n\tau$, η_j^n, u_j^n, φ_j^n are the values of grid functions in these nodes. For difference derivatives the following abbreviations are used:

$$\eta_{t,j}^n = \frac{\eta_j^{n+1} - \eta_j^n}{\tau}, \quad \eta_{x,j}^n = \frac{\eta_{j+1}^n - \eta_j^n}{\Delta x}, \quad \eta_{\bar{x}x,j}^n = \frac{\eta_{j+1}^n - 2\eta_j^n + \eta_{j-1}^n}{\Delta x^2}.$$

We use the predictor-corrector method as the difference scheme and solve the equations of hyperbolic and elliptic parts one after another on each computational time step. In the predictor step the auxiliary quantities $\eta_{j+1/2}^*$, $u_{j+1/2}^*$, $\varphi_{j+1/2}^*$ defined in cell centers are calculated. To do that we first solve the explicit difference equations corresponding to a hyperbolic part

$$\frac{\eta_{j+1/2}^* - \frac{1}{2}(\eta_{j+1}^n + \eta_j^n)}{\tau_{j+1/2}^*} + h_0 u_{x,j}^n = 0, \quad \frac{u_{j+1/2}^* - \frac{1}{2}(u_{j+1}^n + u_j^n)}{\tau_{j+1/2}^*} + g\eta_{x,j}^n = \frac{1}{h_0}\varphi_{x,j}^n, \tag{10}$$

and then calculate $\varphi_{j+1/2}^*$ from

$$\frac{\varphi_{j+3/2}^* - 2\varphi_{j+1/2}^* + \varphi_{j-1/2}^*}{\Delta x^2} - \frac{\varphi_{j+1/2}^*}{\nu} = c_0^2 \frac{\eta_{j+3/2}^* - 2\eta_{j+1/2}^* + \eta_{j-1/2}^*}{\Delta x^2}, \tag{11}$$

where $\tau_{j+1/2}^* = \dfrac{\tau}{2}(1 + \theta_{j+1/2}^n)$, $\theta_{j+1/2}^n$ is the parameter responsible for the TVD-properties.

Calculated predictor values are then used in the second stage – corrector – to calculate values η_j^{n+1}, u_j^{n+1} and φ_j^{n+1}:

$$\eta_{t,j}^n + h_0 \frac{u_{j+1/2}^* - u_{j-1/2}^*}{\Delta x} = 0, \quad u_{t,j}^n + g\frac{\eta_{j+1/2}^* - \eta_{j-1/2}^*}{\Delta x} = \frac{1}{h_0}\frac{\varphi_{j+1/2}^* - \varphi_{j-1/2}^*}{\Delta x}, \tag{12}$$

$$\varphi_{\bar{x}x,j}^{n+1} - \frac{\varphi_j^{n+1}}{\nu} = c_0^2 \eta_{\bar{x}x,j}^{n+1}. \tag{13}$$

It is easy to show that for $\theta \equiv \text{const} \neq 0$ the scheme has the first-order approximation, for $\theta \equiv 0$ it has the second. It is an interesting case of quasi constant values of $\theta = O(\tau, \Delta x)$, when the scheme parameter depends on the mesh size, and decreases in proportion to the decrease of a step, but does not depend on the coordinates of nodes, i.e. behaves as a constant parameter. In the case of quasi-constant parameter θ the scheme is second-order approximation. The necessary condition for stability of the written scheme is as follows:

$$c_0 æ \leq \frac{\sqrt{1 + 4\delta^2\theta/3}}{1 + \theta}, \tag{14}$$

where $0 \leq \theta \leq 1$, $æ = \tau/\Delta x$ and the parameter

$$\delta = \frac{h_0}{\Delta x}$$

characterizes the size of the spatial step Δx in relation to the characteristic depth h_0. The presence of this parameter is the peculiarity of the finite difference schemes for the NLD-equations since it together with the Courant number determines the stability condition and occurs in the formulas describing the dissipative and dispersion properties of numerical schemes. This is one of the differences from the dispersionless case where under similar conditions only one parameter (the Courant number) appears. Increasing δ leads to the fact that the stability condition (14) becomes more and more weak and under $\delta \gg 1$ can be transformed into a time step limitation:

$$\tau \leq \frac{2\sqrt{\theta}}{\sqrt{3}(1+\theta)}\tau_0 < \frac{1}{\sqrt{3}}\tau_0 \approx 0.58\tau_0, \tag{15}$$

where another new parameter $\tau_0 = h_0/c_0$ is introduced, denoting the characteristic time which it takes to wave propagating with the velocity c_0 to pass the distance equal to the characteristic depth h_0.

This fact can be interpreted in such a way that for sufficiently fine grid the stability condition can be replaced by (15), i.e. we get a restriction on the time step τ independent from Δx. Such a situation does not occur in the theory of difference schemes for classical shallow water. Really, for dispersionless linear shallow water equations, i.e. (8) with $\nu = 0$, the predictor-corrector scheme becomes easier due to the fact that there is no need to calculate the dispersion additive φ. The stability condition of such scheme has the form

$$c_0 æ \leq \frac{1}{\sqrt{1+\theta}}. \tag{16}$$

Here we have the stability condition in the form of restrictions on the Courant number, which is the standard for explicit schemes approximating hyperbolic equations. Grinding steps of the spatial grid requires proportional refinement of time steps.

We emphasize the importance of the observed feature consisting in the fact that the stability condition (14) of the difference scheme for dispersion equations is less stringent than (16) for dispersionless ones. This fact also holds for other known schemes [5]. However, it seems that in earlier studies this fact has not been discovered, and the previous developers of numerical algorithms for the NLD-equations chose, as a rule, time steps on the basis of more stringent restrictions: of NLSW-equations, expressed in Courant numbers. The discovered fact was confirmed in the course of experiments conducted for nonlinear case: for NLD-equations a stability remained at bigger time steps than it required for NLSW-equations.

To investigate the dispersion properties of the predictor-corrector scheme let's calculate its phase error

$$\Delta\phi = \mp\frac{c_0 æ}{6}\Big(c_0^2 æ^2(3\theta+1) - 1\Big)\xi^3 + O(\xi^4),$$

where $\xi \in [0,\ \pi],\ \xi = k\Delta x,\ k$ is the wave number of a harmonic. We see that the main part of the phase error on long waves has generally the same order of ξ as the "physical" dispersion Φ of the model (8)

$$\Phi = \pm\left[c_0 æ\xi - \frac{c_0 æ}{6}\delta^2\xi^3 + O(\xi^4)\right],$$

that is a lack. If for a given value $\theta \geq 0$ a Courant number is selected by the formula

$$c_0 æ = \frac{1}{\sqrt{1+3\theta}}, \tag{17}$$

the phase error will be at least one order of ξ less than the "physical" dispersion, and herewith the stability condition (14) is performed. Therefore, under the condition (17) the numerical dispersion for the long wave will not suppress the "physical" one.

4 THE NUMERICAL ALGORITHM FOR NONLINEAR DISPERSIVE EQUATIONS ON A ROTATING SPHERE

The system of equations (1)–(3) of the FNLD-model on the sphere, as the linear system (8) discussed above, is not a system of Cauchy–Kovalevskaya type due to the fact that the equations of motion contain the mixed third-order derivatives of the velocity vector components with respect to time and space. The direct approximation of the derivatives leads to a complex problem which is difficult to solve. As shown above with respect to the linear case, it turned out fruitful to split the system (8) on the scalar equation of elliptic type and the system of hyperbolic equations.

The same approach can be implemented for the system of FNLD-equations (1)–(3) on a rotating sphere. As a result of splitting we obtain the extended system of equations consisting of an elliptic equation for the dispersive component φ of the depth-integrated pressure p and the hyperbolic system of equations (1)–(3) with time derivatives only of the first order that differs from the classical shallow water model on a sphere only by additional terms associated with the dispersive additives φ and ψ in the right part of the motion equations (we consider here φ and ψ as variables, but not as expressions).

The equation for φ is as follows [21]:

$$\Big[\frac{1}{\sin\theta}\Big(\frac{\varphi_\lambda}{H}-\frac{\nabla\varphi\cdot\nabla h}{Hr}h_\lambda\Big)\Big]_\lambda+\Big[\Big(\frac{\varphi_\theta}{H}-\frac{\nabla\varphi\cdot\nabla h}{Hr}h_\theta\Big)\sin\theta\Big]_\theta-k_0\varphi=F, \tag{18}$$

where $k_0=k_{00}+(k_{01})_\lambda+(k_{02})_\theta,\ r=4+\nabla h\cdot\nabla h,$

$$k_{00}=\frac{12(r-3)}{H^3r}R^2\sin\theta,\quad k_{01}=\frac{6h_\lambda}{H^2r\sin\theta},\quad k_{02}=\frac{6h_\theta}{H^2r}\sin\theta.$$

A distinctive feature of the equation (18) is the fact that neither the left nor the right part does not contain time derivatives of the dependent variables H, $\mathbf{u}=(u,v)$. In addition, under conditions $H>0$ and (5) the equation (18) is uniformly elliptic. Therefore, it is possible to construct the difference scheme with the positive definite operator and to use a suitable iterative methods for solving scalar elliptic equations.

For the dispersion component ψ an individual equation is not required, since the function ψ is associated with values of H, $\mathbf{u}$ and φ by expression [21]:

$$\psi=\frac{1}{r}\left(\frac{6\varphi}{H}+HQ+\nabla\varphi\cdot\nabla h\right), \tag{19}$$

where

$$\nabla\varphi\cdot\nabla h=\frac{1}{R^2}\left(\frac{\varphi_\lambda\bar h_\lambda}{\sin^2\theta}+\varphi_\theta\left(\bar h_\theta-\eta_{00,\theta}\right)\right)=\nabla\varphi\cdot\nabla\bar h-\varphi_\theta\frac{\Omega^2}{g}\sin\theta\cos\theta,$$

$$Q=(-g\nabla\bar\eta+\mathbf{a})\cdot\nabla h+\frac{1}{R^2\sin\theta}\left(\frac{u^2}{\sin\theta}\ h_{\lambda\lambda}+2uv\ h_{\lambda\theta}+v^2h_{\theta\theta}\ \sin\theta\right)+B,$$

$$B=h_{tt}+2\Big(\frac{u}{R\sin\theta}h_{\lambda t}+\frac{v}{R}h_{\theta t}\Big),$$

$$\mathbf{a}=(a_1,a_2),\quad a_1=-(2uv\ \mathrm{ctg}\,\theta+fvR)\sin\theta,\quad a_2=u^2\mathrm{ctg}\,\theta+fuR,$$

$f=2\Omega\cos\theta$ is the Coriolis parameter. Functions $\bar\eta$ and $\bar h$ are connected with η and h by relations

$$-h=\eta_{00}-\bar h,\quad \eta=\eta_{00}+\bar\eta$$

and describe the bottom and the free surface not in the form of deviations (4) from the surface of a sphere of radius R, but the deviations $\bar{h}$ and $\bar{\eta}$ from the undisturbed free boundary, as is usually done for solving shallow water equations in plane geometry. In this case the unperturbed free boundary at rest differs from spherical [7] and is described by the equation $r = R + \eta_{00}(\theta)$, where

$$\eta_{00}(\theta) = \frac{1}{2g}\Omega^2 R^2 \sin^2\theta + \text{const}.$$

The numerical algorithm for the extended system of equations (1)–(3), (18) is a direct generalization of the finite-difference scheme (10)–(13) which was discussed above in the context of a linear case. This method is the predictor-corrector scheme, whose steps consist of two stages. On the predictor step the values H^*, $\mathbf{u}^*$ are first determined in the centers of cells as the solution of the explicit difference equations for the system of hyperbolic equations written in non-divergent form. Then we solve the difference equation for φ^* using the new values H^*, $\mathbf{u}^*$ in coefficients and a right-hand side, and define ψ^* by formula (19).

The values H^*, $\mathbf{u}^*$, φ^* and ψ^* found on the predictor step are used then on the corrector step for computation of final values H^{n+1} and $\mathbf{u}^{n+1}$ with the numerical solution of the system (1)–(3) in the divergent form of the left part. In the last turn values of functions φ^{n+1} and ψ^{n+1} are calculated. The similar algorithm can be applied also to the solution of the FNLD-equations in plane geometry [6].

Using a benchmark about the propagation of waves on transoceanic distances it is shown in [21] that in terrestrial conditions the centrifugal force can be ignored and the accounting effects of sphericity can be critical important. Influence of the Coriolis force increases with the horizontal sizes and with propagation distances. The accounting of the sphere rotation lowers the maximum positive amplitudes of waves. Dispersion is also manifested at long distances of waves traveling and can lead to essential reduction of the maximum amplitude of a head wave which disintegrates into a train of solitary waves of smaller length. But, unlike the Coriolis force, dispersion appears more strongly for the smaller size of initial disturbance, at that the accounting of dispersion leads to the deceleration of the waves advance.

5 CONCLUSIONS

This paper sets forth the numerical algorithm for the simulation of propagation of surface waves in the frame of the full nonlinear dispersion model which takes into account the sphericity of the Earth and its rotation. The algorithm is based on the solution of the extended system of equations consisting of the elliptic equation for the dispersive component of the pressure and the hyperbolic first order equations, which differ from the equations of classical shallow water model only by presence of additional terms related to the dispersive additives. The algorithm is implemented as the explicit two-step predictor-corrector scheme with solving the resulting subproblems at each step.

Since the structure of the equations of the full nonlinear dispersive model holds true for the weakly dispersion model on a sphere, as well as the model of weakly dispersive flows over weakly deformable bottom [4], the algorithm for the numerical solution of the full nonlinear dispersive equations can be transferred unchanged also to this approximate models.

Funding: The work was supported by the Russian Scientific Foundation (project No. 14-17-00219).

REFERENCES

[1] Z.I. Fedotova, G.S. Khakimzyanov, Nonlinear-dispersive shallow water equations on a rotating sphere. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **25** (1), 15–26, 2010.

[2] J.T. Kirby, F. Shi, B. Tehranirad, J.C. Harris, S.T. Grilli, Dispersive tsunami waves in the ocean: Model equations and sensitivity to dispersion and Coriolis effects. *Ocean Modelling*, **62**, 39–55, 2013.

[3] Yu.I. Shokin, Z.I. Fedotova G.S. Khakimzyanov, Hierarchy of nonlinear models of the hydrodynamics of long surface waves. *Doklady Physics*, **60** (5), 224-228, 2015.

[4] Z.I. Fedotova, G.S. Khakimzyanov, D. Dutykh, Energy equation for certain approximate models of long-wave hydrodynamics. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **29** (3), 167–178, 2014.

[5] Z.I. Fedotova, G.S. Khakimzyanov, O.I. Gusev, History of the development and analysis of numerical methods for solving nonlinear dispersive equations of hydrodynamics. I. One-dimensional models problems. *Computational Technologies*, **20** (5), 120–156, 2015 (In Russian).

[6] G.S. Khakimzyanov, O.I. Gusev, S.A. Beisel, L.B. Chubarov, N.Yu. Shokina, Simulation of tsunami waves generated by submarine landslides in the Black Sea. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **30** (4), 227–237, 2015.

[7] Z.I. Fedotova, G.S. Khakimzyanov, Nonlinear-dispersive shallow water equations on a rotating sphere and conservation laws. *Journal of Applied Mechanics and Technical Physics*, **55** (3), 404–416, 2014.

[8] Z.I. Fedotova, G.S. Khakimzyanov, Full nonlinear dispersion model of shallow water equations on a rotating sphere. *Journal of Applied Mechanics and Technical Physics*, **52** (6), 865–876, 2011.

[9] F. Lovholt, G. Pedersen, G. Gisler, Oceanic propagation of a potential tsunami from the La Palma Island. *Journal of Geophysical Research*, **113**, C09026, 2008.

[10] F. Lovholt, G. Pedersen, S. Glimsdal, Coupling of dispersive tsunami propagation and shallow water coastal response. *Open Oceanography Journal*, **4**, 71–82, 2010.

[11] S. Glimsdal, G.K. Pedersen, C.B. Harbitz, F. Lovholt, Dispersion of tsunamis: does it really matter? *Natural Hazards and Earth System Science*, **13**, 1507–1526, 2013.

[12] D.H. Peregrine, Long waves on a beach. *Journal of Fluid Mechanics*, **27**, 815–827, 1967.

[13] R.C. Ertekin, W.C. Webster, J.V. Wehausen, Waves caused by a moving disturbance in a shallow channel of finite width. *Journal of Fluid Mechanics*, **169**, 275–292, 1986.

[14] Z.I. Fedotova, G.S. Khakimzyanov, Shallow water equations on a movable bottom. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **24** (1), 31–42, 2009.

[15] J.C. Eilbeck, G.R. McGuire, Numerical study of the regularized long-wave equations 1: Numerical methods. *Journal of Computational Physics*, **19**, 43–57, 1975.

[16] O. Nwogu, Alternative form of Boussinesq equations for nearshore wave propagation. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **119** (6), 618–638, 1993.

[17] G. Wei, J.T. Kirby, Time-dependent numerical code for extended Boussinesq equations. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **121** (5), 251–261, 1995.

[18] G. Wei, J.T. Kirby, S.T. Grilli, R. Subramanya, A fully nonlinear Boussinesq model for surface waves. Part 1. Highly nonlinear unsteady waves. *Journal of Fluid Mechanics*, **294**, 71–92, 1995.

[19] L.A. Kompaniets, Analysis of difference algoritms for nonlinear dispersive shallow water models. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **11** (3), 205–222, 1996.

[20] Z.I. Fedotova, V.Yu. Pashkova, Methods of construction and the analysis of difference schemes for nonlinear dispersive models of wave hydrodynamics. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **12** (2), 127–149, 1997.

[21] O.I. Gusev, G.S. Khakimzyanov, Numerical simulation of long surface waves on a rotating sphere within the framework of the full nonlinear dispersive model. *Computational Technologies*, **20** (3), 3–32, 2015 (In Russian).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL MODELLING OF SURFACE WAVES IN THE FRAMEWORK OF SHALLOW WATER MODEL

Nina Shokina[1], **Gayaz Khakimzyanov**[2]

[1]Computing Center, University of Freiburg
Hermann-Herder-Straße 10, 79104 Freiburg im Breisgau, Germany
e-mail: nina.shokina@gmail.com

[2] Institute of Computational Technologies SB RAS
Lavrentiev Ave. 6, 630090 Novosibirsk, Russia
e-mail: khak@ict.nsc.ru

Keywords: Nonlinear Shallow Water Equations, Finite-Difference Scheme, Adaptive Grid, Surface Waves, Run-Up

Abstract. *An improved adaptive grid method is considered for the numerical solution of the problems on propagation and run-up of surface waves, described by the one-dimensional shallow water model. The modified algorithm for the realization of the explicit predictor-corrector scheme is presented, which is based on the new way of computation of the right-hand side of the shallow water equations. A new method for choosing the scheme parameters on the basis of the analysis of the differential approximation is suggested that guarantees the satisfaction of the TVD-property for the improved predictor-corrector scheme. The presented method for construction of different conservative schemes on moving grids is based on an appropriate choice of the scheme parameters for the predictor–corrector scheme, which represents the canonical form of the two-layer explicit schemes for the shallow water equations. The improved difference boundary conditions are obtained at the moving waterfront point using the known analytical solutions of the shallow water equations in the vicinity of a water-land boundary. These boundary conditions approximate the analytical solutions with a higher accuracy than the conditions used in the earlier works. The numerical results for the improved adaptive grid method are presented.*

1 INTRODUCTION

Nowadays, the methods based on the hierarchy of embedded mathematical models [1] and computational algorithms are widely used for the numerical solution of the problems on propagation of long surface waves in large water areas and their interaction with coasts. The nonlinear shallow water models hold a central position in such a hierarchy. Therefore, the development and investigation of numerical methods for the solution of the problems in the framework of shallow water models is a vital task of the modern computational fluid dynamics [2].

The present work is the continuation of the investigations started in [3] and devoted to some problems of the numerical modelling in the framework of the nonlinear non-dispersive shallow water model (NLSW-model). The NLSW-system is inhomogeneous for an irregular bottom, therefore, many finite-difference schemes, which show good results for a flat bottom, loose the important property of preservation of simple analytical solutions of the NLSW-model [4]. This problem can be solved by a special approximation of the right-hand side of the impulse equation.

It is known, that the solution of the NLSW-equations can become discontinuous even for continuous initial data, often leading to oscillations of the numerical solution [5, 6, 7]. A new approach for choosing the scheme parameters is suggested, which is based on the investigation of the differential approximation of the scheme, guaranteeing the satisfaction of the TVD-property.

In [3] an important conservation property for schemes on moving grids for the nonlinear scalar equation has been achieved by a proper choice of a scheme parameter. In the present work, the same approach is used for automatic construction of conservative schemes for the shallow water equations on moving grids.

An adequate description of the waterfront behavior in the run-up area and a reliable method for the detection of inundation zones are needed in the problems of the propagation of long surface waves in large water areas and their interaction with coasts. The methods for numerical modelling of flooding-drying processes, necessary for the detection of inundation zone after a wave run-up on a coast, can be divided into two main groups [8]. First group consists of the shock-capturing methods with direct application of existing algorithms on fixed spatial grids, where a waterfront has to be evaluated at each time moment. Second group contains the methods on moving grids, which adapt so that a waterfront always coincides with a computational domain boundary. In flooding-drying problems the finite volume methods [9, 10], finite element methods [11, 12, 13], finite difference methods [14, 15], particle method [16] and others are used for the fluid flow modelling.

The present work uses moving adaptive grids for the detection of inundation zones, and a waterfront position is obtained using the exact analytical solutions of the NLSW-model in the waterfront vicinity in a time interval equal to a time step of a numerical scheme. The updated formulas are presented for evaluating a waterfront position and velocity. The results of the numerical modelling of a wave run-up on vertical wall and plane slope are presented.

2 NEW PROCEDURE FOR RIGHT-HAND SIDE COMPUTATION

In [3] a finite-difference scheme on moving grids was suggested and investigated for the one-dimensional shallow water equations:

$$\vec{u}_t + \vec{f}_x = \vec{G}, \tag{1}$$

where

$$\vec{u} = \begin{pmatrix} H \\ Hu \end{pmatrix}; \quad \vec{f}(\vec{u}) = \begin{pmatrix} Hu \\ Hu^2 + gH^2/2 \end{pmatrix}; \quad \vec{G}(x, \vec{u}) = \begin{pmatrix} 0 \\ gHh_x \end{pmatrix};$$

t is the time; $u(x,t)$ is the velocity; $\eta(x,t)$ is the deviation of free surface from a rest level; $h(x)$ is the bottom depth measured from a free surface at rest; $H = \eta + h$ is the full depth; g is the gravity acceleration. It is assumed that some initial boundary value problem is formulated for the system of equations (1) on the interval $x \in [0,\ l]$ and time interval $t \in [0,\ T]$.

The node coordinates of a moving grid, covering the interval $[0,\ l]$ at time moment $t = t^n$, are denoted by x_j^n $(j = 0, \ldots, N)$. These nodes are the images of the nodes q_j of a uniform grid with a step $h = 1/N$ on the interval $q \in [0,\ 1]$ under some non-degenerated mapping $x = x(q,t)$. In the predictor-corrector scheme [3] the flux vector in computed in cell centers on the predictor step:

$$\hat{\vec{f}}_{j+1/2} = \frac{\vec{f}_j^n + \vec{f}_{j+1}^n}{2} - \frac{\tau}{2}\Big(\frac{1}{J}\mathcal{R}\mathcal{D}\bar{\Lambda}(\bar{\Lambda}\vec{P} - \mathcal{L}\vec{G})\Big)_{j+1/2}^n. \tag{2}$$

Then the flux vector is used for computing the full depth H_j^{n+1} and impulse $(Hu)_j^{n+1}$ according the explicit formulas on the corrector step:

$$\frac{(J\vec{u})_j^{n+1} - (J\vec{u})_j^n}{\tau} + \frac{\Big(\hat{\vec{f}} - (x_t\vec{u})^n\Big)_{j+1/2} - \Big(\hat{\vec{f}} - (x_t\vec{u})^n\Big)_{j-1/2}}{h} = \vec{G}_j^*, \tag{3}$$

where τ is the time step; $\vec{f}_j^n = \vec{f}(\vec{u}_j^n)$; $\vec{P}_{j+1/2}^n = (\mathcal{L}\vec{u}_q)_{j+1/2}^n$;

$$\vec{u}_{j+1/2}^n = \frac{\vec{u}_j^n + \vec{u}_{j+1}^n}{2}, \quad \vec{u}_{q,j+1/2}^n = \frac{\vec{u}_{j+1}^n - \vec{u}_j^n}{h}; \quad x_{t,j}^n = \frac{x_j^{n+1} - x_j^n}{\tau}; \quad x_{t,j+1/2}^n = \frac{x_{t,j}^n + x_{t,j+1}^n}{2};$$

$$J_{j+1/2}^n = \frac{x_{j+1}^n - x_j^n}{h} = x_{q,j+1/2}^n; \quad J_j^n = \frac{J_{j+1/2}^n + J_{j-1/2}^n}{2}; \quad \vec{G}_{j+1/2}^n = \begin{pmatrix} 0 \\ g\,(Hh_q)_{j+1/2}^n \end{pmatrix};$$

$$h_{q,j+1/2}^n = \frac{h_{j+1}^n - h_j^n}{h}; \quad h_j^n = h(x_j^n).$$

The matrices $\mathcal{L}$, $\mathcal{R}$, $\bar{\Lambda}$ from the equation (2) are defined as:

$$\mathcal{L}_{j+1/2}^n = \frac{1}{(c_{j+1/2}^n)^2}\begin{pmatrix} -\lambda_{2,j+1/2}^n & 1 \\ -\lambda_{1,j+1/2}^n & 1 \end{pmatrix}, \quad \mathcal{R}_{j+1/2}^n = \frac{c_{j+1/2}^n}{2}\begin{pmatrix} -1 & 1 \\ -\lambda_{1,j+1/2}^n & \lambda_{2,j+1/2}^n \end{pmatrix},$$

$$\bar{\Lambda}_{j+1/2}^n = \Lambda_{j+1/2}^n - x_{t,j+1/2}^n\mathcal{E}, \quad \Lambda_{j+1/2}^n = \begin{pmatrix} \lambda_{1,j+1/2}^n & 0 \\ 0 & \lambda_{2,j+1/2}^n \end{pmatrix},$$

where $\mathcal{E}$ is the identity matrix; $\lambda_{k,j+1/2}^n$ $(k = 1, 2)$ are the eigenvalues of the matrix $\mathcal{A}_{j+1/2}^n$, which approximates the Jacobi matrix of the system (1):

$$\mathcal{A}_{j+1/2}^n = \begin{pmatrix} 0 & 1 \\ -u_j^n u_{j+1}^n + gH_{j+1/2}^n & 2u_{j+1/2}^n \end{pmatrix} = (\mathcal{R}\Lambda\mathcal{L})_{j+1/2}^n, \tag{4}$$

$$\lambda^n_{1,j+1/2} = (u-c)^n_{j+1/2}\,, \quad \lambda^n_{2,j+1/2} = (u+c)^n_{j+1/2}\,, \tag{5}$$

$$c^n_{j+1/2} = \sqrt{(u^n_{j+1/2})^2 - u^n_j u^n_{j+1} + gH^n_{j+1/2}} \geq \sqrt{gH^n_{j+1/2}} > 0\,.$$

For computing the elements of the diagonal matrix

$$\mathcal{D}^n_{j+1/2} = \begin{pmatrix} 1+\theta^1_{j+1/2} & 0 \\ 0 & 1+\theta^2_{j+1/2} \end{pmatrix}$$

the following formula for the scheme parameters was suggested in [3]:

$$\theta^k_{j+1/2} = \begin{cases} 0, & \left|g^k_{j+1/2}\right| \leq \left|g^k_{j+1/2-s_k}\right| \quad g^k_{j+1/2}\cdot g^k_{j+1/2-s_k} \geq 0, \\ \theta^k_{0,j+1/2}(1-\xi^k_{j+1/2}), & \left|g^k_{j+1/2}\right| > \left|g^k_{j+1/2-s_k}\right| \quad g^k_{j+1/2}\cdot g^k_{j+1/2-s_k} \geq 0, \\ \theta^k_{0,j+1/2}, & g^k_{j+1/2}\cdot g^k_{j+1/2-s_k} < 0, \end{cases} \tag{6}$$

where $k = 1, 2;\ \xi^k_{j+1/2} = g^k_{j+1/2-s_k}/g^k_{j+1/2};\ s_k \equiv s^n_{k,j+1/2} = \operatorname{sgn}\bar{\lambda}^n_{k,j+1/2}$;

$$g^k_{j+1/2} = \left(|\bar{\lambda}_k|(1-\mathrm{C}_k)p_k\right)^n_{j+1/2}; \quad \mathrm{C}^n_{k,j+1/2} = æ\Big(\frac{|\bar{\lambda}_k|}{J}\Big)^n_{j+1/2}; \quad \theta^k_{0,j+1/2} = \frac{1}{\mathrm{C}^n_{k,j+1/2}} - 1;$$

$p^n_{k,j+1/2}$ are the components of the vector $\vec{P}^n_{j+1/2}$; $\bar{\lambda}^n_{k,j+1/2} = \left(\lambda_k - x_t\right)^n_{j+1/2}$ are the diagonal matrices $\bar{\Lambda}^n_{j+1/2}$; $æ = \tau/h$. In the computations the modified formula (6) was used, obtained in [3] as the result of the entropy fix on the basis of the differential approximation method. This formula guarantees the non-negativity of approximation viscosity in the vicinity of the sonic point of depression wave.

The values $\mathrm{C}^n_{k,j+1/2}$ can be named as the local Courant numbers corresponding to the eigenvalues $\lambda^n_{k,j+1/2}$. As in the case of uniform fixed grids, the time step τ is chosen from considerations of stability of the computations on moving grids, meaning that the following condition should be fulfilled:

$$\max_{k,j} \mathrm{C}^n_{k,j+1/2} < 1.$$

In [3] the right-hand side of the equation (3) was computed as

$$\vec{G}^*_j = \begin{pmatrix} 0 \\ gH^*_j h^*_{q,j} \end{pmatrix}.$$

Therefore, one more equation had to be solved on the predictor step in addition to two equations of the system (2). This equation was obtained by the approximation of the first equation of the system (1) and used for computing the full depth H^*_j in integer nodes. For this additional equation, the values $\lambda^n_{k,j}$ and θ^k_j had to be evaluated in integer nodes using the complex formulas differing from (5), (6). These additional computations have been increasing the computational time for one-dimensional problems and especially – for two-dimensional problems.

Let us consider a more economic procedure for computing the right-hand side of the equation (3):

$$\vec{G}^*_j = \begin{pmatrix} 0 \\ g(Hh_q)^{n+1/2}_j \end{pmatrix}, \tag{7}$$

where

$$H_j^{n+1/2} = \frac{H_{j+1}^{n+1} + H_{j-1}^{n+1} + H_{j+1}^{n} + H_{j-1}^{n}}{4}; \quad h_{q,j}^{n+1/2} = \frac{h_{j+1}^{n+1} - h_{j-1}^{n+1} + h_{j+1}^{n} - h_{j-1}^{n}}{4h}.$$

It is interesting to note that despite the fact that now the values from the $(n+1)$-th time level are present in (7), the scheme (2), (3) stays explicit. It is not necessary to compute the values H_j^* in the modified algorithm. The corrector step (3) is performed separately for H_j^{n+1} and $(Hu)_j^{n+1}$, i.e. at first, the values H_j^{n+1} are found from the continuity equation, and then the impulse $(Hu)_j^{n+1}$ is obtained by the explicit formula with using of the computed value H_j^{n+1} in the right-hand side (7) of the impulse equation.

Thus, contrary to [3], two equations are solved now on the predictor step, reducing significantly the computation time. In addition, all advantages of the original predictorcorrector scheme [3] remain when the right-hand side is computed by the formula (7). In particular, the preservation a state of rest on a fixed grid is kept valid.

Lemma 1 *If a nonuniform grid $\{x_j^n\}$ is fixed, and a fluid is at rest on the n-th time level*

$$\eta_j^n \equiv 0, \quad u_j^n \equiv 0, \tag{8}$$

then the predictor-corrector scheme (2), (3), (7) preserves a state of rest on the $(n+1)$-th time level.

Proof. From the condition (8) and the condition of grid immovability $x_t \equiv 0$ it follows that

$$\left(\bar{\Lambda}\vec{P} - \mathcal{L}\vec{G}\right)_{j+1/2}^{n} = 0,$$

therefore, on the predictor step

$$\hat{\vec{f}}_{j+1/2} = \frac{\vec{f}_j^n + \vec{f}_{j+1}^n}{2}.$$

Then from the formula (3) of the corrector step, with taking into account (7) and the equalities $h_j^{n+1} = h_j^n$ and $J_j^{n+1} = J_j^n$ for a fixed grid, it follows that $\eta_j^{n+1} \equiv 0$, $u_j^{n+1} \equiv 0$. Thus, a state of rest is preserved on the $(n+1)$-th time level.

3 NEW METHOD FOR SWITCHING SCHEMES

The formula (6) governs the switching from one scheme to another according to a local behavior of the vector $\vec{P} = \mathcal{L}\vec{u}_q$. For example, when the first of two conditions in (6) is satisfied in both neighboring points $q_{j-1/2}$ and $q_{j+1/2}$ and for both values $k = 1$ and $k = 2$, the computation of predictor values (2) is done with the identity matrix $\mathcal{D} = \mathcal{E}$, i.e. the scheme (2), (3) coincides locally (in the node q_j) with the Lax-Wendroff scheme on moving grid. The known or new schemes are obtained for other correlations between the components of the vector $\vec{P}$ in neighboring nodes. In particular, for

$$\theta_{j+1/2}^k = \theta_{0,j+1/2}^k, \quad \theta_{j-1/2}^k = \theta_{0,j-1/2}^k, \quad k = 1, 2, \tag{9}$$

the scheme (2), (3) in the node q_j turns into the upwind scheme of the first approximation order.

The switching from one scheme to another by the formula (6) is actually based on the behavior analysis of the derivative $\vec{u}_q$ of the solution vector, i.e. the derivatives H_q and $(Hu)_q$ of the full depth and full impulse. Considering that $H_q = \eta_q + h_q$, the behavior of the bottom function $h(x)$ can have a strong influence on the choice of a scheme when using this switching method. Nevertheless, relying on their long-term experience of numerical computations of movement of surface waves over irregular bottoms, the authors of the current work came to the conclusion that a switcher of schemes should analyze not the variation of the full depth H_q, but the variation of the free surface elevation η_q and also the velocity derivative u_q. In this new switching method all formulas remain the same, except for the expressions for $g^k_{j+1/2}$ where the components $\widetilde{p}^n_{k,j+1/2}$ of the new vector $\widetilde{\mathbf{P}}$, which do not contain the derivative h_q, are used instead of the components of the vector $\vec{P}$.

From the definition of the vector $\vec{P}$ it follows that

$$\vec{P}^n_{j+1/2} = (\mathcal{L}\vec{u}_q)^n_{j+1/2} = \frac{1}{(c^n_{j+1/2})^2} \begin{pmatrix} -c\eta_q + Hu_q \\ c\eta_q + Hu_q \end{pmatrix}^n_{j+1/2} + \left(\frac{h_q}{c}\right)^n_{j+1/2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Therefore, the vector $\vec{P}$ is represented as the sum of two vectors. One of these vectors depends on the behavior of the solution (the derivatives η_q, u_q), but is independent of the depth variation h_q. This first vector is taken as a new vector $\widetilde{\mathbf{P}}$.

4 CONSERVATIVE SCHEMES ON MOVING GRIDS FOR THE SHALLOW WATER EQUATIONS

It is known [6, 17] that the non-conservative schemes can diverge on discontinuous solutions, therefore, the conservation property is one of the important requirements to numerical schemes. For non-homogeneous systems of (1) type the conservation of a difference scheme means the satisfaction of the difference analogues of the balance equations, which express the variation laws for the so-called conservative variables. For the system of shallow water equations (1) the conservative variables are the full depth H and the full impulse Hu. The balance equations for these variables describe the mass conservation law and the full impulse variation law. For a case of a flat bottom the full impulse variation law turns into the conservation law.

When constructing the conservative schemes on moving grids, the formulas for grid node velocities must be consistent with the formulas for cell areas. The special approximation of these variables guarantees the satisfaction of the difference analogue of the geometric conservation law [18, 19], which is the necessary condition for consistency of the formulas. In [3, 18] an original procedure was suggested for constructing conservative schemes for the scalar conservation law. This procedure was based on a choice of the scheme parameter θ corresponding to a concrete scheme. The same approach can be used for the systems of equations as well.

For example, according to [18], in order to derive the upwind scheme on moving grids for the shallow water equations, it is sufficient to use the formulas (9) for the parameters $\theta^k_{j+1/2}$ in the conservative predictor-corrector scheme (2), (3). Let us provide the obtained conservative upwind scheme on moving grid in the equivalent divergent and non-divergent forms.

If the parameters θ^k, given by the equalities (9), are used in the scheme predictor-corrector (2), (3), then

$$\mathcal{D}\bar{\Lambda} = \frac{hJ}{\tau}\mathcal{S}, \quad \mathcal{D}\bar{\Lambda}^2 = \frac{hJ}{\tau}\left(\bar{\Lambda}^+ - \bar{\Lambda}^-\right), \quad \mathcal{R}\mathcal{D}\bar{\Lambda}^2\mathcal{L} = \frac{hJ}{\tau}\left(\bar{\mathcal{A}}^+ - \bar{\mathcal{A}}^-\right), \quad \mathcal{R}\mathcal{D}\bar{\Lambda}\mathcal{L} = \frac{hJ}{\tau}\mathcal{R}\mathcal{S}\mathcal{L}$$

then the scheme turns into the upwind scheme in the divergent form:

$$\frac{(J\vec{u})_j^{n+1} - (J\vec{u})_j^n}{\tau} +$$

$$+\frac{1}{h}\Big[\Big(\frac{\vec{f}_{j+1}^n + \vec{f}_j^n}{2} - \frac{\left(\bar{\mathcal{A}}^+ - \bar{\mathcal{A}}^-\right)_{j+1/2}^n}{2}\left(\vec{u}_{j+1}^n - \vec{u}_j^n\right) - (x_t\vec{u})_{j+1/2}^n + \frac{h}{2}\left(\mathcal{RSL}\vec{G}\right)_{j+1/2}^n\Big) - \quad (10)$$

$$-\Big(\frac{\vec{f}_j^n + \vec{f}_{j-1}^n}{2} - \frac{\left(\bar{\mathcal{A}}^+ - \bar{\mathcal{A}}^-\right)_{j-1/2}^n}{2}\left(\vec{u}_j^n - \vec{u}_{j-1}^n\right) - (x_t\vec{u})_{j-1/2}^n + \frac{h}{2}\left(\mathcal{RSL}\vec{G}\right)_{j-1/2}^n\Big)\Big] = \vec{G}_j^*,$$

where the right-hand side is calculated using the formula (7),

$$\mathcal{D} = \frac{hJ}{\tau}\begin{pmatrix} 1/|\bar{\lambda}_1| & 0 \\ 0 & 1/|\bar{\lambda}_2| \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}, \quad \bar{\Lambda}^{\pm} = \begin{pmatrix} \bar{\lambda}_1^{\pm} & 0 \\ 0 & \bar{\lambda}_2^{\pm} \end{pmatrix},$$

$$\bar{\lambda}_k^{\pm} = \frac{\bar{\lambda}_k \pm |\bar{\lambda}_k|}{2}, \quad \bar{\mathcal{A}}^{\pm} = \mathcal{R}\bar{\Lambda}^{\pm}\mathcal{L},$$

and the following equalities are satisfied:

$$\bar{\Lambda} = \bar{\Lambda}^+ + \bar{\Lambda}^-, \quad \bar{\mathcal{A}}^+ + \bar{\mathcal{A}}^- = \bar{\mathcal{A}} = A - x_t\mathcal{E}.$$

The indices n and $j + 1/2$ are omitted here in order to simplify the notation.

The left part of the conservative upwind scheme (10) is written in the divergent form. Using the equivalent transformations this scheme can be written in the non-divergent form, which is analogous to the form of upwind schemes for scalar equations. The following equality is used:

$$\vec{f}_{q,j+1/2}^n = \left(\mathcal{A}\vec{u}_q\right)_{j+1/2}^n,$$

which is satisfied for the matrix $\mathcal{A}$ given by the formula (4), and the identity

$$(x_t\vec{u})_{q,j}^n = (x_{tq}\vec{u})_j^n + \frac{1}{2}\left[(x_t\vec{u}_q)_{j+1/2}^n + (x_t\vec{u}_q)_{j-1/2}^n\right],$$

where

$$(x_t\vec{u})_{q,j}^n = \frac{(x_t\vec{u})_{j+1/2}^n - (x_t\vec{u})_{j-1/2}^n}{h}; \quad x_{tq,j}^n = \frac{x_{t,j+1/2}^n - x_{t,j-1/2}^n}{h}.$$

As a result, the difference equation (10) takes on the form

$$\frac{(J\vec{u})_j^{n+1} - (J\vec{u})_j^n}{\tau} + \frac{1}{2}\Big[\mathcal{A}\vec{u}_q - \left(\bar{\mathcal{A}}^+ - \bar{\mathcal{A}}^-\right)\vec{u}_q - x_t\vec{u}_q + \mathcal{RSL}\vec{G}\Big]_{j+1/2}^n +$$
$$+\frac{1}{2}\Big[\mathcal{A}\vec{u}_q + \left(\bar{\mathcal{A}}^+ - \bar{\mathcal{A}}^-\right)\vec{u}_q - x_t\vec{u}_q - \mathcal{RSL}\vec{G}\Big]_{j-1/2}^n - (x_{tq}\vec{u})_j^n = \vec{G}_j^*.$$

Further, using the equality $\mathcal{A} = \mathcal{A}^+ + \mathcal{A}^-$ and the geometric conservation law [18]

$$J_j^{n+1} = J_j^n + \tau x_{tq,j}^n, \quad (11)$$

the final form of the upwind scheme in the non-divergent form is obtained:

$$\frac{\vec{u}_j^{n+1}-\vec{u}_j^n}{\tau}+\frac{1}{J_j^{n+1}}\Big[\left(\mathcal{A}^+\vec{u}_q\right)_{j-1/2}^n+\left(\mathcal{A}^-\vec{u}_q\right)_{j+1/2}^n\Big]=$$

$$=\frac{1}{J_j^{n+1}}\Big[\vec{G}_j^*-\frac{1}{2}\Big(\left(\mathcal{RSL}\vec{G}\right)_{j+1/2}^n-\left(\mathcal{RSL}\vec{G}\right)_{j-1/2}^n\Big)\Big]. \tag{12}$$

The explicit upwind scheme (12) is implemented as follows. At first, the full depth H_j^{n+1} is obtained from the first equation of the system (12) in all grid nodes. Then, using new values H^{n+1} in the right-hand side of (7), the full impulse $(Hu)_j^{n+1}$ is calculated. This scheme is conservative and preserves a state of rest.

Lemma 2 *The upwind scheme (12) preserves a state of rest (8) on a fixed grid.*

Proof. The preservation of a state of rest $\eta_j^{n+1}\equiv 0,\ u_j^{n+1}\equiv 0$ on the $(n+1)$-th time level follows from the formula (12) with taking into account the conditions (8), the equalities $x_t\equiv 0$ and $h_j^{n+1}=h_j^n$ for a fixed grid, and the formula for the right-hand side (7).

5 TESTING OF SCHEMES ON EXACT SOLUTION OF SHALLOW WATER EQAUTIONS

Some properties, which are inherent to the numerical solutions of linearized systems or scalar nonlinear equations, can not be strictly proved for the difference equations approximating the system of nonlinear shallow water equations. Therefore, it is very important to investigate the scheme properties using numerical tests.

In numerical modelling of a wave run-up on a coast one of the most popular tests is the problem of a wave run-up on a plane slope adjacent to a horizontal bottom with a constant depth [20, 22]. An initial wave is positioned above the horizontal bottom, propagates over it in the direction of the plane slope and runs up on the slope. For reliable estimates of inundation zones not only the quality algorithms for the detection of waterfront movement are needed, but also the reliable numerical methods for modelling of the initial stage of a wave movement including a propagation over a flat bottom. Let us present the results of this modelling using the described methods on adaptive grids.

In the considered test problem it is assumed that a bottom is horizontal $h(x)\equiv h_0$ and the initial conditions are given for $t=0$:

$$\eta(x,0)=\eta_0(x),\quad u(x,0)=u_0(x). \tag{13}$$

The numerical solution is compared with the exact solution of the Cauchy problem (1), (13), which can be found in [5]. In the case of the nonlinear dispersive equations [23, 24] the problem with the exact solution in the form of a solitary wave propagating over a flat bottom is included into the list of imperative tests. Yet, the exact solution [5] of the non-dispersive shallow water equations is rarely used for testing the numerical algorithms. The authors assume that the reason is the insufficient description of the method for obtaining the exact solution in [5]. Therefore, another way to derive the formulas of the exact solution is presented now.

Let us write down the shallow water equations (1) in Riemann invariants [7]:

$$r_t+\frac{3r+s}{4}r_x=0,\quad s_t+\frac{r+3s}{4}s_x=0, \tag{14}$$

where

$$r = u - 2c; \quad s = u + 2c; \quad c = \sqrt{gH}; \quad H = \eta + h_0. \tag{15}$$

The solution of the equations (14) is sought in the form of the r-wave propagating from right to left [25], i.e. under the assumption that $s \equiv s_0 = \text{const}$. Assuming that a fluid at infinity is in a state of rest, it follows that $s_0 = 2\sqrt{gh_0} = 2c_0$. Therefore, the formula $u + 2c = s_0$ leads to the expression of the velocity in terms of the full depth:

$$u(x,t) = 2c_0 - 2\sqrt{gH(x,t)} = 2c_0 - 2\sqrt{g\left[\eta(x,t) + h_0\right]}, \tag{16}$$

$$u_0(x) = u(x,0) = 2c_0 - 2\sqrt{gH_0(x)} = 2c_0 - 2\sqrt{g\left[\eta_0(x) + h_0\right]}. \tag{17}$$

If a new dependent variable is introduced:

$$p = \frac{3r + s_0}{4}, \tag{18}$$

then the equation for the invariant r is transformed into the Hopf-equaiton $p_t + pp_x = 0$. The solution of the Cauchy problem for this equation is found using the characteristics method and given implicitly by the formula [25]

$$p(x,t) = p_0(x - pt), \tag{19}$$

where p_0 is the initial function,

$$p_0(x) = \frac{3r(x,0) + s_0}{4} = \frac{3\left(u_0(x) - 2\sqrt{gH_0(x)}\right) + 2c_0}{4} = 2c_0 - 3\sqrt{gH_0(x)}. \tag{20}$$

Therefore, from (19) it follows the value $p(x,t)$ is the root p of the equation

$$p = 2c_0 - 3\sqrt{g\left(h_0 + \eta_0(x - pt)\right)} \tag{21}$$

for the given x, t. The obtained p is used for computing the free surface elevation $\eta(x,t)$. Taking into account the equalities (15), (16), (18), it is obtained that

$$\eta(x,t) = \left(\frac{2c_0 - p(x,t)}{3\sqrt{g}}\right)^2 - h_0. \tag{22}$$

Thus, the exact solution of the equations (1) is found using the following algorithm. The initial elevation $\eta_0(x)$ in the form of a continuous function is given for $t = 0$. The initial velocity $u_0(x)$, which is consistent with the initial elevation, is calculated using the formula (17). The root of the equation (21) is found at the time moment t in the point x. The elevation $\eta(x,t)$ is obtained using the formula (22), then the velocity $u(x,t)$ is calculated using (16).

With the help of the described method of characteristics it is possible to find the exact solution either for all $t > 0$ or until a gradient catastrophe moment t_*. Let us consider the initial elevation in the form of a solitary wave with the length λ [20]:

$$\eta_0(x) = \begin{cases} \dfrac{a}{2}\left[1 + \cos\left(\dfrac{2\pi(x - x_w)}{\lambda}\right)\right], & |x - x_w| \le \lambda/2, \\ 0, & |x - x_w| > \lambda/2, \end{cases} \tag{23}$$

where x_w is the abscissa of the wave top at $t = 0$, a is the wave amplitude. For $0 < t < t_*$ the length of the wave, moving to the left with the velocity $-c_0$, and its amplitude stay constant and equal to the corresponding values of the initial wave (23). The profile of the propagating wave will be deformed with time so that its leading edge is steepening and the trailing edge is flattening. This happens because the characteristics $dx/dt = p_0(x)$, going out from the interval $[x_w - \lambda/2,\, x_w]$, form a convergent bundle, and from the interval $[x_w,\, x_w + \lambda/2]$ – the divergent bundle. Therefore, the solution includes a compression wave and a rarefaction wave, which propagates before a compression wave and leads to the gradient catastrophe.

Fig. 1, *a* shows the exact solution $y = \eta(x, t)$ at $t = 5$ s (thin solid line). The non-dimensional (obtained by the division by h_0) values of parameters, defining the initial wave, are taken as follows:

$$a = 0.2, \quad x_w = 30, \quad \lambda = 10, \quad h_0 = 1.$$

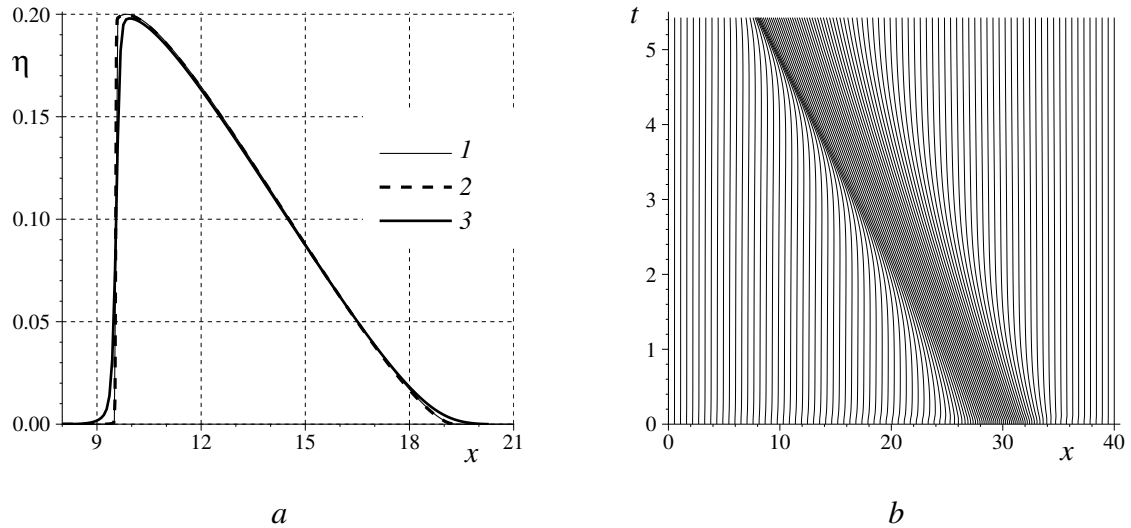
For these parameter values the gradient catastrophe occurs at $t_* \approx 5.57$ s, therefore, the leading edge becomes close to a vertical line at $t = 5$ s.

Let us note that the equation (21) has the unique solution p for $t < t_*$, because the function

$$f(p;\, x, t) = p + 3\sqrt{g\,(h_0 + \eta_0(x - pt))} - 2c_0 \tag{24}$$

is a monotonically increasing function of the variable p for all x and $t \in [0, t_*)$. Fig. 2 shows the function $f(p;\, x, t)$ for the fixed values $t = 5$ s and $x \in [0,\, L]$. Here $L = 40$ and the non-dimensional variable $\bar{p} = p/c_0$ is plotted along the axis Op. The solid line denotes the line of level $f = 0$. It is clear that the equation $f = 0$ has one root p for each fixed value of the variable x. Some method for solving the nonlinear equations can be used for finding this root. $[p_1,\, p_2]$ can be taken as the initial interval, which contains the root, where

$$p_1 = 2c_0 - 3\sqrt{g\,(h_0 + a)}, \quad p_2 = -c_0.$$



a *b*

Figure 1: Solitary wave (23) propagation over the horizontal bottom: *a* — exact solution (*1*) at $t = 5$ s and numerical solution, obtained using the predictor-corrector scheme (*2*) and the upwind scheme (*3*); *b* — the trajectories of the adaptive grid nodes

Fig. 1, *a* also shows the numerical solution, obtained using two schemes on adaptive grids on the interval $[0,\ L]$. The adaptive grids were constructed by the equidistribution method [18] with the control function

$$w^n_{j+1/2} = 1 + \alpha_0 \left|\eta^n_{j+1/2}\right| + \alpha_1 \frac{\left|\eta^n_{j+1} - \eta^n_j\right|}{x^n_{j+1} - x^n_j} \tag{25}$$

with the coefficient $\alpha_0 = 10$, $\alpha_1 = 10$. From Fig. 1, *b*, which presents the grid dynamics in the computation using the predictor-corrector scheme, it is seen that at $t = 0$ the nodes condense in the initial wave zone, then the continuous automatic redistribution of nodes occurs with time, and the condensation increases in the zone of the steepening leading wave edge. Such behavior of the adaptive grid increases the computational accuracy. Fig. 1, *a* shows that the exact solution and numerical solution, obtained by the predictor-corrector scheme, are visually indistinguishable, despite a rather small number of grid nodes ($N = 100$). The using of fixed uniform grids with the same number of nodes leads to a significant decrease of accuracy, being in good agreement with the conclusions of the works [3, 18, 19] where the solution of scalar equations on moving grids was considered.

Regarding to the upwind scheme, the using of adaptive grids also increases the accuracy in comparison with uniform grids, though not as much as for the predictor-corrector scheme (see Fig. 1, *a*, where the smearing of the wave borders and the decrease of its amplitude are seen for the upwind scheme).

6 ADAPTIVE GRID METHOD IN WAVE RUN-UP PROBLEMS

In [22] the adaptive grid method was used for the computation of the wave run-up in one-dimensional cross-sections drawn from some isobath to a chosen isobath on the dry land. The analytical solutions of the shallow water equations [20] were used for defining the waterfront point movement. These solutions were represented in the form of infinite series for one of possible cases of wave interaction with a coast and as the solution of the system of ordinary differential equations for two other cases. In the numerical computation [22] few first terms from the obtained series were used, and the Euler method of the second accuracy order was

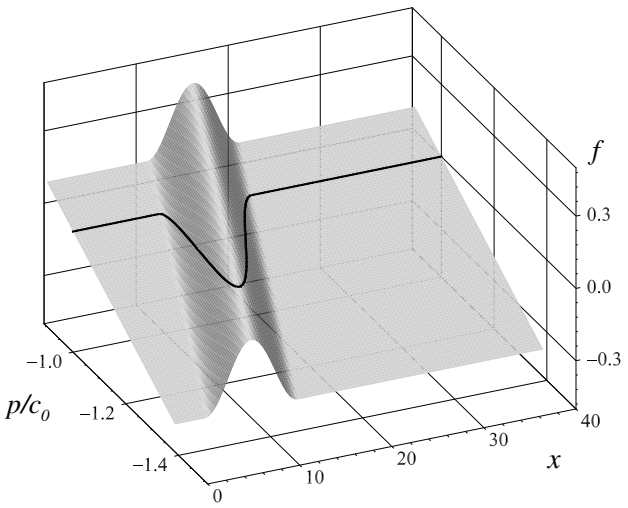


Figure 2: The function (24) at $t = 5$ s

used for the solution of the ordinary differential equations systems. In the present work, a larger number of terms in the partial sums of the series and a preciser methods for solving the ordinary differential equations are used.

6.1 Refined formulas for computing a waterfront point movement

Let us consider the wave run-up problem, where an unknown function $x = x_0(t)$, describing the law of the waterfront point movement, and the solution of the system of equations (1) for $x \geq x_0(t)$ are sought. Further, it is assumed that the initial data for the equations (1) are given not at $t = 0$, but at $t = t_0$, i.e. the initial position of the waterfront point $x_0(t_0) = x_{00}$ and the initial functions for $x \geq x_{00}$ are known

$$u(x, t_0) = u_0(x), \quad H(x, t_0) = H_0(x). \tag{26}$$

It is necessary to find $x = x_0(t)$, $u(x, t)$, $H(x, t)$ at $t > t_0$. The additional boundary condition in the waterfront point consists in the requirement that the full depth in this point is equal to zero at all time moments:

$$H(x_0(t), t) \equiv 0. \tag{27}$$

Depending on the initial data (26) three regimes are possible for wave interaction with a coast [20]. The first case is the run-up of a non-breaking wave on a coast:

$$H_x \mid_{x=x_0(t)} \neq 0. \tag{28}$$

Let us show that for the velocity of the waterfront point movement $U(t) = u(x_0(t), t)$ the equality:

$$U(t) = \dot{x}_0(t), \quad t \geq t_0. \tag{29}$$

is satisfied. Here and further a dot above is used to denote a time derivative of a function which depends on one variable t only. Let us introduce a new independent variable $\xi = x - x_0(t)$ and write down the system (1) in new coordinates (ξ, t):

$$\tilde{H}_t + (\tilde{u} - \dot{x}_0)\tilde{H}_\xi + \tilde{H}\tilde{u}_\xi = 0, \quad \tilde{u}_t + (\tilde{u} - \dot{x}_0)\tilde{u}_\xi + g\tilde{H}_\xi = g\tilde{h}_\xi, \quad \xi \geq 0, \tag{30}$$

where

$$h(x) = h\left(x(\xi, t)\right) = \tilde{h}(\xi, t),\ H(x, t) = H(x(\xi, t), t) = \tilde{H}(\xi, t),\ u(x, t) = u(x(\xi, t), t) = \tilde{u}(\xi, t).$$

The waterfront point is immovable in new variables (ξ, t) and it has the same coordinate $\xi = 0$ for all $t \geq t_0$, and here

$$\tilde{H}(0, t) \equiv 0. \tag{31}$$

Using the identity (31) in the first equation of the system (30) for $\xi = 0$, the identity's corollary $\tilde{H}_t(0, t) = 0$ and the condition (28), it is obtained that $\dot{x}_0 - \tilde{u}|_{\xi=0} = 0$, which coincides with the equality (29).

The law of movement of the waterfront point is sought in the form of the power series in time

$$x_0(t) = x_{00} + x_{01}(t - t_0) + x_{02}\frac{(t - t_0)^2}{2} + x_{03}\frac{(t - t_0)^3}{6} + x_{04}\frac{(t - t_0)^4}{24} + O\left((t - t_0)^5\right). \tag{32}$$

From the equality (29) it follows that

$$U(t) = \dot{x}_0(t) = x_{01} + x_{02}(t-t_0) + x_{03}\frac{(t-t_0)^2}{2} + x_{04}\frac{(t-t_0)^3}{6} + O\left((t-t_0)^4\right), \qquad (33)$$

i.e. $x_{01} = U(t_0) = u_0(x_0(t_0)) = u_0(x_{00}) = u_{00}$, where u_{00} denotes the fluid velocity in the waterfront point at the initial time moment.

The recursive formulas [20] can be used to find other coefficients. However, it is not convenient for the numerical realization of the algorithm, because one and the same term of the series has to be approximated with a different order, depending on how many terms are taken in the partial sum for using in the difference boundary condition. A more preferable approach is to use the explicit formulas for computing the coefficients x_{0k} ($k \geq 2$) in terms of the initial data of the problem.

In order to calculate the coefficient x_{02}, the second equation of the system (30) is considered for $\xi = 0$ and $t = t_0$: $\dot{U}(t_0) = -g\eta_0'(x_{00})$. On the other hand, from (33) it follows that $\dot{U}(t_0) = x_{02}$, therefore

$$x_{02} = -g\eta_0'(x_{00}). \qquad (34)$$

Here and further a prime denotes a space derivative of a function which depends on one variable x only.

For finding x_{03} the equations, obtained as the result of the differentiation of the first equation of the system (30) with respect to the variable ξ, and the second equation — with respect to the variable t:

$$\begin{aligned} \tilde{H}_{\xi t} + (\tilde{u} - \dot{x}_0)\tilde{H}_{\xi\xi} + 2\tilde{H}_\xi\tilde{u}_\xi + \tilde{H}\tilde{u}_{\xi\xi} = 0, \\ \tilde{u}_{tt} + (\tilde{u}_t - \ddot{x}_0)\tilde{u}_\xi + (\tilde{u} - \dot{x}_0)\tilde{u}_{\xi t} + g\tilde{H}_{\xi t} = g\dot{x}_0 h''. \end{aligned} \qquad (35)$$

Taking $\xi = 0$ in these equations and then taking $t = t_0$, the formula

$$x_{03} = \ddot{U}(t_0) = g\left(2u_0'H_0' + u_{00}h''\right)\Big|_{x=x_{00}} \qquad (36)$$

is derived.

The coefficient x_{04} is computed using the equations (30), (36), and the following equations, which are obtained from them by the differentiation with respect to the variables t and ξ:

$$\begin{aligned} &\tilde{H}_{\xi tt} + (\tilde{u}_t - \ddot{x}_0)\tilde{H}_{\xi\xi} + (\tilde{u} - \dot{x}_0)\tilde{H}_{\xi\xi t} + 2\tilde{H}_{\xi t}\tilde{u}_\xi + 2\tilde{H}_\xi\tilde{u}_{\xi t} + \tilde{H}_t\tilde{u}_{\xi\xi} + \tilde{H}\tilde{u}_{\xi\xi t} = 0, \\ &\tilde{u}_{ttt} + (\tilde{u}_{tt} - \dddot{x}_0)\tilde{u}_\xi + 2(\tilde{u}_t - \ddot{x}_0)\tilde{u}_{\xi t} + (\tilde{u} - \dot{x}_0)\tilde{u}_{\xi tt} + g\tilde{H}_{\xi tt} = g\left[\ddot{x}_0 h'' + (\dot{x}_0)^2 h'''\right], \\ &\tilde{u}_{\xi t} + (\tilde{u}_\xi)^2 + (\tilde{u} - \dot{x}_0)\tilde{u}_{\xi\xi} + g\tilde{H}_{\xi\xi} = gh''. \end{aligned} \qquad (37)$$

If $\xi = 0$ is taken in these equations, and the equalities (29), (31) and their corollary $\dot{U}(t) = \ddot{x}_0(t)$, $\ddot{U}(t) = \dddot{x}_0(t)$, $\tilde{H}_t(0,t) = 0$ are taken into account, then the system of equations in the point $\xi = 0$ is obtained:

$$\begin{gathered} \tilde{H}_{\xi tt} + 2\tilde{H}_{\xi t}\tilde{u}_\xi + 2\tilde{H}_\xi\tilde{u}_{\xi t} = 0, \quad \tilde{H}_{\xi t} + 2\tilde{H}_\xi\tilde{u}_\xi = 0, \quad \tilde{u}_{\xi t} + (\tilde{u}_\xi)^2 + g\tilde{H}_{\xi\xi} = gh'', \\ \ddot{U} + g\tilde{H}_{\xi tt} = g\left[\ddot{x}_0 h'' + (\dot{x}_0)^2 h'''\right]. \end{gathered}$$

If the value $\tilde{H}_{\xi tt}$ is eliminated from the last equation using the first three equations, and $t = t_0$ is set, then the expression is obtained for the coefficient x_{04} in terms of the initial data:

$$x_{04} = \ddot{U}(t_0) = g\Big[u_{00}^2 h''' - \eta_0' h'' - 2H_0' \left(g\eta_0'' + 3(u_0')^2\right)\Big]\Big|_{x=x_{00}}. \tag{38}$$

If at the time moment $t = t_0$ the condition (28) is not satisfied, then two situations are possible: either $H_x(x_{00}, t_0) = 0$ or $H_x(x_{00}, t_0) = \infty$. The first case corresponds to a non-breaking wave run-up, for which the tangential line to free surface at the waterfront point coincides at $t = t_0$ with the tangential line to the bottom surface. The second case corresponds to a breaking wave run-up. In both cases the motion law for the waterfront point is defined as the solution of the system of ordinary differential equation [20]

$$\dot{x}_0 = U(t), \quad \dot{U} = gh'(x_0(t)) \tag{39}$$

with the initial conditions

$$x_0(t_0) = x_{00}, \quad U(t_0) = u_*, \tag{40}$$

here $u_* = u_{00}$ for a wave with tangency and $u_* = u_{00} - 2\sqrt{gH_{00}}$ for a breaking wave.

6.2 Computational algorithm

For the numerical solution of the wave run-up problem the above predictor-corrector scheme on adaptive grid $\left\{x_j^n\right\}$ $(j = 0, \ldots, N)$ is used. At every time level the left nodes x_0^n coincides with the moving waterfront point $x_0(t^n)$. It is assumed that the velocity, the full depth, and the position of the waterfront point are known at the time level $t = t^n$, and it is required to find these values on the next time level $t^{n+1} = t^n + \tau$.

In order to use the predictor-corrector scheme for computing the free surface elevation η_j^{n+1} and the velocity u_j^{n+1}, it is first necessary to generate the grid $\left\{x_j^{n+1}\right\}$. Therefore, a new position of the waterfront point, i.e. the far left node x_0^{n+1}, has to be known. The right boundary is assumed to be fixed, therefore, $x_N^{n+1} = L$. The difference non-reflecting boundary condition [22] is given in this node and used to obtain the values η_N^{n+1} and u_N^{n+1}.

In the non-degenerate case (28), the difference analogues of the partial sums of the series (32) and (33) are used to compute the coordinate x_0^{n+1} and the velocity u_0^{n+1}:

$$x_0^{n+1} = x_0^n + u_0^n\tau + x_{02}\frac{\tau^2}{2} + x_{03}\frac{\tau^3}{6} + x_{04}\frac{\tau^4}{24}, \quad u_0^{n+1} = u_0^n + x_{02}\tau + x_{03}\frac{\tau^2}{2} + x_{04}\frac{\tau^3}{6}. \tag{41}$$

The difference formulas for the coefficients x_{0k} $(k = 2, 3, 4)$ are obtained as the result of the approximation of the derivatives from the right-hand sides of the equalities (34), (36), (38) using the one-sided differences of corresponding order in the left node x_0^n of the non-uniform grid.

The first formula (41) shows that in order to calculate the position of the waterfront point x_0^{n+1} with the accuracy order $O(\tau^5)$ it is sufficient to approximate the coefficient x_{02} with the error $O(h^3)$, the coefficient x_{03} — with $O(h^2)$, and x_{04} — with $O(h^1)$. Here it is assumed that the law of passage to the limit has the form æ = const. Therefore, in the formula for the coefficient

$$x_{02} = -g\frac{\eta_{q,0}^n}{x_{q,0}^n} \tag{42}$$

the finite differences of the third order are used

$$\eta^n_{q,0} = \frac{-11\eta^n_0 + 18\eta^n_1 - 9\eta^n_2 + 2\eta^n_3}{3h}, \quad x^n_{q,0} = \frac{-11x^n_0 + 18x^n_1 - 9x^n_2 + 2x^n_3}{3h}.$$

As in [18], it is assumed that the nodes x^n_j of a non-uniform moving grid are the images of the nodes $q_j \in [0,\, 1]$ of a uniform fixed grid with the step $h = 1/N$ under some coordinate mapping $x = x(q,t)$. At each time moment t the mapping transforms in one-to-one manner the interval $[0,\, 1]$ to $[x_0(t),\, L]$. Using the transformation $x = x(q,t)$ the derivatives with respect to x can be expressed in terms of the derivatives with respect to q, as for instance $\eta_x = \eta_q/x_q$. On the discrete level it means that the problem of the approximation of the derivatives on a non-uniform moving grid x^n_j can be replaced by a simpler problem of their approximation of a uniform fixed grid q_j. This is used in the formula (42) according to which the approximation formula for the coefficient (36) has the following form:

$$x_{03} = g\frac{1}{(x^n_{q,0})^2}\left[2u^n_{q,0}H^n_{q,0} + u^n_0\left(h^n_{qq,0} - h^n_{q,0}\frac{x^n_{qq,0}}{x^n_{q,0}}\right)\right]. \tag{43}$$

As it was stated above, it is sufficient to approximate the derivatives entering into this formula by the one-sided differences of the second order:

$$u^n_{q,0} = \frac{-3u^n_0 + 4u^n_1 - u^n_2}{2h}, \quad H^n_{q,0} = \frac{-3H^n_0 + 4H^n_1 - H^n_2}{2h}, \quad x^n_{q,0} = \frac{-3x^n_0 + 4x^n_1 - x^n_2}{2h},$$

$$h^n_{qq,0} = \frac{2h^n_0 - 5h^n_1 + 4h^n_2 - h^n_3}{h^2}, \quad x^n_{qq,0} = \frac{2x^n_0 - 5x^n_1 + 4x^n_2 - x^n_3}{h^2}.$$

As far as the derivatives from the formula (38) have to be calculated with the first order only, it is easier to write down the approximation formula for x_{04} immediately on a non-uniform grid $\{x^n_j\}$:

$$x_{04} = g\Big[(u^n_0)^2 h^n_{xxx,0} - \eta^n_{x,0}h^n_{xx,0} - 2H^n_{x,0}\left(g\eta^n_{xx,0} + 3(u^n_{x,0})^2\right)\Big], \tag{44}$$

where the following formulas are used for the first difference derivatives:

$$H^n_{x,0} = \frac{H^n_1 - H^n_0}{x^n_1 - x^n_0}, \tag{45}$$

and the known formulas [17] of the first order on a non-uniform grid are used for the second and third difference derivatives.

For a plane slope

$$z = -h(x) = \begin{cases} z_0 - x\,\mathrm{tg}\,\theta & \text{for} \quad 0 \le x \le x_s, \\ -1 & \text{for} \quad x_s \le x \le L, \end{cases} \tag{46}$$

adjacent to a horizontal bottom of a constant depth h_0, the obtained formulas significantly simplify, because $h' = \mathrm{tg}\theta$, $h'' = h''' = 0$. In the formula (46) $\theta > 0$ is the slope angle, $z_0 > 0$ is the height of the "dry land" in the point $x = 0$, $x_s = (z_0 + 1)\mathrm{ctg}\,\theta$ is the abscissa of the slope bed. The value of z_0 is adjusted so that the maximal wave run-up on a coast does not exceed this value.

As it was shown above, for the second and third regimes of wave interaction with a coast, the motion law of the waterfront point is defined as the solution of the system of two ordinary differential equations (39). The Euler method was used to solve this system in [20]. In the present work the Runge-Kutta method of the fourth accuracy order is used with the initial conditions $x_0(t^n) = x_0^n$, $U(t^n) = u_*$ for the second regime, and $u_* = u_1^n - 2\sqrt{gH_1^n}$ – for the third one.

In the case of a plane slope (46) the system (39) has the exact analytical solution

$$x_0^{n+1} = x_0^n + \tau u_* + \frac{\tau^2}{2} g \,\mathrm{tg}\theta, \quad u_0^{n+1} = u_* + \tau g \,\mathrm{tg}\theta \tag{47}$$

and both methods (Euler and Runge-Kutta) give the same results.

As in [22], in numerical computation a concrete regime of wave interaction with a coast is defined according to the value of the difference derivative $H_{x,0}^n$ in the node x_0^n of a moving grid. In [22] this derivative was calculated with the help of the one-sided difference of the first order (45). In the present work the formula of the second order

$$H_{x,0}^n = \frac{-3H_0^n + 4H_1^n - H_2^n}{-3x_0^n + 4x_1^n - x_2^n}$$

is used. If this difference derivative satisfies the condition $m \le \left|H_{x,0}^n\right| \le M$, where m, M are the given numbers ($0 < m \ll M$), then it is assumed that the first regime is realized, and for the condition $\left|H_{x,0}^n\right| < m$ – the second regime. The inequality $\left|H_{x,0}^n\right| > M$ is used as the criteria for the third regime.

The above formulas for the position of a moving waterfront point x_0^{n+1} and its velocity u_0^{n+1} depend on a regime of wave interaction with a coast, whereas the full depth in the far left grid node is equal to zero $H_0^{n+1} = 0$ for all regimes.

Lemma 3 *If the position of the waterfront point and its velocity are defined by the formulas (41)–(44), then the predictor-corrector scheme (2), (3), (7) preserves a state of rest (8).*

Proof. Under the rest condition (8) the first regime of wave interaction with a coast is realized ($m \le \left|H_{x,0}^n\right| \le M$). It is easy to check that with the use of the formulas (41)–(44) the waterfront point has a zero velocity ($u_0^{n+1} = 0$) and the same position ($x_0^{n+1} = x_0^n$) at the time moment t^{n+1}. Then, from the boundary condition $H_0^{n+1} = 0$ it follows that the free surface elevation does not change in this point: $\eta_0^{n+1} = \eta_0^n = 0$. Further, the proof is the same as in the lemma 1.

6.3 Some numerical results of modelling of a solitary wave run-up on a plane slope

Let us consider the results of the numerical solution of the problem on a solitary wave run-up on a plane slope. A solitary wave is given at the initial time moment $t = 0$:

$$\eta_0(x) = a_0 \mathrm{ch}^{-2}\left(\frac{\sqrt{3a_0 g}}{2U_0}(x - x_w)\right), \quad H_0(x) = h(x) + \eta_0(x), \quad u_0(x) = -U_0 \frac{\eta_0(x)}{H_0(x)}, \tag{48}$$

where a_0 is the amplitude of the wave, $x = x_w$ is the abscissa of the wave top, $U_0 = \sqrt{g(h_0 + a_0)}$, $x \in [0,\, L]$. A plane slope is given by the formula (46) with the value $z_0 = 1$.

At the initial time moment the wave top (48) was well spaced from the slope bed x_s and from the right boundary $x = L$. This guaranteed the position of the wave main part above the horizontal part of the bottom and the independence of the run-up picture on the boundary conditions on the right boundary up to some time moment. As the computations have shown,

the wave can break in a long-continued propagation above a horizontal bottom. Thus, if the interaction of a non-deformed solitary wave with a slope has to be investigated in the framework of the shallow water model, then a wave should not be initially positioned too far from a slope.

For a small amplitude a_0 (for instance, $a_0 = 0.01$) the initial solitary wave has a larger effective length, therefore, the following values of the top abscissa of the initial wave and the right boundary were chosen in the numerical experiments:

$$x_w = x_s + 30, \quad L = x_w + 30.$$

For shorter waves (for instance, $a_0 = 0.1$) the point $x = x_w$ was positioned closer to the slope and the domain length was smaller:

$$x_w = x_s + 20, \quad L = x_w + 20.$$

Fig. 3 shows the maximal wave amplification (the maximal wave run-up on a plane slope taken relative to the amplitude a_0 of the initial wave). The markers *1* correspond to the refined formulas for computing the waterfront point movement, derived in the present work. The markers *4* are obtained when using the formulas from [20, 22], i.e. the first three terms in the partial sums (41). For the small amplitude a_0 the results are nearly the same (Fig. 3, *a*). For the initial amplitude, increased ten times, the difference becomes visible, especially for small slope angles (Fig. 3, *b*). The computations with other wave amplitudes have proved the conclusion that for small slope angles and large values a_0 the refined formulas gave the results which differ from those obtained by the previous formulas.

In many works, devoted to the investigation of a wave run-up on a coast, the comparison is done with the values obtained by the analytical formula [21]

$$\frac{R}{a_0} = 2.831\sqrt{\operatorname{ctg}\theta}\left(\frac{a_0}{h_0}\right)^{1/4}. \tag{49}$$

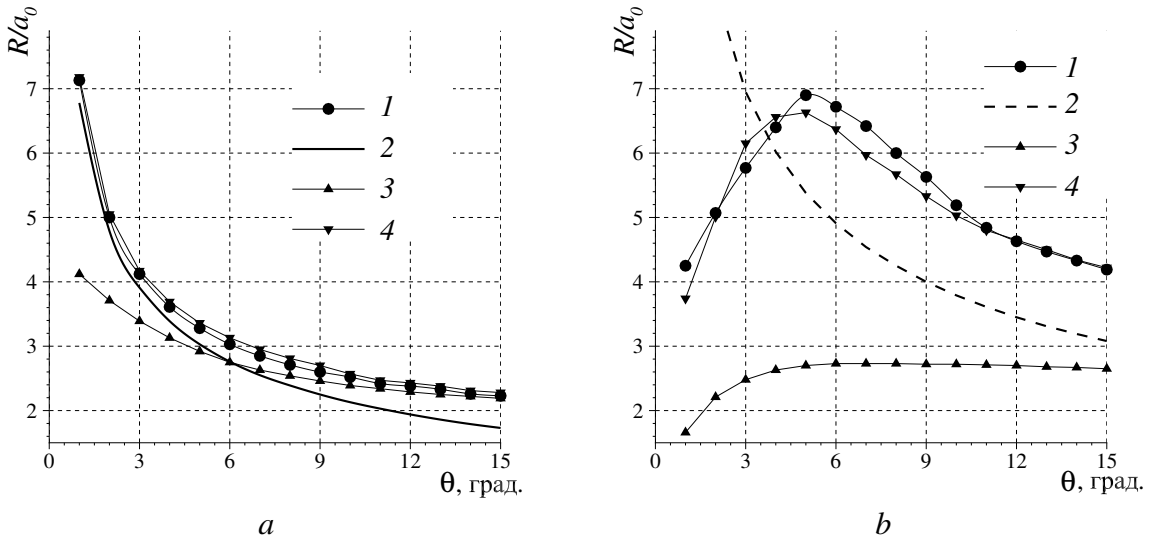


Figure 3: The maximal wave amplification with respect to a slope angle: $a_0/h_0 = 0.01$ (*a*), 0.1 (*b*). *1* (the refined formulas for the waterfront point movement,) *4* (the formulas from [20, 22]) — the numerical computation for the plane slope; *2* — the analytical dependency (49); *3* — the numerical computation for the vertical wall, positioned at the depth $h_{\text{wall}} = a_0$

However, this formula has to be used carefully, because it has a very narrow domain of applicability. According to [21, 26] the following restrictions are the conditions of applicability:

$$\frac{a_0}{h_0} \gg (0.288 \operatorname{tg}\theta)^2, \tag{50}$$

$$\frac{a_0}{h_0} < 0.479(\operatorname{tg}\theta)^{10/9}. \tag{51}$$

Fig. 3 shows the values (lines *2*) of the maximal wave amplification obtained by the formula (49). If it is held that the lower boundary of the applicability domain is defined by the expression in the right-hand side of the inequality (50), then nearly the whole line *2* in Fig. 3, *a* (except for the small upper part corresponding to angles in the vicinity $\theta = 1^\circ$) lies in the domain of applicability.

It is known that the value of the maximal amplification R/a_0 for a solitary wave run-up on a vertical wall is greater than two [27], and on a plane slope – greater than on a vertical wall [14]. However, for the angles greater than 11° the formula (49) gives the value less than two (Fig. 3, *a*, line *2*), i.e. even less than for the run-up on the vertical wall (markers *3*). This causes doubts in the physical consistency of the formula for $a_0/h_0 = 0.01$ and $\theta > 11^\circ$. Thus, the analytical formula (49) is applicable only for small angles θ and, therefore, in the restriction (50) the sign $\gg$ means that the expression in the right-hands side must be significantly smaller than the expression on the left. Hence, even for such small initial amplitude of a running wave it is possible to compare with the analytical formula only for this small range of angles θ, and the range can not be determined by the equality (50).

If this range exists, then Fig. 3, *b* shows that the refined formulas for the waterfront point movement (markers *1*) give the numerical run-up values which are slightly closer to analytical ones than those obtained by the formulas from [20, 22] (markers *4*). For the amplitude $a_0/h_0 = 0.1$ the analytical formula (49) is not applicable because of the restrictions (50), (51). Therefore, it makes no sense to use it for the comparison with numerical results. In order to emphasize this fact, the analytical dependency (49) is shown in Fig. 3, *b* by the dash line *2*.

The interesting peculiarity of the dependency of the maximal wave amplification R/a_0 on the angle θ is its non-monotonic character for large amplitudes (markers *1* in Fig. 3, *b*). The reason for the non-monotonicity is the wave breaking on the plane slope. Fig. 4, *a* shows the wave run-up on a plane slope and the subsequent run-down. Let us note, that the wave with negative polarity and significant amplitude is generated near the initial waterfront point position after the main run-down. This wave produces the whole series of run-ups and run-downs with damping amplitudes, that is physically explainable but has not been noted in the previous publications on the subject. The same figure shows that the leading edge of the initial wave is steepening when approaching to the waterfront point.

For the detailed investigation of this stage of run-up processes, Fig. 4, *b* shows the free surface elevation profiles at the initial time moment, wave breaking moment, and run-up start moment. By the moment of breaking the wave amplitude (line *2*) has become greater than the initial one (line *1*), and its leading edge has become nearly vertical. Here the wave has not yet reached the initial waterfront point position (left circle on the axis Ox), and the wave top has made the way slightly greater than one third of the slope length, counting from the starting point of the slope (right circle on the axis Ox).

After the wave breaking moment the wave amplitude and velocity decrease, and the wave actually approaches the waterfront point with a significantly smaller height than before the

breaking moment and, in this case, even smaller than the initial height a_0 (line *3*). This causes the fact that for small slope angles θ the run-up can be smaller than for larger angles. Let us note that this does not occur for small initial amplitudes a_0. Here the wave is very long, has low gradient slopes, and is not able to break even for small slope angles θ, i.e. on the long slope.

Let us also note that for the considered amplitude $a_0/h_0 = 0.1$ the wave breaking on the flat slope occurs for the slope angle $\theta_* \approx 6^\circ$. The formula (51) gives the threshold value $\theta_* \approx 12^\circ$. There exist other theories which provide another criteria for breaking and another threshold values accordingly. In the considered case the deviation of threshold values in numerical and analytical computations can be explained by the fact that the already deformed wave, with a steeper leading edge which is greater than that one of the initial wave (48), actually reaches the slope.

In many works, where the real problems are solved, the modelling is done for a run-up not on a coast but on a vertical wall positioned on a coast line. This significantly simplifies the computational algorithm, but leads to the errors in computations of the maximal wave amplification and, consequently, to the errors in the definition of inundation zones.

In one-dimensional case, a coast is substituted by a vertical wall in the initial position of the waterfront point $x_{00} = z_0 \mathrm{ctg}\theta$, and the impermeability condition $u(x_{00}, t) = 0$ is set meaning that the mobility of the waterfront point is ignored. The bottom (46) in the vicinity of the wall is modified so that if the depth becomes less than some given value h_{wall}, then it is assumed that $h(x) = h_{\text{wall}}$. In other words, the bottom in the vicinity of the wall becomes horizontal with the depth h_{wall}, i.e. the bottom is "cut" and straighten in the vicinity of the point x_{00}. If the depth is greater than h_{wall}, then the bottom form does not change.

Multiple computations were done for the problem on a wave run-up on a vertical wall. Some results are presented in Fig. 3 (markers *3*). It is seen that, as in the case of a wave run-up on a coast, the dependency of the maximal wave amplification R/a_0 on bottom slope angles has monotonic character for small amplitudes and non-monotonic – for large ones. In the last case the reason is still the breaking of the wave on the slope even before approaching the vertical

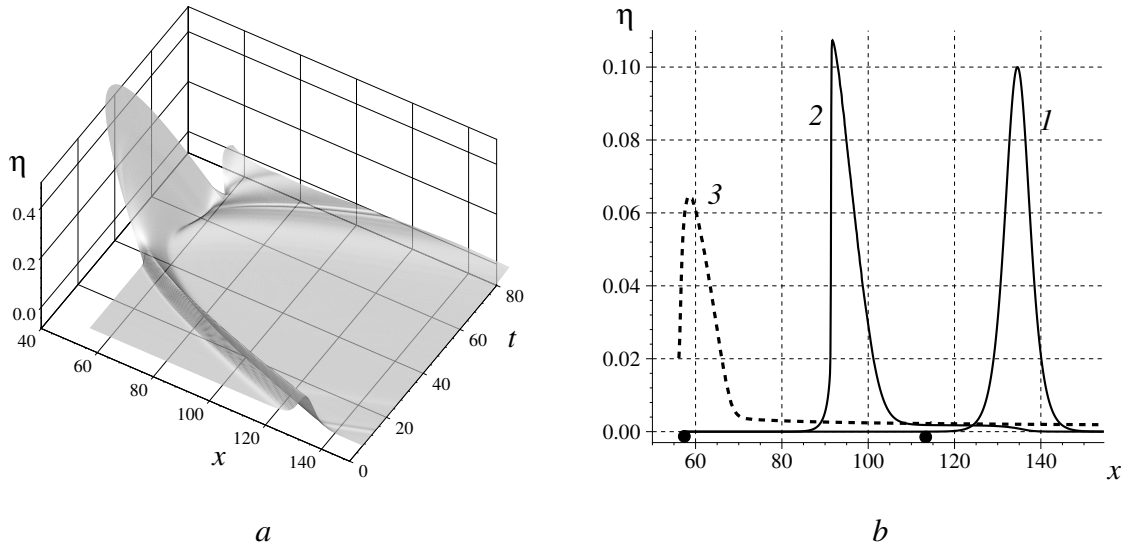


Figure 4: The run-up of the solitary wave (48) with the amplitude $a_0/h_0 = 0.1$ on the plane slope with the slope angle $\theta = 1^\circ$: *a* — the free surface elevation $z = \eta(x, t)$; *b* — the free surface elevation $\eta(x, t)$ at the initial time moment $t = 0$ (*1*), at the wave breaking moment $t = 12.64$ s (*2*), at the run-up start moment $t = 29.97$ s (*3*)

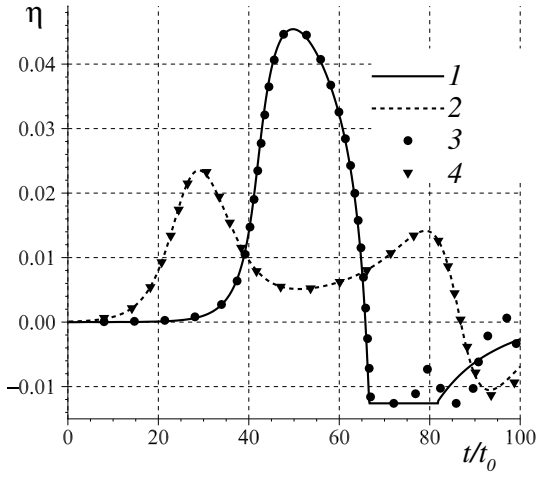


Figure 5: The mareograms, obtained analytically (lines) and numerically (markers) in the points, positioned to the right of the initial position of the waterfront point at the distances 0.25 (*1, 3*) and 9.95 (*2, 4*)

wall. For small amplitudes and large slope angles the values of the maximal wave amplification are practically indistinguishable for the run-up on the wall and for the run-up on the coast, but for large amplitudes they differ even for large slope angles. The greatest difference is observed for $\theta \approx 6^\circ$ (see Fig. 3, *b*). Here the run-up on the coast is significantly larger than on the vertical wall.

By now, a large set of test problems [28] has been formed, which are imperative for validation check of the results obtained by the numerical algorithms and for comparison of the characteristics of different algorithms [29]. One of these tests specifies the comparison of numerical results with the analytical solution of the problem on the run-up of the solitary wave with the small amplitude $a_0/h_0 = 0.019$ on the low gradient plane slope (46) with the slope angle $\theta = \arctan(1/19.85) \approx 2.88^\circ$.

Fig. 5 presents the mareograms, recorede by two virtual mareographs positioned at the distances 0.25 and 9.95 to the right of the initial position of the waterfront point x_{00}. For the first mareograph ($x = x_{00} + 0.25$), positioned near the initial waterfront point, there exist a time interval when the bottom drying occurs. The figure depicts this fact by the horizontal line drawn on the height $\eta \approx -0.012594$. In the position of the second mareograph ($x = x_{00} + 9.95$) the bottom drying does not occur. Figure shows that the numerical results, obtained using the predictor-corrector scheme, are visually indistinguishable from the analytical solution in the run-up phase and in the major part of the run-down phase. The difference can be seen only in the time interval from $t = 75$ to $t = 100$, where t is the non-dimensional time obtained by the division of the dimensional time by the value $t_0 = \sqrt{g/h_0}$. This difference can be attributed to the fact that the analytical formula for the exact solution, which was derived for non-breaking waves, becomes incorrect on this time interval because of the wave breaking at the end of the run-down phase. In fact, according to the conclusions in [26], if the condition (51) is fulfilled, then the wave does not break in the run-down phase. But in the considered case this condition is violated, because $a_0/h_0 = 0.019$ and the right-hand side of the inequality (51) takes the value equal to 0.0173.

7 CONCLUSIONS

An improved adaptive grid method is considered for the numerical solution of the problems on propagation and run-up of surface waves, described by the one-dimensional shallow water model. The modified algorithm for the realization of the explicit predictor-corrector scheme is presented, which is based on the new way of computation of the right-hand side of the shallow water equations on irregular bottom. The algorithm provides savings in computational time in comparison with its earlier version [3] while preserving the approximation order. Also the algorithm guarantees the preservation of the state of rest in transition from one time level to a next one. A new method for choosing the scheme parameters on the basis of the analysis of the differential approximation is suggested that guarantees the satisfaction of the TVD-property for the improved predictor-corrector scheme. The presented method for construction of different conservative schemes on moving grids is based on an appropriate choice of the scheme parameters for the predictor–corrector scheme, which represents the canonical form of the two-layer explicit schemes for the shallow water equations. As an example, a conservative upwind scheme on moving grid is provided in the divergent and non-divergent forms. The properties of the upwind scheme and the predictor–corrector scheme on dynamically adaptive grids are demonstrated for the exact solution of the nonlinear shallow water equations

Using the known analytical solutions of the shallow water equations in the vicinity of the water-land boundary the improved difference boundary conditions are obtained at the moving waterfront point. These boundary conditions approximate the analytical solutions with a higher accuracy than the conditions used earlier [20]. It is proved that if a fluid is at rest and has a non-perturbed free boundary at the initial time moment, then the difference predictor-corrector scheme on adaptive grid preserves the state of rest at all subsequent time moments when the newly obtained conditions are used. This is one of the advantages of the developed boundary conditions in comparison with the known shock-capturing methods, where the preservation of the state of rest is usually problematic for the run-up problems. The numerical experiments have shown that for the run-up problems the substitution of a slope by a vertical wall in the initial position of the waterfront point leads to the significant change of the wave amplification in the case of very smooth slopes even if a wall embedding is small.

The obtained results will find the application in the solution of two-dimensional problems in the framework of the shallow water model and in the algorithms for solution of the nonlinear dispersive equations [23, 24].

Funding: the work was supported by the Russian Scientific Foundation (project No. 14-17-00219).

REFERENCES

[1] Yu.I. Shokin, Z.I. Fedotova, G.S. Khakimzyanov, Hierarchy of nonlinear models of the hydrodynamics of long surface waves. *Doklady Physics*, **60(5)**, 168–172, 2015.

[2] E. Toro, P. García-Navarro, Godunov-type methods for free-surface shallow flows: A review. *Journal of Hydraulic Research*, **45(6)**, 736–751, 2007.

[3] G.S. Khakimzyanov, N.Yu. Shokina, Adaptive grid method for one-dimensional shallow water equations. *Computational Technologies*, **18(3)**, 54–79, 2013. (in Russ.)

[4] M.E. Vazquez-Cendon, Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry. *Journal of Computational Physics*, **148(2)**, 497–526, 1999.

[5] N.E. Voltsinger, E.N. Pelinovskii, K.A. Klevannyi, *The long-wave dynamics of the coastal zone*. Leningrad: Gidrometeoizdat, 1989. (in Russ.)

[6] R.J. LeVeque, *numerical methods for conservation laws*. Berlin: Birkhäuser Verlag, 1992.

[7] E.F. Toro, *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Berlin: Springer-Verlag, Springer-Verlag, 2009.

[8] Lynch D.R. Gray W.G., Finite element simulation of flow in deforming regions. *Journal of Computational Physics*, **36**, 135–153, 1980.

[9] J. Hou, Q. Liang, H. Zhang, R. Hinkelmann, An efficient unstructured MUSCL scheme for solving the 2D shallow water equations. *Environmental Modelling & Software*, **66**, 131–152, 2015.

[10] J. Hou, Q. Liang, . Simons, R. Hinkelmann, A stable 2D unstructured shallow flow model for simulations of wetting and drying over rough terrains. *Computers & Fluids*, **82**, 132–147, 2013.

[11] A. Duran, F. Marche, Recent advances on the discontinuous Galerkin method for shallow water equations with topography source terms. *Computers & Fluids*, **101**, 88–104, 2014.

[12] S.W. Funke, C.C. Pain, S.C. Kramer, M.D. Piggott, A wetting and drying algorithm with a combined pressure/free-surface formulation for non-hydrostatic models. *Advances in Water Resources*, **34(11)**, 1483–1495, 2011.

[13] T. Kärnä, B. de Brye, O. Gourgue, J. Lambrechts, R. Comblen, V. Legat, E. Deleersnijder, A fully implicit wetting-drying method for DG-FEM shallow water models, with an application to the Scheldt Estuary. *Computer Methods in Applied Mechanics and Engineering*, **200(5-8)**, 509–524, 2011.

[14] Y. Li, F. Raichlen, Non-breaking and breaking solitary wave run-up. *Journal of Fluid Mechanics*, **456**, 295–318, 2002.

[15] E.T. Flouri, N. Kalligeris, G. Alexandrakis, N.A. Kampanis, C.E. Synolakis, Application of a finite difference computational model to the simulation of earthquake generated tsunamis. *Applied Numerical Mathematics*, **67**, 111–125, 2013.

[16] Yu.I. Shokin, S.A. Beisel, A.D. Rychkov, L.B. Chubarov, Numerical simulation of the tsunami runup on the coast using the method of large particles. *Mathematical Models and Computer Simulations*, **7(4)**, 339–348, 2015.

[17] A.A. Samarskii, *The theory of difference schemes*. USA: Marcel Dekker, Inc., 2001.

[18] G.S. Khakimzyanov, N.Yu. Shokina, Some notes on monotonicity preserving schemes. *Computational Technologies*, **17(2)**, 78–98, 2012. (in Russ.)

[19] N.Yu. Shokina, To the problem of construction of difference schemes on movable grids. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **27(6)**, 603–626, 2012.

[20] S.P. Bautin, S.L. Deryabin, A.F. Sommer, G.S. Khakimzyanov, N.Yu. Shokina, Use of analytic solutions in the statement of difference boundary conditions on a movable shoreline. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **26(4)**, 353–377, 2011.

[21] C.E. Synolakis, The runup of solitary waves. *Journal of Fluid Mechanics*, **185**, 523–545, 1987.

[22] S.A. Beizel, N.Yu. Shokina, G.S. Khakimzyanov, L.B. Chubarov, O.A. Kovyrkina, V.V. Ostapenko, On some numerical algorithms for computation of tsunami runup in the framework of shallow water model. I. *Computational Technologies*, **19(1)**, 40–62, 2014. (in Russ.)

[23] O.I. Gusev, N.Yu. Shokina, V.A. Kutergin, G.S. Khakimzyanov, Numerical modelling of surface waves generated by underwater landslide in a reservoir. *Computational Technologies*, **18(5)**, 74–90, 2013. (in Russ.)

[24] Yu.I. Shokin, S.A. Beisel, O.I. Gusev, G.S. Khakimzyanov, L.B. Chubarov, N.Yu. Shokina, Numerical modelling of dispersive waves generated by landslide motion. *Bulletin of the South Ural State University*, **7(1)**, 121–133, 2014. (in Russ.)

[25] B.L. Rozhdestvenskiy, N.N. Yanenko, *Systems of quasilinear equations and their application to gas dynamics*. Moscow: Nauka, 1978. (in Russ.)

[26] G. Pedersen, B. Gjevik, Run-up of solitary waves. *Journal of Fluid Mechanics*, **135**, 283–299, 1983.

[27] M.J. Cooker, P.D. Weidman, D.S. Bale, Reflection of a high-amplitude solitary wave at a vertical wall. *Journal of Fluid Mechanics*, **342**, 141–158, 1997.

[28] C.E. Synolakis, E.N. Bernard, V.V. Titov, U. Kanoglu, F.I. Gonzalez, Validation and verification of tsunami numerical models. *Pure and Applied Geophysics*, **165**, 2197–2228, 2008.

[29] J. Horrillo, S.T. Grilli, D. Nicolsky, V. Roeber, J. Zhang, Performance benchmarking tsunami models for NTHMP's inundation mapping activities. *Pure and Applied Geophysics*, **170(3-4)**, 1333–1359, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON NUMERICAL METHODS FOR SOLVING RUN-UP PROBLEMS. COMPARATIVE ANALYSIS OF NUMERICAL ALGORITHMS AND NUMERICAL RESULTS

Leonid B. Chubarov[1,2], Alexandr D. Rychkov[1], Gayaz S. Khakimzyanov[1,2], and Yurii I. Shokin[1]

[1]Institute of Computational Technologies SB RAS
6, Lavrentiev Avenue, 630090, Novosibirsk, Russia
e-mail: (chubarov, rych, khak, shokin)@ict.nsc.ru

[2] Novosibirsk State University
2, Pirogova Str., 630090, Novosibirsk, Russia
e-mail: dec@mmf.nsu.ru

Keywords: Surface Waves, Shallow Water, Run-up, Numerical Simulation.

Abstract. *The numerical simulation of the run-up of long surface waves on a plane slope is presented. Using a method based on the combination of the TVD scheme and the SPH method the shallow water approximation is applied to the solution of the well known model problem of a run-up of a wave approaching from an area of constant depth towards a plane slope. The numerical method has proved to be reliable and effective not only in the range of small amplitudes, but also outside of the theoretical limits of applicability of the shallow water theory, such as for the modelling of breaking waves. The qualitative and partially quantitative comparison with the results of numerical calculations of other authors are presented. The differences in the results caused by the differences in the numerical algorithms are highlighted.*

1 INTRODUCTION

The interaction of waves with coasts and coastal installations is one of the most interesting and complex phenomena studied in the framework of the mathematical models of wave hydrodynamics. The investigation of the final phase of tsunami waves' existence – their run-up on coast and run-down – belongs to this class of problems. It is necessary to define the extreme values of run-up, distances of wave propagation to a coast, drying areas, inundation depths, and duration of land's being under water with necessary precision and well ahead of time.

The development of reliable algorithms for the solution of the considered class of problems requires not only the theoretical analysis of the mathematical models and the numerical methods used for their implementation, but also the validation on the test problems. One of the "canonical" problems of this type is the problem on the definition of the parameters of a long wave run-up on a plane slope with the given slope angle, where a wave runs up a slope from the side of the bottom part with a constant depth [1]. It is customary to assume that such problem reflects the main physical aspects of the considered phenomenon with a sufficient completeness.

The current paper is devoted to the application of our combined TVD+SPH method [2] to the solution of the above described "canonical" problem under the conditions, which reconstruct the run-ups of long waves with various configurations and amplitudes on the slopes with different slope angles. Our task is to identify the basic characteristics of the phenomena for weakly nonlinear (non-breaking) as well as for highly nonlinear (breaking) waves. The stability (robustness) of the algorithm has also to be estimated for a wide range of parameter variation.

The first part of the paper briefly describes the problem formulation; the second part presents the results of the numerical modelling of the wave processes, which were investigated experimentally in the Large Wave Flume of the University of Hannover [3]. The obtained results are compared with those of the numerical modelling by the research group of Prof. E.N. Pelinovsky [4].

2 PROBLEM FORMULATION AND NUMERICAL ALGORITHMS

One-dimensional problem of wave run-up on a plane slope adjacent to a horizontal bottom with a constant depth h_0 is considered in the framework of the shallow water model. The Cartesian coordinate system Oxz is used with the vertical axis Oz directed upward and with the coordinate line $z = 0$ coinciding with the unperturbed free surface of an ideal fluid layer. A fluid layer is bounded from below by a fixed bottom $z = -h\,(x)$, and from above – by a moving free boundary $z = \eta\,(x, t)$, where t is the time. In this case the system of shallow water equations is written in the form of the conservation laws:

$$\mathbf{q}_t + \mathbf{F}_x = \mathbf{G}, \qquad t > 0, \tag{1}$$

where

$$\mathbf{q} = \begin{pmatrix} H \\ Hu \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} Hu \\ Hu^2 + gH^2/2 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 0 \\ gHh_x \end{pmatrix},$$

g is the gravity acceleration. The sought values are the full depth of a fluid layer $H(x,t) = \eta(x,t) + h(x) \geq 0$ and $u(x,t)$ – its velocity, which is vertically averaged from bottom to free surface.

In [2] the channel was considered, where the plane slope was positioned to the left from the bottom part with a constant depth. However, in the present work the plane slope is placed to the right from the horizontal segment, making the comparison to the results of other authors more convinient. In this case the solution domain is limited from the left by a impermeable wall at the

point $x = 0$, and from the right – by a moving boundary $x_0(t)$, which separates water from land (waterfront point). A moving boundary is unknown, therefore, the number of sought functions increases.

The problem for the equation (1) is closed by the initial and boundary conditions. Under the assumptions that the initial position $x_0(0)$ of the waterfront point is known and the wave moves from left to right, the boundary conditions take the following form:

$$H(x_0(t), t) = 0, \quad u(0, t) = 0, \quad t \geq 0; \tag{2}$$

$$\eta(x, 0) = \eta_0(x), \quad u(x, 0) = u_0(x), \quad 0 \leq x \leq x_0(0). \tag{3}$$

The relief of bottom and adjacent land is given by the function

$$z = -h(x) = \begin{cases} -h_0, & 0 \leq x \leq x_s, \\ -h_0 + (x - x_s)\tan\ \beta, & x_s \leq x \leq L_x, \end{cases} \tag{4}$$

where $\beta > 0$ is the plane slope angle, $x_s > 0$ is the given abscissa of the transition point between the inclined and the horizontal bottom segments, $L_x = x_s + (z_0 + h_0)\cot\beta$, $z_0 > 0$ is the land height at the point $x = L_x$. The value z_0 is chosen so that the maximal vertical run-up of a wave on a coast is less than z_0. Therefore, the inequality $x_0(t) < L_x$ is satisfied for all $t > 0$.

In our investigations presented in this paper, we have used two numerical methods of the second accuracy order for modelling of wave run-up on a coast. The first method is based on the approach suggested in [5] and represents an original combination of the SPH (smoothed-particle hydrodynamics) method and the finite-difference scheme with TVD properties (named below as the TVD+SPH method). We supplemented this method with the algorithm for computing a moving waterfront point. The comparative analysis of the methods for modelling of long surface wave run-up in the framework of the shallow water theory [2] has shown that the TVD+SPH method appears to be the most advanced for this class of problems.

The second method, conventionally named as "exact", is presented in [6, 7]. It uses a moving grid, and the analytical solution of the problem in a small vicinity of a waterfront point is applied to compute its position and velocity.

3 NUMERICAL RESULTS

Let us consider some results of the numerical solution of the test problem, where the characteristics of run-ups of positive and negative pulses (positive and negative polarity correspondingly) on a slope are defined for a wide range of initial wave amplitudes. We compare our numerical results with the results of other authors. In particular, qualitative and quantitative comparison is done with the results of the research group of Prof. E.N. Pelinovsky [4, 8, 9], which include the analytical solutions (for non-breaking waves) and the data of laboratory and numerical investigations. The experiments were performed in the Large Wave Flume of the University of Hannover, using a wave flume consisting of a segment with a constant depth $h_0 = 3.5$ m and length $x_s = 250$ m, adjacent to a plane slope $1 : 6$ which was positioned near the right boundary of the flume [3]. The numerical computation in [4] were done using the software package CLAWPACK [10].

The initial data were given by the relations

$$\eta(x, 0) = A\cosh^{-2}\Big[(x - x_w)/L\Big], \quad u(x, 0) = 2\left[\sqrt{g\left(h_0 + \eta(x, 0)\right)} - \sqrt{gh_0}\right]. \tag{5}$$

As it is shown in [11, 12], these relations correspond to the exact solution (Riemann wave) of the shallow water equations

$$H(x,t) = H_0\left[x - V(H)\,t\right], \quad V(H) = 3\sqrt{gH} - 2\sqrt{gh_0}, \quad H_0(x) = h_0 + \eta(x,0).$$

The authors of [4] interpret the value L as the half of the length of the initial elevation and take $L = 11$ m in their computations. The extremal values of surface displacements (wave amplitudes) were initially positioned near the point $x_w = 50$ m, and they varied in the range from 0.05 to 3.5 m for the waves with positive polarity, and in the range from -0.05 m to -3.49 m – for the waves with negative polarity.

In comparison to [4], we have additionally modelled the waves with smaller initial amplitudes, namely 0.001 m, 0.0025 m, 0.005 m, 0.01 m. Let us note that the applicability limits for the shallow water theory, which are defined by the possibility of using the analytical formula of Synolakis, are bounded in our computations by the interval of initial amplitude variation $0.008064 \leq A < 0.228976576$. However, the results for the amplitudes which fall outside the limits of the given interval, can be used for testing new algorithms for wave run-up modelling in the framework of the shallow water model. At the least, these results explicitly point out that the used numerical methods can work even outside the theoretical limits of applicability of the approximated mathematical model.

We have chosen the display format for the results according to the examples provided in [4]. Unfortunately, some peculiarities of the formulations of numerical modelling problems [4] do not allow direct quantitative comparison of the results, but qualitative comparison is possible.

The first series of graphs (Fig. 1) shows the profiles of free surface computed for the same set of amplitudes as in [4]. Each graph presents the waves at the initial time moment (lines (*1*)); near the transition point between the inclined and the horizontal bottom segments (lines (*2*), which correspond to the time moment $t = 30$ s for the amplitudes $A = 0.1$ m and $A = 0.5$ m, and $t = 20$ s – for $A = 1.5$ m and $A = 3.5$ m); at the moment of nearly vertical run-up (lines (*3*), which correspond to the time moment $t = 40$ s for the amplitudes $A = 0.1$ m and $A = 0.5$ m, and $t = 30$ s – for $A = 1.5$ m and $A = 3.5$ m); and at the moment of approach of the wave, reflected by the slope, to the opposite boundary of the computational domain (lines (*4*), which correspond to the time moment $t = 70$ s for the amplitudes $A = 0.1$ m and $A = 0.5$ m, and $t = 60$ s – for $A = 1.5$ m and $A = 3.5$ m). Let us note that for small amplitudes our results are not only in qualitative agreement with the results from [4], but in quantitative agreement as well; whereas the differences in free surface profiles, which evolve after the interaction of waves with slope, significantly increase with the increase of initial amplitudes.

The influence of the non-linearity can be already seen for the wave with the smallest initial amplitude. The leading edge of the wave gradually becomes steeper, but wave breaking (gradient catastrophe) does not happen (Fig. 1 (*a*)), the amplitude of the wave, which propagates above the horizontal bottom segment, does not essentially decline. After the reflection from the slope the wave form changes significantly and in fact changes into the N-wave, which contains the fragment of depression lower than zero level.

With the increase of an initial wave amplitude the influence of non-linear effects on wave characteristics is also enhanced. Under the impact on non-linearity the leading edge of waves become significantly steeper, whereas trailing edges flatten out so that the wave lengths increase and the amplitudes decline. As a result the already "breaking" waves approach a plane slope and run up to it. These considerations are proved by the Figure 2, which presents the characteristics of the computed incident and reflected waves above the flume segment with a constant

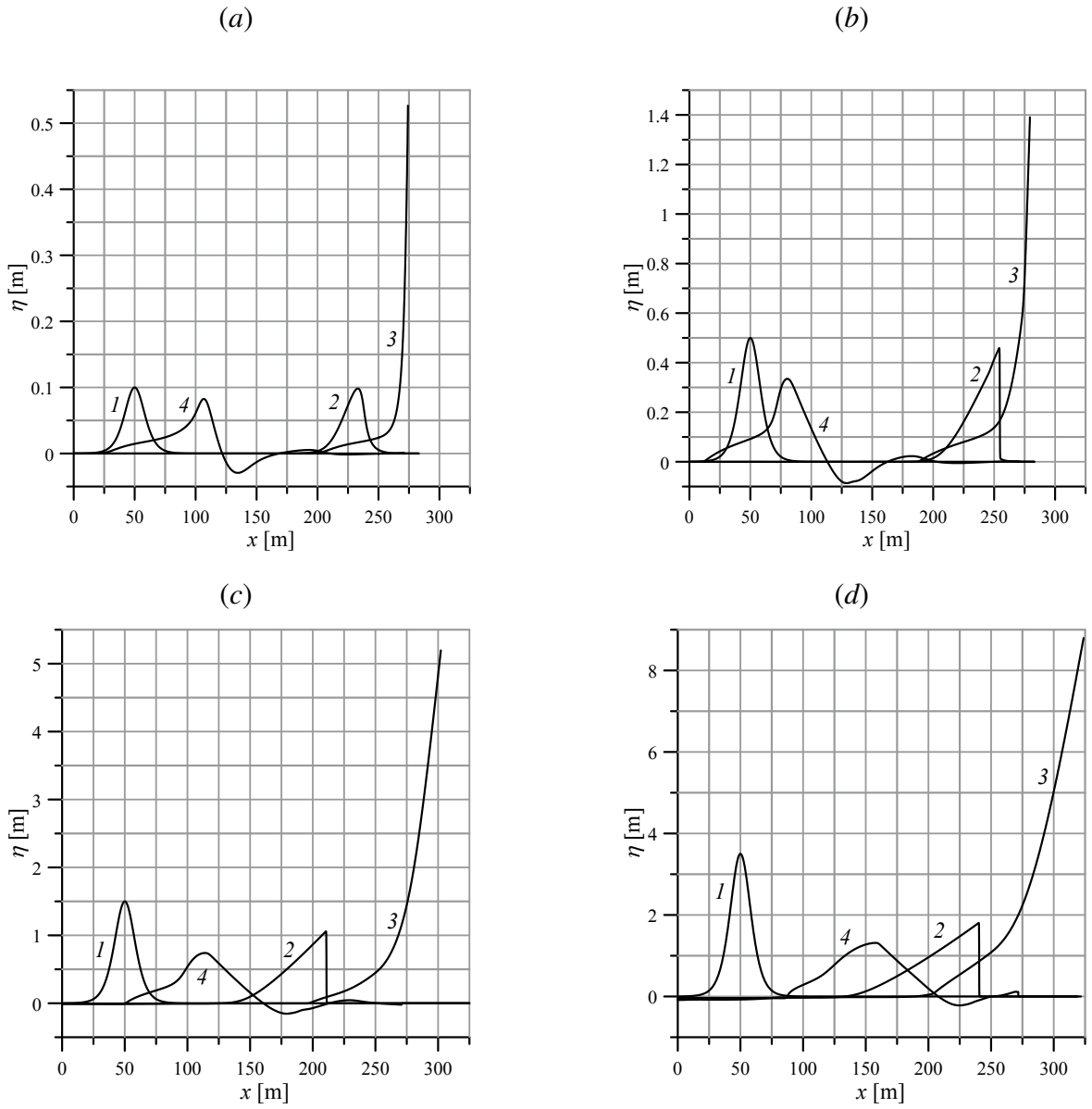


Figure 1: The free surface profiles for run-up on a plane slope for the waves with positive polarity and the amplitudes $A = 0.1$ m (*a*), 0.5 m (*b*), 1.5 m (*c*), 3.5 m (*d*) at the initial time moment, near the transition point between the inclined and the horizontal bottom segments, the moment of nearly vertical run-up, the moment of approach of the wave, reflected by the slope, to the opposite boundary of the computational domain. The values of time moments are given in the text

depth (Fig. 2 (*a*)) and above the transition point between the inclined and the horizontal bottom segments (Fig. 2 (*b*)). Fig. 2 (*a*) shows that incident and reflected waves are time-spaced, the incident waves have vertical leading edges, whereas the reflected waves have a smooth form. Above the transition point between the inclined and the horizontal bottom segments the waves are not time-spaced, here an incident wave meets a reflected wave before completely passing this point $x = x_s$.

The comparison of the results from the Fig. 3 (*a*) with the results from [4] demonstrates not only quantitative but also significant qualitative differences. The run-up heights computed with the TVD+SPH method are much larger, whereas the duration of run-up is much shorter. Besides, in contrast to the results of [4], the dependencies of maximal (in absolute magnitude) run-

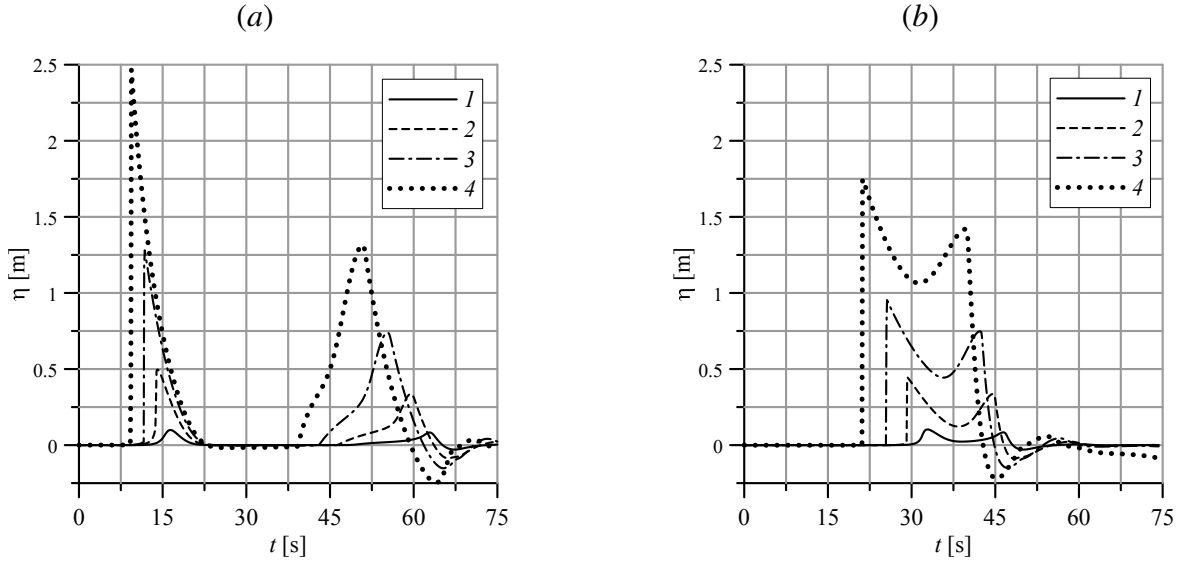


Figure 2: The mareograms, computed above the segment with constant depth of the virtual wave flume at the point $x = 150$ m (*a*), the transition point $x = 250$ m between the inclined and the horizontal bottom segments (*b*), for the run up of the waves with initial amplitudes $A = 0.1$ m (*1*), 0.5 m (*2*), 1.5 m (*3*), 3.5 m (*4*)

downs on initial amplitudes are monotone, and the dependencies of vertical run-ups are smooth for all amplitudes. Finally, in [4] the absolute values of run-downs change non-monotonically and decrease with tending to zero (with that being in poor agreement with the ideas of the physics of modelled processes) for the largest initial amplitudes starting from $A = 2$ m. In the results presented in 3 (*a*) these values increase slowly, but monotonically. There are also differences in velocities: for the same conditions the run-up and run-down processes happen significantly slower in [4].

Some of the data obtained by the solution of the considered problem is presented in Table 1.

The next results were obtained for the modelling of run-up of the waves with negative polarity ($A < 0$). The initial positions coincided with those of the waves with positive polarity. The

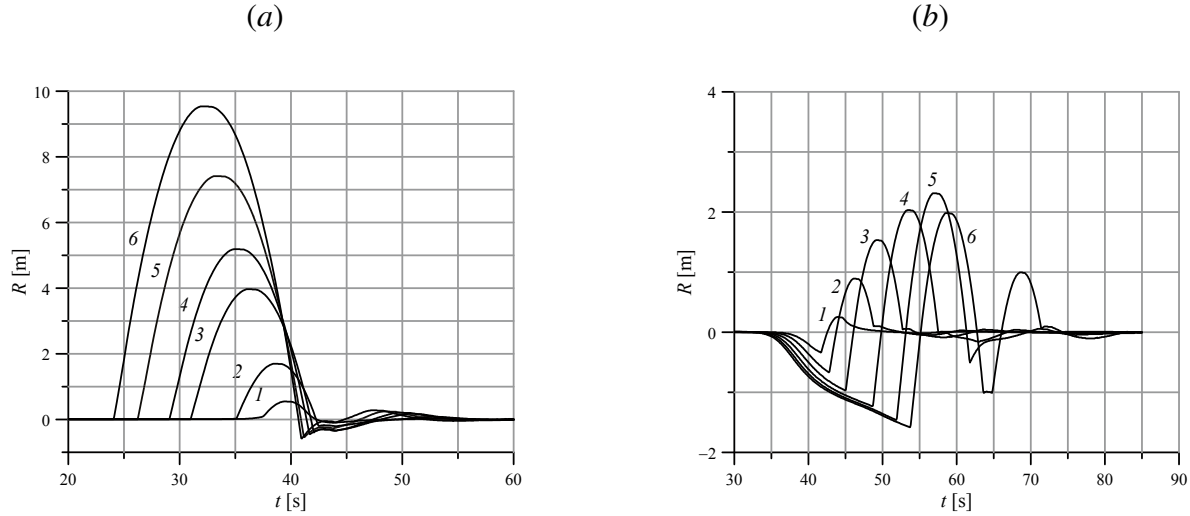


Figure 3: The dynamics of the vertical run-up on a plane slope of the waves with positive (*a*, $A > 0$) and negative (*b*, $A < 0$) polarity with the intial amplitudes $A = \pm 0.1$ m (*1*), ± 0.3 m (*2*), ± 0.7 m (*3*), ± 1.5 m (*4*), ± 2.5 m (*5*), $A = 3.5$ m (*a*) and $A = -3.49$ m (*b*) – (*6*)

	Ratio of maximal run-up to initial amplitude		Ratio of maximal run-down to initial amplitude		Ratio of maximal run-down to maximal run-up	
A, m	TVD+SPH	[4]	TVD+SPH	[4]	TVD+SPH	[4]
0.05	4.868	4.7354	1.104	1.0053	0.2268	0.2365
0.1	5.499	4.6779	1.024	1.0015	0.1862	0.2211
0.3	5.658	3.5827	0.682	0.7363	0.1205	0.2055
0.5	5.0096	2.9408	0.5778	0.5733	0.1153	0.1962
0.7	4.503	2.5674	0.4829	0.4672	0.1072	0.1820
1	3.9676	2.2246	0.396	0.3653	0.0998	0.1642
1.5	3.4581	1.9056	0.2983	0.1877	0.0863	0.0983
2	3.1606	1.7105	0.2547	0.0899	0.0806	0.0526
2.5	2.9644	1.5826	0.2144	0.0229	0.0723	0.0144
3	2.8293	1.4976	0.189	0.0065	0.0668	0.0044
3.5	2.7244	1.3807	0.1663	0.0009	0.0611	0.0007

Table 1: Main characteristics of the run-ups of the waves with positive polarity on a plane slope, computed using two numerical algorithms

initial form of the waves coincided as well with an accuracy up to the plane reflection. As in the case of the positive pulses, the results obtained by the TVD+SPH method are close to the results from [4] for small amplitudes, but significantly differ for $|A| > 1$ m. The primary analysis of the computed profiles shows that the for large amplitudes the times of arrivals of the extremal values to the same points do not coincide. Thus, the profiles in Fig. 4 correspond to the time moments other than those in [4]. Each graph shows the waves at the initial time moment (lines (*1*)); near the transition point between the inclined and the horizontal bottom segments (lines (*2*), which correspond to the time moment $t = 20$ s); at the moment of nearly vertical run-up (lines (*3*), which correspond to the time moment $t = 41$ s for the amplitude $A = -0.1$ m, $t = 43$ s – for $A = -0.5$ m, $t = 46.5$ s – for $A = -1$ m, $t = 53.86$ s – for $A = -3.49$ m); and at the moment of approach of the wave, reflected by the slope, to the opposite boundary of the computational domain (lines (*4*), which correspond to the time moment $t = 75$ s for the amplitudes $A = -0.1$ m, -0.5 m and -1 m, and $t = 85$ s – for $A = -3.49$ m).

Thus, for $A = 1.0$ m some advance is seen in the positions of run-up and run-down profiles in comparison to the results of the work [4], where as for $A = -3.49$ m – the delay is present. The numerical experiment with such amplitude of the wave with negative polarity can be hardly considered to be reasonable because of the obvious inapplicability of the shallow water theory. However, the possibility of performing the computations of this kind testify to the high working efficiency of the algorithm. The corresponding fragment of the figure (Fig. 4, *d*) demonstrates the significant differences for all time moments with the meaningfully close figure from the work [4], which demonstrate themselves in much more complicated forms of incident and reflected waves.

The general idea about the waterfront dynamics during the run-up of waves with negative polarity on a plane slope can be given by the results (Fig. 3 (*b*)), obtained by the TVD+SPH on the grid with 6000 nodes. However, these results appeared to be practically identical to those obtained on the grid with 4000. The graphs on this figure demonstrate the monotonic increase of the values of both primary and secondary run-downs, yet the value of run-up increases up to the initial amplitude -2.5 m and then decreases for the amplitudes $A = -3$ m and $A = -3.499$ m.

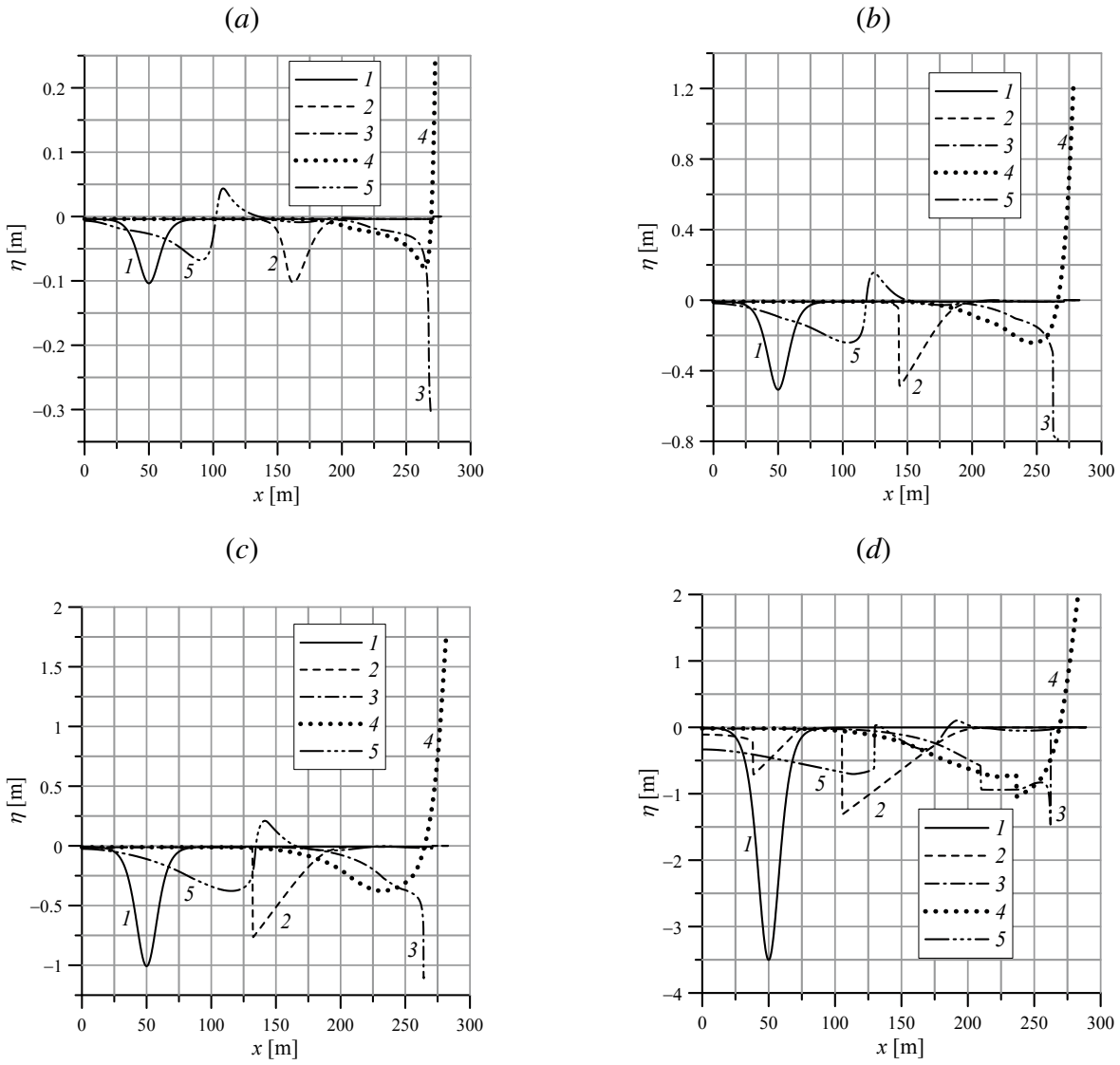


Figure 4: The free surface profiles for the run-up on a plane slope of the waves with negative polarity with initial amplitudes $A = -0.1$ m (*a*), -0.5 m (*b*), -1 m (*c*), -3.49 m (*d*) at the initial time moment, and at the moment of approach of the wave to the transition point between the inclined and the horizontal bottom segments, at the moment of nearly vertical run-up, and at the moment of approach of the wave, reflected by the slope, to the opposite boundary of the computational domain. The values of time moments are given in the text

Further we compare our results with those of the work [4]. Our results were obtained by the TVD+SPH method and the algorithm, based of the definition of the waterfront point using the analytical solutions of the shallow water equations for the formulation of the difference boundary conditions on a moving waterfront [6]. The values from [4] were obtained partially from its text and partially by the digitization of the graphs presented there.

The first graphs (Fig. 5) demonstrate the dependency of the maximal values of R/A on the initial amplitude in the phase of run-up of the waves with positive polarity. These values, obtained using the TVD+SPH method, are significantly larger than the results from the work [4] for practically all considered amplitudes. Yet, they are close to the results obtained using the methodology from the work [6], which, however, happen to be somewhat smaller for the largest amplitudes. Let us note, that if for the absolute values of the run-up values R the monotonic increase takes place, then the corresponding relative values R/A increase first (up to the ampli-

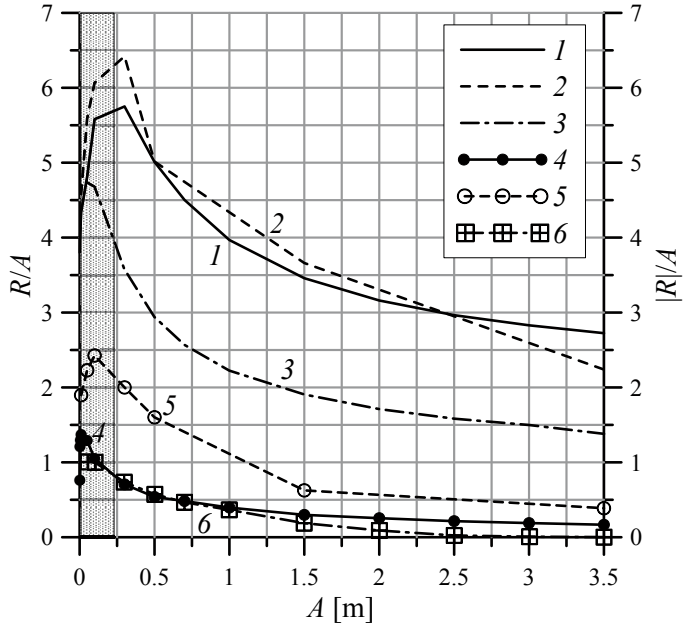


Figure 5: Relative characteristics of the run-ups of the wave with positive polarity with respect to their initial amplitudes: maximal values of R/A during run-up phases (*1–3*) and $|R|/A$ – during run-down phases (*4–6*), computed with the TVD+SPH method (*1*, *4*), the methods from [6] (*2*, *5*) and [4] (*3*, *6*); the domain of applicability of the analytical solution [13] is marked by grey color.

tude 0.25 m), and only after that they begin to decrease monotonically. This can be explained by the strengthening of the non-linear effects' influence, which lead to the increase of the leading edge steepness, decrease of the heights and increase of lengths for the waves approaching a slope. In the results [4] the extremum is not found because of the absence of the data for small amplitudes. The significantly smaller values from [4] can be explained by the fact that the implicit algorithm is implemented in the software package CLAWPACK used by the authors. The absolute stability of the algorithm allows computing with a rather large time step, leading to the significant growth of the numerical dissipation and a subsequent decrease of the results quality.

Considering the characteristics in the run-down phases, some changes can be noted. For small amplitudes (nearly up to $A = 1$ m) the values, obtained by the TVD+SPH method, increase monotonically and practically coincides with the results from [4], which starts to decrease sharply and tend to zero for the largest amplitude. On the contrary, the results, obtained by the TVD+SPH algorithm, continue the stable monotonic increase and stay smaller than the results from [6] for the whole range of amplitude variation. The analysis of the graphs with the maximal values $|R|/A$ in the run-down phase points in the first place to the preserving anomalous tendency of the results from [4] to zero for $A > 1$ m and their closeness to the results, obtained by the TVD+SPH algorithm, up to the amplitude $A = 1$ m. The behavior of the TVD+SPH results are qualitatively close to the distribution computed by the methodology from [6], which however is positioned significantly higher. The results obtained for the smaller initial amplitudes, as in the run-up phase, demonstrate the presence of the extremum, which is slightly shifted to the direction of the smallest values of A. Unfortunately, the results, computed in the discussed range of the argument variation and needed for the comparison, are absent on the work [4].

The relations of the maximal values of $|R|$ in the run-down phase to the maximal values of R in the run-up phases decrease with the increase of initial amplitude of the wave approaching a slope. As before, the results from [4] tend to zero; the results from [6] have the largest values; and the results obtained by the TVD+SPH algorithm demonstrate the presence of extremum in the domain of the very small amplitudes.

The analogues of the above considered characteristics of the run-up of waves with negative polarity show that, as in the case of the positive pulses, the results obtained by the TVD+SPH algorithm are larger than the results from [4]. This is expressed to a greater degree in the run-up characteristics, whereas the corresponding run-down characteristics appear to be very close. Both sets of relative values for the run-up phase (Fig. 6) possess the maximums in the range of small initial amplitudes. After reaching these maximums the corresponding distributions decrease monotonically. The relative characteristics of the run-down phase do not have extremes and decrease monotonically with the increase of initial amplitude. Let us note some peculiarity of the value R variation, computed using the TVD+SPH algorithm. This peculiarity consists in the fact that near the initial amplitude $A = 2.5$ m the monotonous increase of the maximal vertical position of the moving waterfront point is changed to equally monotonous decrease. We can assume that this is caused by the specificity of the non-linear effects' simulation by the TVD+SPH algorithm, which leads to an early breaking of the waves with large amplitudes, approaching a slope, with the simultaneous decrease of their positive fragments' heights. This conclusion finds its proof on the profiles of the corresponding free surfaces (Fig. 4) as well as on the graphs of the dynamics of the moving waterfront point (Fig. 3 (*b*)). The distributions of the relations of the maximal values of $|R|$ in the run-down phase to the maximal values of $|R|$ in the run-up phase are essentially close qualitatively. At first, they decrease rather steeply and take the minimal value for the initial amplitude $A = -1$ m, and then they increase slowly. Here the distributions from the work [4] appear to be higher.

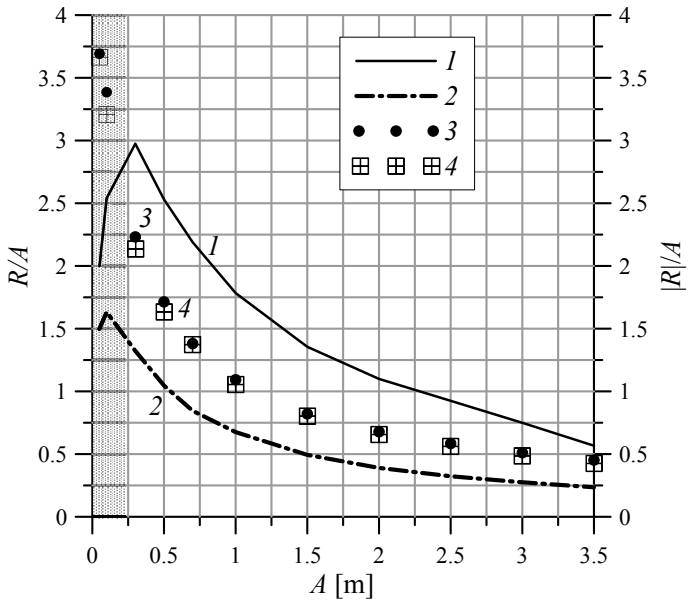


Figure 6: The dependence of the relative characteristics of the run-ups of the waves with negative polarity on their initial amplitudes: maximal values of R/A in the run-up phases (*1*, *2*) and $|R|/A$ – in the run-down phases (*3*, *4*), computed with the TVD+SPH method (*1*, *3*) and [4] (*2*, *4*); the domain of applicability of the analytical solution [13] is marked by grey color

The analysis of the peculiarities of the run-ups of waves with different polarities (positive and negative pulses) has shown that the waves with positive polarity lead to the larger values of the vertical run-down R. However, the velocities of the waterfront point for the run-up of waves with negative polarity are significantly larger. This leads to the steeper trailing edges of waves, being in the qualitative agreement with the results of the work [8]. Let us also note that for the run-up of waves with positive polarity the vertical displacements of the waterfront points in the run-up phase exceed significantly the displacements in the run-down phase, whereas for the waves with negative polarity these values appear to be much closer. The non-monotonicity of free surface profiles above the breakpoint of the bottom relief (Fig. 2), which is absent in the results from [8], is the consequence of the interaction of the approaching wave and the wave reflected from the slope.

Summarizing the comparison with the results from the work [4], which are connected to some extent with the conditions of the physical experiments in the wave flume in the University of Hannover, let us note that even the application of the well-known computer code, created by highly professional specialists, requires the detailed investigation of the underlying algorithm, understanding of its peculiarities, and correct parameter settings. Otherwise it can happen that the numerical effects, which are intrinsic for the considered algorithm, can corrupt significantly the values of the computed variables, complicate the adequate interpretation of the obtained results, and lead to the physically wrong conclusions.

Let us also note that the explicit algorithm, which is used in the TVD+SPH method, guarantees the necessary computational accuracy and proves the correctness of the choice of numerical schemes for the modelling of fine effects arising in wave run-ups even on simple model slopes.

4 CONCLUSIONS

The presented results have demonstrated the possibilities of the combined TVD+SPH numerical method in the definition of the main characteristics of run-ups of the waves with various configurations in the “canonical” problem on long wave run-up on a plane slope. This method has proved to be functional and effective not only in the range of small amplitudes where the shallow water approximation os theoretically allowed, but also beyond its theoretical applicability in the case of breaking waves.

The analysis of the run-ups of solitary waves with various polarity (positive and negative pulses) has shown that the positive pulses have lead to the larger values of vertical run-up in comparison to the negative pulses. This conclusion is in the qualitative agreement with the results of the work [4]. It was also shown that the interaction of the incident and reflected waves above the breakpoint of the bottom relief has lead to the non-monotonicity of free surface profiles, which has influenced the run-up characteristics.

We consider the development and careful analysis of the version of the TVD+SPH algorithm for the modelling of run-ups of catastrophic waves of various configurations in the spatial formulations and with taking into account the real properties of water areas and adjacent coasts to be a promising research direction.

Funding: The study was supported by the Russian Scientific Foundation (project No. 14-17-00219).

REFERENCES

[1] S. Tadepalli, C.E. Synolakis, The Run-up of N-waves on sloping beaches. *Proceedings of Royal Society of London. A*, **445**, 99–112, 1994.

[2] Yu.I. Shokin, A.D. Rychkov, G.S. Khakimzyanov, L.B. Chubarov, On numerical methods for solving run-up problems. I. Comparative analysis of numerical algorithms for one-dimensional problems. *Computational Technologies*, **20** (5), 214–232, 2015 (In Russian).

[3] P. Denissenko, I. Didenkulova, A. Rodin, M. Listak, E. Pelinovsky, Experimental statistics of long wave runup on a plane beach. *Journal of Coastal Research*, **65** (Special issue), 195–200, 2013.

[4] A. Rodin, I. Didenkulova, E. Pelinovsky, Numerical study for run-up of breaking waves of different polarities on a sloping beach. E. Pelinovsky, C. Kharif eds. *Extreme Ocean Waves*, Springer International Publishing, 155–172, 2016.

[5] S.S. Hrapov, A.V. Khoperskov, N.M. Kuzmin, A.V. Pisarev, I.A. Kobelev, Numerical scheme for simulating the dynamics of surface water on the basis of the combined SPH-TVD approach. *Computational Methods and Programming Techniques*, **12**, 282–297, 2011 (In Russian).

[6] S.P. Bautin, S.L. Deryabin, A.F. Sommer, G.S. Khakimzyanov, N.Yu. Shokina, Use of analytic solutions in the statement of difference boundary conditions on a movable shoreline. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **26** (4), 353–377, 2011.

[7] G. Khakimzyanov, N.Yu. Shokina, D. Dutykh, D. Mitsotakis, A new run-up algorithm based on local high-order analytic expansions. *Journal of Computational and Applied Mathematics*, **298**, 82–96, 2016.

[8] I.I. Didenkulova, E.N. Pelinovsky, O.I. Didenkulov, Run-up of long solitary waves of different polarities on a plane beach. *Izvestiya, Atmospheric and Oceanic Physics*, **50** (5), 532–538, 2014.

[9] A.A. Rodin, Numerical calculations of run-up of solitary breaking waves on a plane beach. *Transactions of Nizhni Novgorod State Technical University named by R.E. Alekseev*, **1** (98), 36–43, 2013 (In Russian).

[10] Clawpack-5 home page // http://www.clawpack.org/

[11] E.N. Pelinovsky, A.A. Rodin, Nonlinear deformation of a large-amplitude wave on shallow water. *Doklady Physics*, **56** (5), 305–308, 2011.

[12] E.N. Pelinovsky, A.A. Rodin, Transformation of a strongly nonlinear wave in a shallow-water basin. *Izvestiya, Atmospheric and Oceanic Physics*, **48** (3), 383–390, 2012.

[13] C.E. Synolakis, The runup of solitary waves. *Journal of Fluid Mechanics*, **185** (6), 523–545, 1987.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EULERIAN CALCULATIONS OF MULTIMATERIAL FLOWS WITH SUB-CELL RECONSTRUCTION OF INTERFACES

Igor S. Menshov[1,2] **and Pavel P. Zakharov**[2]

[1] Keldysh Institute for Applied Mathematics RAS
Miusskaya 4, 125047 Moscow Russia
e-mail: menshov@kiam.ru

[2] VNIIA ROSATOM
Sushchevskaya 22, 127055 Moscow Russia
jaquecostou@mail.ru

Keywords: multi-material fluid flow, interface capturing, composite Riemann problem, Godunov method

Abstract. *In the present paper we consider an Eulerian numerical method for calculating multi-material fluid. This method relates to the class of interface capturing methods that resolve material interfaces not exactly, but with some numerical smearing in space. To suppress this numerical smearing, we suggest a numerical flux approximation based on the sub-cell material interface reconstruction by simple patterns and the solution of the composite Riemann problem (CRP). It is found that the improved scheme captures the interface within one computational cell in 1D calculations. For multi-dimensional calculations we implement an approach of directional splitting that sufficiently decreases smearing of material interfaces what is demonstrated in computations of various multi-material applications.*

1 INTRODUCTION

The present paper is devoted to an Eulerian finite-volume method for numerical calculations of multimaterial fluid flows. We consider a heterogeneous medium that consists of different materials (components) separated with interfaces. Each component is described by its own EOS and governed by the system of compressible Euler equations. The mathematical model is applied that treats this multimaterial fluid flow as the flow of a heterogeneous multiphase fluid under the assumption of equilibrium in velocity, pressure, and temperature.

The method we develop belongs to the class of interface capturing methods. The interface is represented by the distribution in space of component mass fractions. The conventional Godunov method is applied for numerical flux approximation with exact or approximate Riemann solvers. This approach inevitably diffuses the interfaces between different materials because of the numerical viscosity that smears initially sharp distribution of mass fractions. The interface is given by a set of mixed cells occupied by the heterogeneous mixture of the components. As the calculation proceeds the domain of mixture cells expands involving more and more computational cells so that eventually the position of the interface in space is completely lost. This is the basic drawback of the conventional Godunov method.

Another drawback appears when we try to compute flows of condensed and gaseous materials with rarefaction waves. The EOSs in this case are quite different and admit different regions of admissible values of thermodynamic parameters. Solids admit negative pressure while the gaseous phase not. Because of the equilibrium assumption the pressure in mixed cells should be always positive. This leads to an artificial numerical effect that we name as pseudo-fracture. It shows up in the formation of a small region of mixed cells near the pure solid region where the solid volume fraction drastically decreases [1].

In order to fix these drawbacks one can sharpen smearing interfaces by implementing some techniques commonly referred to as anti-diffusive [2-3]. We suggest an alternative way that is based on a numerical flux approximation that takes into account the interface local position or in other words the material sub-cell structure in mixed cells. Following this way we come to generalized formulation of the Riemann problem for the medium that contains two materials separated with the interface. This formulation includes an initial discontinuity in one component and also the contact point of the interface [4]. We refer this problem as Composite Riemann Problem (CRP). The solution of the CRP provides more accurate approximation of the numerical flux in mixed cells. Various numerical results demonstrate that the CRP technique allows us to reduce the mixed zone to only one computational cell in 1D calculations and to a few cells in multidimensional calculations.

2 MODELING MULTI-MATERIAL FLOWS

We consider a heterogeneous multi-material compressible medium that in general consists of N components (materials). We assume that each component is thermodynamically a two-parameter medium and is governed by mechanical and thermal EOSs in the form of functional relations between density ρ_i^o, pressure p_i, specific internal energy e_i, and temperature T_i as $e_i = e_i(\rho_i^o, p_i)$ and $T_i = T_i(\rho_i^o, p_i)$, respectively. Here, $i = 1, \ldots, N$ indicates the index of the component. The mathematical model of the multi-phase medium with velocity, pressure, and temperature equilibrium, so called the PT model,

$$\mathbf{v} = \mathbf{v}_1 = \ldots = \mathbf{v}_N,\ p = p_1 = \ldots = p_N,\ T = T_1 = \ldots = T_N, \tag{1}$$

with the assumption of additivity of internal energies is used to describe the multi-material fluid flow [18]. The governing equations of this model can be written in the conservative form as follows:

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial \mathbf{f}_1}{\partial x} + \frac{\partial \mathbf{f}_2}{\partial y} + \frac{\partial \mathbf{f}_3}{\partial z} = 0\,, \tag{2}$$

where $\mathbf{q}$ is the vector of conservative variables, and $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ are the flux vectors along the x-axis, y-axis, and z-axis, respectively:

$$\mathbf{q} = \begin{pmatrix} \rho \\ \rho v_1 \\ \rho v_2 \\ \rho v_3 \\ \rho E \\ \rho\beta_1 \\ \dots \\ \rho\beta_N \end{pmatrix},\ \mathbf{f}_1 = \begin{pmatrix} \rho v_1 \\ \rho v_1 v_1 + p \\ \rho v_1 v_2 \\ \rho v_1 v_3 \\ v_1(\rho E + p) \\ \rho v_1 \beta_1 \\ \dots \\ \rho v_1 \beta_N \end{pmatrix},\ \mathbf{f}_2 = \begin{pmatrix} \rho v_2 \\ \rho v_2 v_1 \\ \rho v_2 v_2 + p \\ \rho v_2 v_3 \\ v_2(\rho E + p) \\ \rho v_2 \beta_1 \\ \dots \\ \rho v_2 \beta_N \end{pmatrix},\ \mathbf{f}_1 = \begin{pmatrix} \rho v_3 \\ \rho v_3 v_1 \\ \rho v_3 v_2 \\ \rho v_3 v_3 + p \\ v_3(\rho E + p) \\ \rho v_3 \beta_1 \\ \dots \\ \rho v_3 \beta_N \end{pmatrix}. \tag{3}$$

Here, ρ is the mixture density, $\rho = \sum_{i=1}^{N} \alpha_i \rho_i^o$, $\beta_i = \alpha_i \rho_i^o / \rho$; α_i, β_i are the volume and the mass fractions of i-th component, respectively, $\mathbf{v} = (v_1, v_2, v_3)^{\mathrm{T}}$ is the velocity vector in Cartesian coordinates, $E = e + 0.5\mathbf{v}^2$ is the mixture specific total energy, $e = \sum_{i=1}^{N} \beta_i e_i$ is the mixture specific internal energy. Along with the vector of conservative variables $\mathbf{q}$, we shall also consider the vector of primitive variables $\mathbf{z} = (\rho, v_1, v_2, v_3, p, \beta_1, \dots, \beta_N)^{\mathrm{T}}$.

The system of equations (2) and (3) is closed by mixture EOSs that are derived from the assumption of pressure and temperature equilibrium:

$$e = \sum_{i=1}^{N} \beta_i e_i(\rho_i^o, p_i) = \sum_{i=1}^{N} \beta_i e_i(p_i, T_i) = e(p, T, \beta_1, \dots, \beta_N)\,, \tag{4}$$

$$1 = \sum_{i=1}^{N} \alpha_i = \rho \sum_{i=1}^{N} \frac{\beta_i}{\rho_i^o(p_i, T_i)} = \rho \sum_{i=1}^{N} \frac{\beta_i}{\rho_i^o(p, T)}\,. \tag{5}$$

The multi-material fluid flow model can be considered as the multiphase model with a particular distribution of mass fractions given by

$$\beta_i(\mathbf{r}) = 1,\ \beta_j(\mathbf{r}) = 0,\ \forall j \neq i,\ \mathbf{r} = (x, y, z) \in \Omega_i(t),\ i = 1, \dots, N. \tag{6}$$

If the number of materials N is equal to one, Equations (2) and (3) reduce to the standard Euler equations for a single-material fluid flow.

If the mixture is characterized at initial ($t = 0$) by a distribution of type (6), then the same type of distribution is theoretically kept for $t > 0$, and two or more different components are never at the same point of space. However, this is not the case in numerical simulations.

When an Eulerian approach is employed, the captured material interface is found to be a region of mix cells, separating sub-domains of a pure fluid flows. The width of this region depends on the numerical scheme to be used and can span from a few to dozens of grid cells. A challenging problem in this approach is to reduce the width of numerical interface smearing as much as possible. As a solution to this problem, one can apply high-order numerical schemes or so-called anti-diffusive techniques. We suggest an alternative method to cope with numerical smearing of material interfaces, which is based on a special procedure of resolving materials at the sub-cell level.

3 THE DISCRETE MODEL

The spatial discretization of Equations (2) and (3) is performed with the finite volume method. With the explicit time marching scheme, this results in the following system of discrete equations:

$$\mathbf{q}_i^{n+1} = \mathbf{q}_i^n - \frac{\Delta t}{V_i} \sum_\sigma T_\sigma^{-1} \mathbf{F}_\sigma S_\sigma \,. \tag{7}$$

where the subscripts i and σ denote the cell and its face, respectively, V_i is the volume of the i-th cell, S_σ is the face area, $\mathbf{T}$ is the transforming matrix from the absolute system of coordinates to a coordinate system associated with the face. The vector $\mathbf{F} = \mathbf{F}(\mathbf{Q})$, $\mathbf{Q} = \mathbf{Tq}$, has the form of a locally 1D flux in the direction of the outward face normal

$$\mathbf{F} = \mathbf{F}(\mathbf{Q}) = \left(\rho v_n, \rho v_n^{\,2} + p, \rho v_n v_k, \rho v_n v_l, \rho v_n E + p v_n, \rho v_n \beta_1, \ \ldots\ , \rho v_n \beta_N\right)^T , \tag{8}$$

where the subscripts n, k, and l indicate normal and tangential components of the velocity vector.

In the conventional Godunov method the numerical flux $\mathbf{F}_\sigma$ is calculated as follows:

$$\mathbf{F}_\sigma = \mathbf{F}\left(\mathbf{Q}^{RP}\left(\mathbf{Z}_i^\sigma, \mathbf{Z}_{\sigma(i)}^\sigma\right)\right) = \mathbf{F}\left(\mathbf{Q}^{RP}\left(\mathbf{Tz}_i^\sigma, \mathbf{Tz}_{\sigma(i)}^\sigma\right)\right) \tag{9}$$

where $\mathbf{Q}^{RP}$ denotes the exact solution to the Riemann problem at the line $x = 0$ with initial data $\mathbf{Z}_i^\sigma$ and $\mathbf{Z}_{\sigma(i)}^\sigma$ on the left and on the right of the initial discontinuity, respectively. The vector $\mathbf{z}$ is the primitive vector. The superscript σ indicates that these values are taken at the cell face depending of the scheme order of accuracy.

In the case of single-material fluid flow, the numerical flux can be also calculated by means of an approximate Riemann solver as Rusanov method [5], HLL [6] and others. These flux approximations can be applied straightforwardly to solve multi-material fluid flow modeled by Equations (2) and (3). Thus, calculations in the whole domain can be executed in the same manner, both in pure and mix cells. We refer this straightforward extension of the conventional numerical flux approximation to the multi-material case as a *standard flux approximation*.

The usage of the standard flux approximations results in significant numerical smearing of material interfaces. These approximations do not regard the presence of the sharp material interface inside mix cells. Thus, if a non-zero mass flux appears from a mix cell to an adjacent pure cell it inevitably would result in non-zero mass fluxes of all components containing in the mix cell.

As a solution to this problem, we suggest modifications for the numerical flux approximation for those cell faces that border at least one mix cell. These modifications are made on the base of the sub-cell material interface reconstruction and the CRP solution. We call this

flux approximation a *CRP-based flux approximation.* If no mix cell borders the face, numerical flux is calculated by one of standard flux approximations.

3.1 The CRP-based flux approximation

In order to introduce the modifications of the flux approximation for the cell faces bordering at least one mix cell, we have to consider two possible situations when one cell is pure and another is mixed or both are mixed. For clarity, let the pure cell is filled with the material "A", and the mix cell is filled with the materials "A" and "B".

If the connectivity at a cell face σ is of the type "mix-pure", we assume that the contact surface between materials "A" and "B" in the mix cell is parallel to the face σ and located at a distance δ, and material "A" in the mix cell borders the material "A" in the pure cell. The distance $\delta = \alpha_A V/S_\sigma$, where α_A is the material "A" volume fraction in the mix cell, V is the mix cell volume and S_σ is the face σ area.

For a given mixture state vector $\mathbf{z} = (\rho, \mathbf{v}, p, \beta_A, \beta_B)^T$ we can derive vectors $\mathbf{z}_A$ and $\mathbf{z}_B$, characterizing the state of materials "A" and "B" in the mix cell: $\mathbf{z}_A = (\rho^o{}_A, \mathbf{v}, p, 1, 0)^T$ and $\mathbf{z}_B = (\rho^o{}_B, \mathbf{v}, p, 0, 1)^T$, with $\rho^o{}_{A/B} = \rho^o{}_{A/B}(p, T)$ being material densities defined with the corresponding EOSs. The material densities are related with volume fractions $\alpha_{A/B}$, $\alpha_{A/B} = \rho\beta_{A/B}/\rho^o{}_{A/B}$. Thus, the position of the material interface in the mix cell is uniquely located. Note that in 1D such a reconstruction of the interface is natural.

Applying the interface reconstruction in the mix cell, we then calculate the numerical flux $\mathbf{F}_\sigma$ at the face σ by means of the CRP solution with the proper initial data $\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}$ as follows:

$$\mathbf{F}_\sigma = \mathbf{F}^{CRP}\left(\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}, \delta\right) \tag{10}$$

where the superscripts (L) and (R) indicate the left and right cells, respectively. It is obviously that we can define the term $\mathbf{F}^{CRP}\left(\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}, \delta\right)$ for $\delta = 0$ ($\alpha_A = 0$) as

$$\mathbf{F}^{CRP}\left(\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}, 0\right) = \mathbf{F}\left(\mathbf{Q}^{RP}\left(\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(R)}\right)\right), \tag{11}$$

and for $\delta = h = V/S_\sigma$ ($\alpha_A = 1$) as

$$\mathbf{F}^{CRP}\left(\mathbf{Z}_B^{(L)}, \mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}, h\right) = \mathbf{F}\left(\mathbf{Q}^{RP}\left(\mathbf{Z}_A^{(L)}, \mathbf{Z}_A^{(R)}\right)\right). \tag{12}$$

Details of the CRP solver are given can be found in [1].

If the connectivity at a cell face σ is of the type "mix-mix", a reasonable assumption is that the contact surface should in a some way cross the cell face σ, passing through the mix cells. Following the procedure mentioned above, one can calculate component vectors $\mathbf{z}_{A/B}$ and volume fractions $\alpha_{A/B}$ for each cell and then reconstruct contact surface in the cells as shown in Figure 1b. In general, reconstructed contact surfaces do not match at the cell face ($\alpha_A^{(L)} < \alpha_A^{(R)} < 0.5$). One can suggest patterns for the sub-cell interface reconstruction with continuous behavior of the contact surface at the face σ. We assume that the contact surface crosses the face in an intermediate point between points "1" and "2" as shown in Figure 1c.

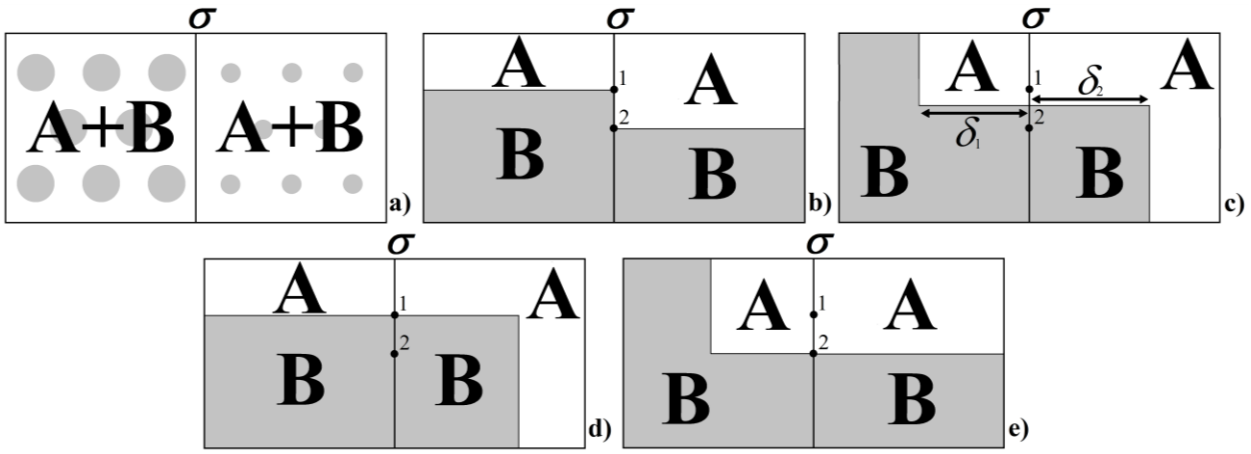


Figure 1. Patterns for the sub-cell interface reconstruction in the case of "mix-mix " cell connectivity.

For the patterns of Figure 1c-1e, we can calculate the numerical flux as a linear combination of two fluxes defined on corresponding CRP solutions as follows:

$$\mathbf{F}_\sigma = \tilde{\alpha}\mathbf{F}^{CRP}\left(\mathbf{Z}_{\mathrm{B}}^{(\mathrm{L})}, \mathbf{Z}_{\mathrm{A}}^{(\mathrm{L})}, \mathbf{Z}_{\mathrm{A}}^{(\mathrm{R})}, \delta_1\right) - (1-\tilde{\alpha})\mathbf{F}^{CRP}\left(\mathbf{Z}_{\mathrm{A}}^{(\mathrm{R})}, \mathbf{Z}_{\mathrm{B}}^{(\mathrm{R})}, \mathbf{Z}_{\mathrm{B}}^{(\mathrm{L})}, \delta_2\right), \tag{13}$$

where $\tilde{\alpha} = w\alpha_{\mathrm{A}}^{(\mathrm{L})} + (1-w)\alpha_{\mathrm{A}}^{(\mathrm{R})}$, $w \in [0, 1]$. The parameter w defines the intermediate point in Figure 1c. Thus, the numerical flux across a face σ is calculated by means of solving one or two appropriate CRPs.

It should be noted that the interface reconstruction with local 1D patterns is not a real one. It is implemented independently for each cell face on the base of only volume fraction values in mix cells bordering the considered face. This results in a very simple algorithm. However, it has certain drawbacks that are a penalty of these simplifications. We discuss this issue in section of numerical results. In order to overcome these drawbacks, we implement a method with the directional splitting technique which allows to better utilize a 1D nature of the proposed flux approximation.

An open issue is the choice of w in Equation (13) when a "mix-mix" pattern is used. In that case the sub-cell interface reconstruction can be done uniquely just on the base of volume fractions. Most calculations presented in this paper are done with $w = 0$ that corresponds to a "mix-mix" pattern of Figure 1e. In some cases we use also values $w = 0.5$ and $w = 1$. Anyway, the parameter is used as constant in the calculations. A dynamical approach where it is evaluated depending on local distributions of mass and volume fractions is the purpose of our future research.

3.2 The scheme with splitting in directions

The sub-cell interface reconstruction of the previous subsection is by nature locally one-dimensional and carried out independently for all cell faces. To adapt this approach to multi-dimensional calculations, we develop also its implementation based on the method of splitting in directions. In the case of Cartesian grids the explicit time integration numerical scheme is written as

$$\mathbf{q}_i^{n+1} = \mathbf{q}_i^n - \frac{\Delta t}{V_i}(\mathbf{F}_{1x} + \mathbf{F}_{2x} + \mathbf{F}_{1y} + \mathbf{F}_{2y}), \tag{14}$$

where $\mathbf{F}_{1x}$, $\mathbf{F}_{2x}$ and $\mathbf{F}_{1y}$, $\mathbf{F}_{2y}$ are numerical fluxes along the x-axis and the y-axis, respectively. For certainty, we consider the 2D case here.

A directionally splitted scheme can be written as follows:

$$\mathbf{q}_i^{nx} = \mathbf{q}_i^n - \frac{\Delta t}{V_i}(\mathbf{F}_{1x} + \mathbf{F}_{2x}), \mathbf{q}_i^{n+1} = \mathbf{q}_i^{nx} - \frac{\Delta t}{V_i}(\mathbf{F}_{1y} + \mathbf{F}_{2y}), \tag{15}$$

where the fluxes $\mathbf{F}_{1x}$ and $\mathbf{F}_{2x}$ are calculated with values of the n-th time layer $\{\mathbf{q}^n\}$, and the fluxes $\mathbf{F}_{1y}$ and $\mathbf{F}_{2y}$ are calculated with those of the nx-th time layer $\{\mathbf{q}^{nx}\}$.

4 NUMERICAL RESULTS

In our previous paper [1], numerical examples mostly concerned one-dimensional problems. It was shown that CRP-based flux approximation always allows to capture the interface within just one computational cell. In the 1D case there should not be two neighboring mix cells and only a "mix-pure" pattern for the sub-cell interface reconstruction can exist. This is not the case in 2D calculations. We have to use "mix-mix" pattern for the sub-cell interface reconstruction shown in Figure 2c, which is not so definite procedure as in the case of "mix-pure" cell connectivity. In this section we present numerical results of calculations with various implementations of the "mix-mix" interface reconstruction and discuss their interface capturing properties.

4.1 The problem of shock/contact interaction

The problem statement is sketched in Figure 2. At initial the lead-gas interface has a triangle disturbance with the depth of 0.0035 cm. The computational domain is $0 \le x \le 1.6$ cm, $0 \le y \le 0.2$ cm occupied by a lead sample with a length of 0.5 cm and a gas. Boundary conditions are following: a rigid wall boundary condition is imposed at $x = 0$, $y = 0$, $y = 0.2$ cm and a non-reflecting boundary condition is imposed at $x = 1.6$ cm. The computational grid consists of 800 and 100 along the x-axis and the y-axis, respectively. The CFL number is equal to 0.5.

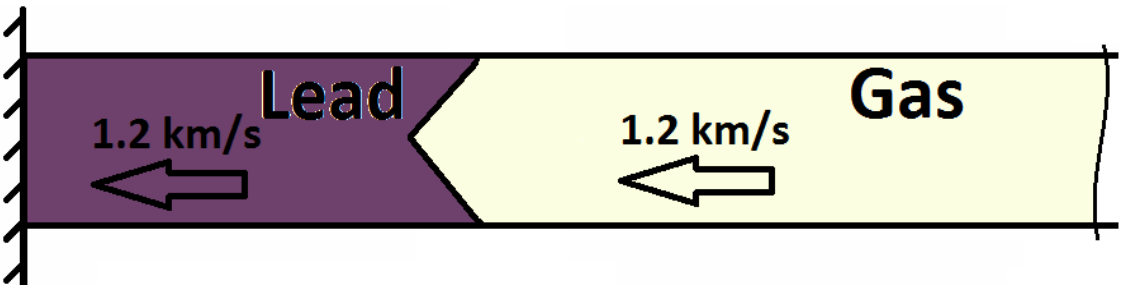


Figure 2. A sketch of the shock/contact interaction problem.

To highlight the importance of keeping the interface region as small as possible, calculations of this problem are carried out with the first order of accuracy using Rusanov and CRP-based flux approximation. In the CRP-based calculation we use the directional splitting technique and a "mix-mix" pattern of Figure 1e.

As can be seen from Figure 3, the Rusanov flux calculation results in a numerical effect of pseudo-fracturing in lead (a narrow region of small density near the interface). The jet formation is found to be underestimated (see Figure 3). Figure 4 shows the jet formation in the CRP-based flux calculation with no any pseudo-fracture in lead. The comparison of the

cell type distribution for both calculations at a time moment t = 7.25 μs is presented in Figure 5. Black color indicates mix cells, white and grey colors indicate pure gas cells and pure lead cells, respectively. We can observe that the lead-gas interface is captured within few computational cells in the CRP-based flux calculation, while it turns to be an expanded domain in the Rusanov flux calculation.

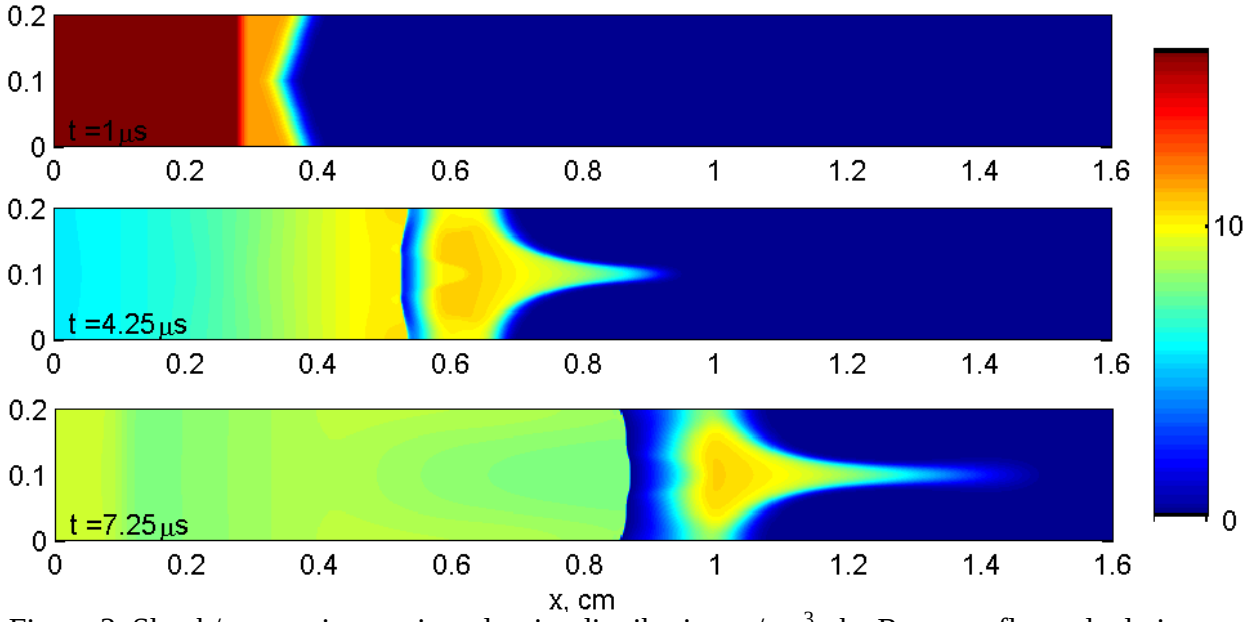


Figure 3. Shock/contact interaction: density distribution, g/cm^3, the Rusanov flux calculation.

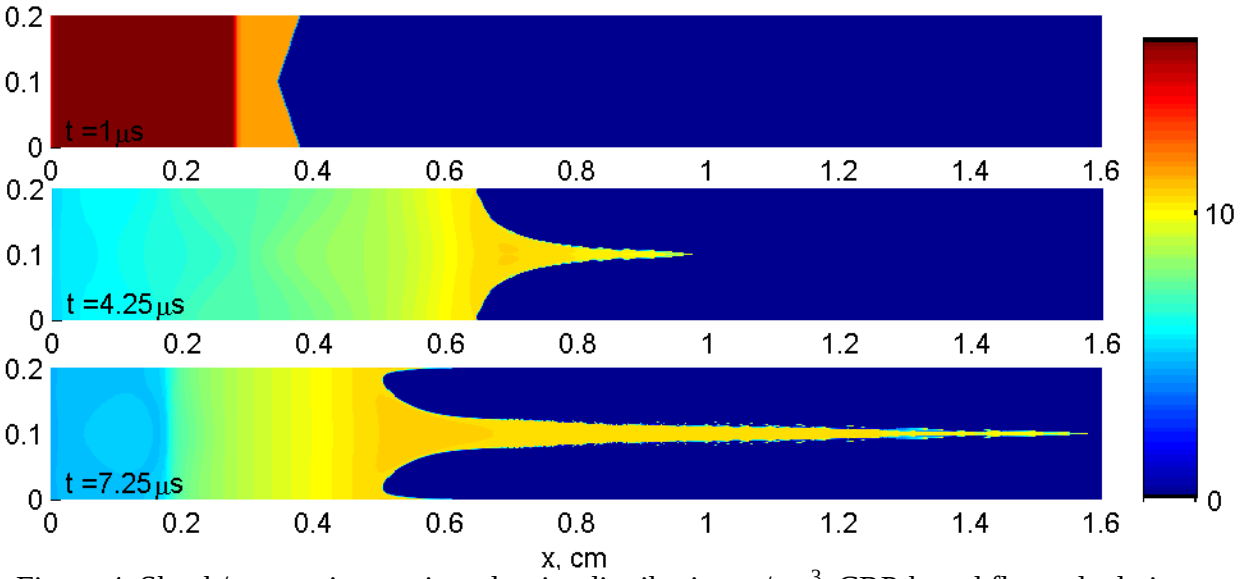


Figure 4. Shock/contact interaction: density distribution, g/cm^3, CRP-based flux calculation.

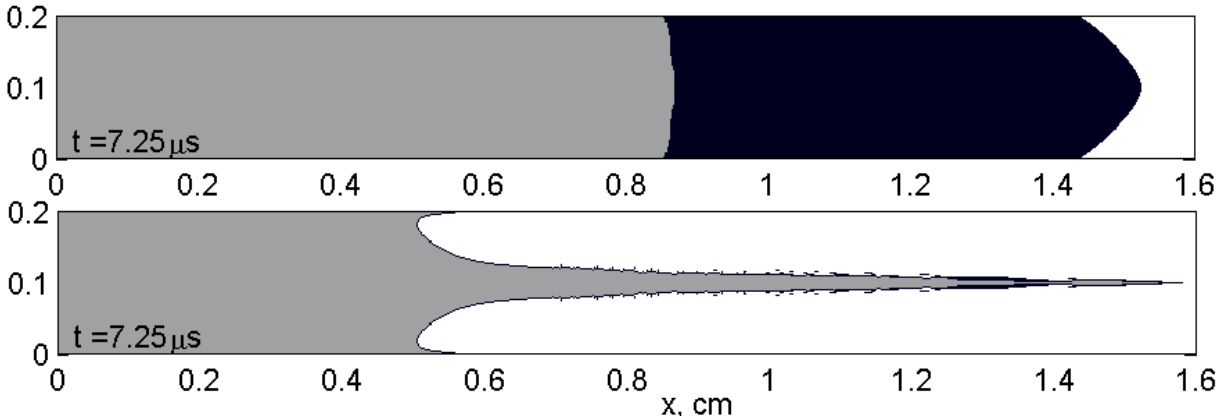


Figure 5. Shock/contact interaction: cell type distribution,
the Rusanov flux (top) and the CRP-based flux (lower), t = 7.25 μs.

4.2 The problem of underwater air bubble collapse

The problem statement is the same as in [7]. The computational domain is $0 \le x \le 1.2$ cm, $0 \le y \le 1.2$ cm. The air bubble diameter is 0.6 cm, and its center is at (0.6, 0.6) cm. The EOS of air and water is taken in the form of a stiffened gas EOS. Periodical boundary conditions are imposed on the bottom and top boundary of the computational domain. The computational grid consists 200 of computational cells in each direction. The CFL number is equal to 0.5.

Figure 6 shows the cell type distributions at a time moment t = 3.25 μs for six variants of calculation. Black color indicates mix cells, white and grey colors indicate pure air cells and pure water cells, respectively. The top row of figures corresponds to calculations with the directional splitting off. The lower row corresponds to calculation with directional splitting. The left column corresponds to a "mix-mix" pattern of Figure 1c with w = 0.5, the middle column - a "mix-mix" pattern of Figure 1d and the right column - a "mix-mix" pattern of Figure 1e.

One can conclude from Figure 6 that the directional splitting technique improves interface capturing properties. It can be also observed that different "mix-mix" patterns for the sub-cell interface reconstruction result in different forms of numerical smearing, especially in the case of non splitting calculations.

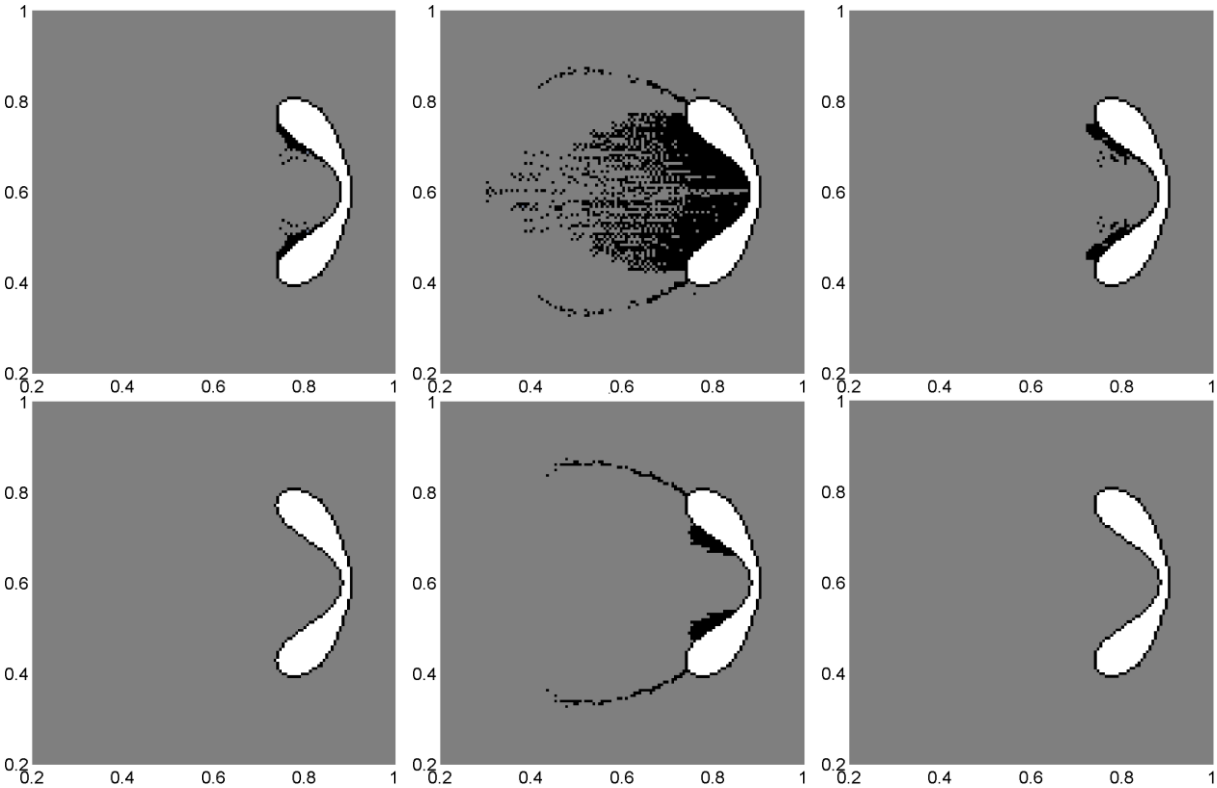
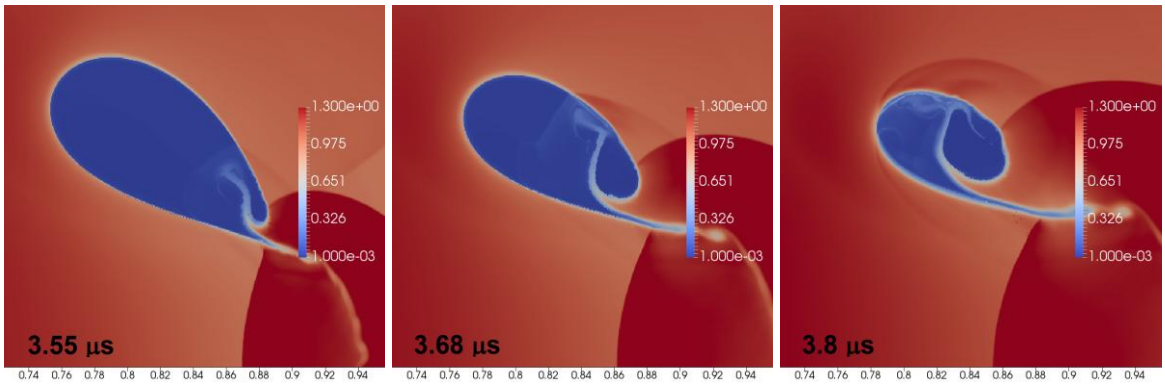


Figure 6. The underwater air bubble collapse: the cell type distribution at t = 3.25 μs. The comparison of different "mix-mix" patterns and directional splitting.

Figure 7 shows results of calculations with the CRP-based flux on a fine grid consisting of 2000 computational cells in each direction. Such a grid resolution allows to capture the formation of small scale secondary jets. One can see a narrow water jet that slits the bubble and hits the water from the base. These calculations have been performed with the directional splitting, a "mix-mix" pattern of Figure 1e in the sub-cell interface reconstruction, and the second order of accuracy.



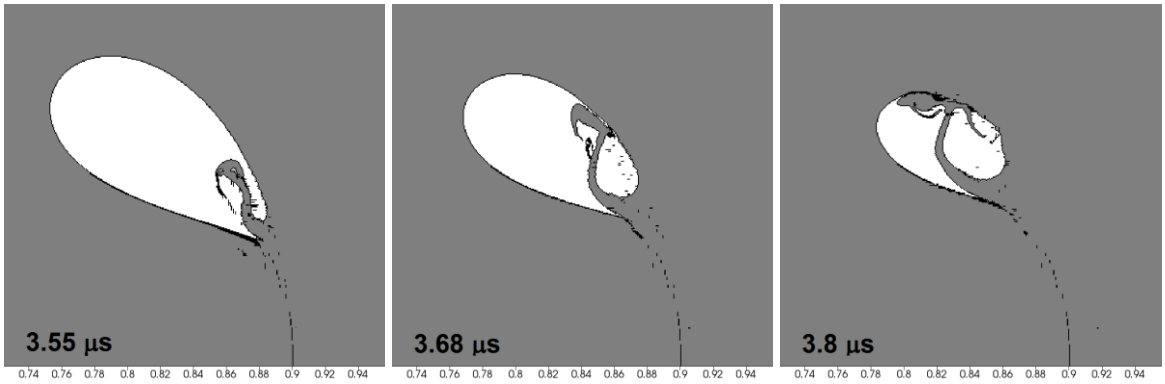


Figure 7. The underwater air bubble collapse: the formation of secondary jets.
The density distribution, g/cm^3 (top) and the cell type distribution (low).

4.3 The triple point problem

The problem statement is taken from [8]. The computational domain is $0 \le x \le 7$ and $0 \le y \le 3$. The first sub-domain ($0 \le x \le 1$, $0 \le y \le 3$) contains a high-pressure high-density material. The second sub-domain ($1 < x \le 7$, $0 \le y < 1.5$) contains a low-pressure high-density material. The third sub-domain ($1< x \le 7$, $1.5 \le y \le 3$) a low-pressure low-density material. All three materials are governed by the ideal gas EOS. The rigid wall condition is imposed at all boundaries of the computational domain. We solve this problem as a two-material problem and do not distinguish materials in the first and the third sub-domain.

Figure 8 shows results with the directional splitting, a "mix-mix" pattern of Figure 1e, and the second order of accuracy on a computational grid of 4200 and 1800 cells along the x-axis and the y-axis, respectively. The CFL number is equal to 0.5. One can observe that the material interface undergoes a Kelvin-Helmholtz instability that results in a large deformation of the interface with the formation of small-scale jet-like structures.

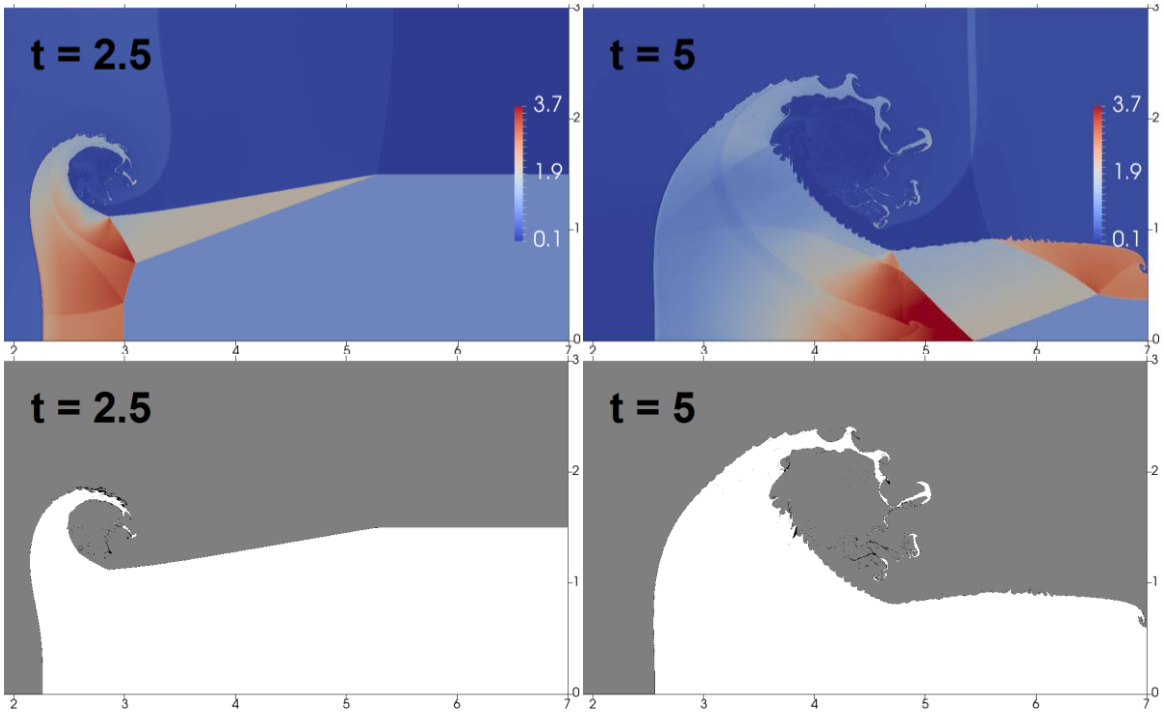


Figure 8. The triple point problem: the formation of Kelvin-Helmholtz instability.
The density distribution, g/cm^3 (top) and the cell type distribution (low).

5 CONCLUSIONS

An Eulerian finite-volume method for multi-material fluid flows modeled by a multiphase medium with pressure, temperature and velocity equilibrium is discussed. The method belongs to the class of interface capturing approaches and aims to reduce numerical smearing of material interfaces. To achieve this goal, we have introduced the sub-cell interface reconstruction procedure based on simple interfaces patterns. The proposed sub-cell interface structure is taken into account in calculating the numerical flux across a cell face bordering a mix cell. This is performed with a proper IVP - a Composite Riemann Problem (CRP) that involves both a point of initial discontinuity and a material contact point. The numerical flux approximation based on this CRP solution strongly reduces the interface smearing region. Numerical experiments have shown the reduction of the interface region up to a few computational cells in 2D tests.

The work reported in this paper was partially supported by the grant No 14-11-00872 from the Russian Scientific Fund.

REFERENCES

[1] I.S. Menshov, P.P. Zakharov. On the composite Riemann problem for multimaterial fluid flows. *International Journal for Numerical Methods in Fluids*, **76**(2), 109-127, 2014.

[2] T. Nonomura, K. Kitamura, K. Fujii. A simple interface sharpening technique with a hyperbolic tangent function applied to compressible two-fluid modeling. *Journal of Computational Physics,* **258,** 95–117, 2014.

[3] K.K. So, X.Y. Hu, N.A. Adams. Anti-diffusion interface sharpening technique for two-phase compressible flow simulations. *Journal of Computational Physics,* **231(11),** 4304–4323, 2012.

[4] I. Menshov and P. Zakharov. A Composite Riemann Solver for Improving Interface Capturing in Multimaterial Calculations. *MULTIMAT 2015*, Wurzburg, Germany, September 7 - 11, 2015.

[5] V.V. Rusanov. Calculation of Interaction of Non-Steady Shock Waves with Obstacles. *USSR Computational Mathematics and Mathematical Physics*, **1(2)**, 267-279, 1961.

[6] A . Harten, P.D. Lax, and B . Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Review*, **25(1),** 35-61, 1983.

[7] R.R. Nourgaliev, T.N. Dinh, T.G. Theofanous. Adaptive characteristics-based matching for compressible multifluid dynamics. *Journal of Computational Physics,* **213(2),** 500–529, 2006.

[8] M. Kucharik, R.V. Garimella, S.P. Schofield, M.J. Shashkov . A comparative study of interface reconstruction methods for multi-material ALE simulations. *Journal of Computational Physics.,* **229(7),** 2432-2452, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SOME ACOUSTIC-TRANSPORT SPLITTING SCHEMES FOR TWO-PHASE COMPRESSIBLE FLOWS

Simon PELUCHON[1], Gérard GALLICE[2], and Pierre-Henri MAIRE[2]

[1] CEA-CESTA
15 avenue des Sablières, CS 60001, 33116 Le Barp cedex, FRANCE
e-mail: simon.peluchon@CEA.FR

[2] CEA-CESTA
15 avenue des Sablières, CS 60001, 33116 Le Barp cedex, FRANCE
e-mail: {gerard.gallice, pierre-henri.maire}@CEA.FR

Keywords: Compressible flow, Two-phase flow, Finite volume method, Carbuncle, Splitting operator.

Abstract. *In this paper, we present a splitting strategy to simulate compressible two-phase flows using the five-equation model [1]. The main idea of the splitting [2] is to separate the acoustic and transport phenomena. The acoustic step is solved as a Lagrangian step by using different schemes [3, 4, 5] based on approximate Riemann solvers. On the one hand, since the acoustic time step driven by the fast sound velocity is very restrictive, an implicit treatment of the Lagrangian step is performed. On the other hand, an explicit scheme is used for the transport step driven by the slow material waves. The global scheme resulting from this splitting operator strategy is conservative, positive, and preserves contact discontinuities. Numerical simulations of compressible diphasic flows are presented on 2d-structured grids. The implicit-explicit strategy allows large time steps, which do not depend on the fast acoustic waves.*

On supersonic velocity test case, a carbuncle phenomenon can occur with usual schemes for the Lagrangian part. However, we show that this carbuncle does not appear when the Lagrangian step is solved with a genuinely two-dimensional scheme based on a nine-point stencil [5].

1 INTRODUCTION

We are interested in solving compressible two-phase flows using a splitting strategy. Flows involving an almost incompressible liquid and a compressible gas are considered here. Using a diffuse interface approach [1, 6, 7], the same equations are solved in the entire domain. Therefore, the liquid is modelled by a compressible fluid in which the Mach number is very low. Several time scales are involved in this configuration: the scale of the fast acoustic waves which leads to an important CFL restriction of the time step for explicit schemes, and the scale of slow material waves.

In this paper, we propose an extension of the splitting strategy presented in [2, 8] to compressible two-phase flows using the five-equation system [1]. The splitting strategy is done with a Lagrange Projection algorithm. The acoustic part is solved in a Lagrangian form using several schemes [3, 5]. In Low Mach computations, an implicit treatment of this step should be done since the fast acoustic waves induce very small time steps. We are interested here in supersonic test cases, thus the acoustic step will be solved using an explicit scheme. The transport step is treated explicitly using a finite volume scheme.

The outline of this paper is as follow. In the second section we present the governing equations of two-phase flows. In section 3, the operator splitting in the one dimensional case is presented. We also describe the extension of the numerical scheme to two-dimensional structured grids in section 4. Several numerical schemes used in the Lagrangian part are recalled in section 5. Finally some numerical results for two-dimensional two-phase flows are presented in section 6. On supersonic computations, a carbuncle phenomenon appears with schemes based on a five-point stencil for the acoustic part. The shock instabilities disappear when a genuinely two-dimensional scheme is used in the acoustic step [5].

2 FIVE-EQUATION SYSTEM

We denote by ρ_i, ϵ_i and p_i the density, the internal energy and the pressure of the fluid phase $i = 1, 2$. Each fluid is equipped with an Equation Of State (EOS) of the form $p_i = p_i(\rho_i, \epsilon_i)$. The sound velocity c_i of each phase is defined by $c_i^2 = \frac{\partial p_i}{\partial \rho_i}\Big|_{s_i}$, where s_i is the entropy.

We introduce the volume fraction z_i of each fluid and we have the relation $z_1 + z_2 = 1$. In the sequel we denote by $z = z_1$ the volume fraction of the first fluid. The mixture density and mixture internal energy can be defined through the volume fraction, the density and internal energy of each fluid:

$$\rho = z_1\rho_1 + z_2\rho_2, \tag{1}$$
$$\rho\epsilon = z_1\rho_1\epsilon_1 + z_2\rho_2\epsilon_2. \tag{2}$$

Both phases share the same velocity $\mathbf{u}$, and the same pressure p. The five-equation system with isobaric closure derived in [1] reads:

$$\begin{aligned}
&\partial_t(\rho_1 z_1) + \nabla\cdot(\rho_1 z_1\mathbf{u}) = 0, \\
&\partial_t(\rho_2 z_2) + \nabla\cdot(\rho_2 z_2\mathbf{u}) = 0, \\
&\partial_t(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \nabla p = 0, \\
&\partial_t(\rho e) + \nabla\cdot((\rho e + p)\,\mathbf{u}) = 0, \\
&\partial_t z + \mathbf{u}\cdot\nabla z = 0, \\
&p = p_1 = p_2,
\end{aligned} \tag{3}$$

where $e = \epsilon + \frac{\mathbf{u}^2}{2}$ is the total energy of the mixture.

The evolution of the volume fraction is governed by a non-conservative equation. This formulation preserves contact discontinuities, i.e. the evolution of constant pressure and velocity profiles, see [1].

We only consider here a stiffened gas EOS for each fluid,

$$p_i = \rho_i \epsilon_i \left(\gamma_i - 1\right) - \gamma_i \pi_i, \tag{4}$$

where $\gamma_i > 1$ is the adiabatic exponent and $\pi_i \geq 0$ a reference pressure.

The mixture pressure of two stiffened gases is $p = \left(\gamma - 1\right)\rho\epsilon - \gamma\pi$ where the mixture parameters γ and π are defined by:

$$\gamma = 1 + \frac{1}{\sum_{i=1}^{2} \frac{z_i}{\gamma_i - 1}} \quad \text{and} \quad \pi = \frac{\gamma - 1}{\gamma} \sum_{i=1}^{2} \frac{z_i \gamma_i \pi_i}{\gamma_i - 1}. \tag{5}$$

The speed of sound for a stiffened gas reads:

$$c_i^2 = \gamma_i \frac{p + \pi_i}{\rho_i}.$$

3 SPLITTING STRATEGY FOR ONE DIMENSIONAL CASE

For the sake of simplicity, in this section a one dimensional problem is considered. The idea is to separate the acoustic and the transport phenomena according to their own propagation speed [2]. First, note that the evolution equation of $\rho_2 z_2$ in (3) can be replaced by the evolution equation of the mixture density ρ which is $\partial_t \rho + \partial_x(\rho u) = 0$. Then, the system (3) is expanded to:

$$\begin{aligned}
\partial_t \rho + \rho \partial_x u + u \partial_x \rho &= 0, \\
\partial_t \left(\rho_1 z\right) + \rho_1 z \partial_x u + u \partial_x (\rho_1 z) &= 0, \\
\partial_t \left(\rho u\right) + \rho u \partial_x u + \partial_x p + u \partial_x (\rho u) &= 0, \\
\partial_t \left(\rho e\right) + \rho e \partial_x u + \partial_x \left(p u\right) + u \partial_x (\rho e) &= 0, \\
\partial_t z + u \partial_x z &= 0.
\end{aligned} \tag{6}$$

This system is split into the following two systems. The first system takes into account the propagation of the acoustic waves only:

$$\begin{aligned}
\partial_t \rho + \rho \partial_x u &= 0, \\
\partial_t \left(\rho_1 z\right) + \rho_1 z \partial_x u &= 0, \\
\partial_t \left(\rho u\right) + \rho u \partial_x u + \partial_x p &= 0, \\
\partial_t \left(\rho e\right) + \rho e \partial_x u + \partial_x \left(p u\right) &= 0, \\
\partial_t z &= 0.
\end{aligned} \tag{7}$$

The transport system takes into account the propagation of the material waves through the fluid:

$$\begin{aligned}
\partial_t \rho + u \partial_x \rho &= 0, \\
\partial_t \left(\rho_1 z_1\right) + u \partial_x \left(\rho_1 z_1\right) &= 0, \\
\partial_t \left(\rho u\right) + u \partial_x \left(\rho u\right) &= 0, \\
\partial_t \left(\rho e\right) + u \partial_x \left(\rho e\right) &= 0, \\
\partial_t z + u \partial_x z &= 0.
\end{aligned} \tag{8}$$

The overall algorithm for a time step between t^n and t^{n+1} reads:

Step 1: From a state $(\rho, \rho_1 z, \rho u, \rho e, z)^n$, compute the approximation of the acoustic system (7) $(\rho, \rho_1 z, \rho u, \rho e, z)^{n+1,L}$.

Step 2: Find the fluid state $(\rho, \rho_1 z, \rho u, \rho e, z)^{n+1}$ by solving the transport system (8) with the initial state $(\rho, \rho_1 z, \rho u, \rho e, z)^{n+1,L}$.

Before going any further, we introduce some classical notations. We consider a discretization of the real line into cells $\Big([x_{i-1/2}, x_{x+1/2}]\Big)_{i\in Z}$. The mesh interfaces are $x_{i+1/2} = i\Delta x$ for $i \in Z$, where $\Delta x > 0$ is the grid step. The centre of the cell i is $x_i = \frac{x_{i+1/2}+x_{i-1/2}}{2}$. The time variable is discretized by $t^n = n\Delta t$ for $n \in N$, where $\Delta t > 0$ is the time step. We use a Finite Volume method, in which a_i^n is the approximation of

$$\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} a(x, t^n)\mathrm{d}x,$$

where $(x, t) \mapsto a(x, t)$ is a fluid parameter.

3.1 Acoustic step

First, the non-conservative system (7) is rearranged. The second, third and fourth equations of acoustic system are combined with the evolution equation of the mixture density. Then, the first equation is divided by the mixture density. Finally, the acoustic system reads:

$$\begin{aligned}
\partial_t \vartheta - \frac{1}{\rho}\partial_x u &= 0, \\
\partial_t y &= 0, \\
\partial_t u + \frac{1}{\rho}\partial_x p &= 0, \\
\partial_t e + \frac{1}{\rho}\partial_x (pu) &= 0, \\
\partial_t z &= 0,
\end{aligned} \tag{9}$$

where $\vartheta = \frac{1}{\rho}$ is the specific volume, and $y = \frac{\rho_1 z}{\rho}$ is the mass fraction of fluid 1.

Let us introduce the mass variable m defined by $\mathrm{d}m = \rho \mathrm{d}x$. The acoustic step consists in resolving a hyperbolic system of the form

$$\partial_t V + \partial_m G(V) = 0, \tag{10}$$

with $V = {}^t(\vartheta, y, u, e, z)$ and $G = {}^t(-u, 0, p, pu, 0)$.

This system is strictly hyperbolic and the five real eigenvalues are $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (-\rho c, 0, 0, 0, \rho c)$. Let us denote by $a = \rho c$ the Lagrangian speed of sound. The first and the fifth characteristic fields are genuinely non-linear while the waves associated to λ_2, λ_3 and λ_4 are linearly degenerate.

A classical approach to solve hyperbolic system like (10) is to use a Godunov type scheme. The update formula for the Lagrangian step reads

$$V_i^{n+1,L} = V_i^n - \frac{\Delta t}{\Delta m_i}\left(H_{i+1/2}^n - H_{i-1/2}^n,\right), \tag{11}$$

where $H = {}^t\Big(h^{(1)}, 0, h^{(3)}, h^{(4)}, 0\Big)$ is the numerical flux, and $\Delta m_i = \rho_i^n \Delta x$.

In the sequel, we use introduce the notation $[\![x]\!]_i = x_{i+1/2} - x_{i-1/2}$ for any quantity $(x_{i+1/2})_i$ defined at an interface.

The update formulae for the Eulerian variables are directly given by:

$$\begin{aligned}
&L_i \rho_i^{n+1,L} = \rho_i^n, \\
&L_i \left(\rho_1 z\right)_i^{n+1,L} = \left(\rho_1 z\right)_i^n, \\
&L_i \left(\rho u\right)_i{}^{n+1,L} = (\rho u)_i^n - \frac{\Delta t}{\Delta x}[\![h^{(3)}]\!]_i, \\
&L_i \left(\rho e\right)_i{}^{n+1,L} = (\rho e)_i^n - \frac{\Delta t}{\Delta x}[\![h^{(4)}]\!]_i, \\
&z_i^{n+1,L} = z_i^n,
\end{aligned} \tag{12}$$

with $L_i = 1 + \frac{\Delta t}{\Delta x}[\![-h^{(1)}]\!]_i$.

The acoustic step is stable under the classical Courant-Friedrichs-Lewy stability condition, given by

$$\frac{\Delta t}{\Delta x} \max_{i \in Z} \left(\frac{a_{i+1/2}}{\min(\rho_i^n, \rho_{i+1}^n)} \right) \leq \frac{1}{2}. \tag{13}$$

where $a_{i+1/2}$ is the Lagrangian sound speed at the interface $i + 1/2$.

3.2 Transport step

The transport system (8) is a convection system at the fluid velocity u. Therefore, it is simply approximated by an upwind Finite Volume scheme:

$$\phi_i^{n+1} = \phi_i^{n+1,L} - \frac{\Delta t}{\Delta x} \left(\left(\tilde{u}_{i+1/2}\right)^- \left(\phi_{i+1}^{n+1,L} - \phi_i^{n+1,L}\right) + \left(\tilde{u}_{i-1/2}\right)^+ \left(\phi_i^{n+1,L} - \phi_{i-1}^{n+1,L}\right) \right), \tag{14}$$

where ϕ denotes $\rho, \rho_1 z, \rho u, \rho e, z$, and with the classical notation $u^{\pm} = \frac{u \pm |u|}{2}$.

In order to have a fully conservative scheme for $\rho, \rho_1 z, \rho u$ and ρe, the choice of the propagation velocity $\tilde{u}_{i+1/2}$ should depend on the flux of the acoustic step.
Indeed, let us rewrite (14) as

$$\phi_i^{n+1} = \phi_i^{n+1,L} \left(1 + \frac{\Delta t}{\Delta x}[\![\tilde{u}]\!]_i\right) - \frac{\Delta t}{\Delta x}[\![\phi^{n+1,L}\tilde{u}]\!]_i,$$

where $\phi_{i+1/2}^{n+1,L}$ is defined as the upwind value of ϕ with respect to the sign of the velocity $\tilde{u}_{i+1/2}$, i.e. $\phi_{i+1/2}^{n+1,L} = \phi_i^{n+1,L}$ if $\tilde{u}_{i+1/2} > 0$, and $\phi_{i+1}^{n+1,L}$ otherwise.

We also denote by $\tilde{L}_i$ the volume variation $1 + \frac{\Delta t}{\Delta x}[\![\tilde{u}]\!]_i$. Combining this equation with explicit formulae (12) for the acoustic step, one can see that the overall scheme to compute the state at the time t^{n+1} from the state at t^n reads:

$$\begin{aligned}
\rho_i^{n+1} &= \frac{\tilde{L}_i}{L_i}\rho_i^n - \frac{\Delta t}{\Delta x}[\![\tilde{u}\rho^{n+1,L}]\!]_i, \\
(\rho_1 z)_i^{n+1} &= \frac{\tilde{L}_i}{L_i}(\rho_1 z)_i^n - \frac{\Delta t}{\Delta x}\left([\![\tilde{u}(\rho_1 z)^{n+1,L}]\!]_i\right), \\
(\rho u)_i^{n+1} &= \frac{\tilde{L}_i}{L_i}(\rho u)_i^n - \frac{\Delta t}{\Delta x}\left([\![\tilde{u}(\rho u)^{n+1,L}]\!]_i + \frac{\tilde{L}_i}{L_i}[\![h^{(3)}]\!]_i\right),
\end{aligned} \tag{15}$$

$$(\rho e)_i^{n+1} = \frac{\tilde{L}_i}{L_i}(\rho e)_i^n - \frac{\Delta t}{\Delta x}\left(\llbracket \tilde{u}(\rho e)^{n+1,L} \rrbracket_i + \frac{\tilde{L}_i}{L_i}\llbracket h^{(4)} \rrbracket_i\right),$$
$$z_i^{n+1} = z_i^{n+1,L} + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} z^n \rrbracket_i - z_i^n \frac{\Delta t}{\Delta x}\llbracket h^{(1)} \rrbracket_i.$$

One can see that the global scheme is conservative if $\tilde{L}_i = L_i$. This choice leads to define the propagation velocity at an interface as the opposite of the first component of the acoustic flux:

$$\tilde{u}_{i+1/2} = -h^{(1)}_{i+1/2}. \tag{16}$$

With this choice, the global scheme reads:

$$\begin{aligned}
\rho_i^{n+1} &= \rho_i^n + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} \rho^{n+1,L} \rrbracket_i, \\
(\rho_1 z)_i^{n+1} &= (\rho_1 z)_i^n + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} (\rho_1 z)^{n+1,L} \rrbracket_i, \\
(\rho u)_i^{n+1} &= (\rho u)_i^n + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} (\rho u)^{n+1,L} - h^{(3)} \rrbracket_i, \\
(\rho e)_i^{n+1} &= (\rho e)_i^n + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} (\rho e)^{n+1,L} - h^{(4)} \rrbracket_i, \\
z_i^{n+1} &= z_i^n + \frac{\Delta t}{\Delta x}\llbracket h^{(1)} z^n \rrbracket_i - z_i^n \frac{\Delta t}{\Delta x}\llbracket h^{(1)} \rrbracket_i.
\end{aligned} \tag{17}$$

It is fully conservative with respect to $\rho, \rho_1 z, \rho u, \rho e$, but not for the color function equation.

The stability condition associated with this transport step reads

$$\frac{\Delta t}{\Delta x} \max_{i \in R}\left(\tilde{u}^+_{i-1/2} - \tilde{u}^-_{i+1/2}\right) < 1. \tag{18}$$

4 EXTENSION TO TWO DIMENSIONS

Without loss of generality, a two dimensional problem discretized over a structured mesh is considered. Let us denote by $\Omega_{l,m}$ a cell of our mesh, $\Gamma_{l+1/2,m}$ (resp. $\Gamma_{l,m+1/2}$) the face of the cell $\Omega_{l,m}$ toward the direction x (resp. y) and $\vec{n}_{l+1/2,m}$ the unit vector normal to $\Gamma_{l+1/2,m}$, see Fig. 1.

System (3) is again split into two parts. The first part is the acoustic step:

$$\begin{aligned}
\partial_t \rho + \rho \nabla \cdot \mathbf{u} &= 0, \\
\partial_t (\rho_1 z) + \rho_1 z \nabla \cdot \mathbf{u} &= 0, \\
\partial_t (\rho \mathbf{u}) + \rho \mathbf{u} \nabla \cdot \mathbf{u} + \nabla \cdot p &= 0, \\
\partial_t (\rho e) + \rho e \nabla \cdot \mathbf{u} + \nabla \cdot (p \mathbf{u}) &= 0, \\
\partial_t z &= 0.
\end{aligned} \tag{19}$$

The second system is the transport step:

$$\begin{aligned}
\partial_t \rho + \mathbf{u} \cdot \nabla \rho &= 0, \\
\partial_t (\rho_1 z) + \mathbf{u} \cdot \nabla (\rho_1 z) &= 0, \\
\partial_t (\rho \mathbf{u}) + \mathbf{u} \cdot \nabla (\rho \mathbf{u}) &= 0, \\
\partial_t (\rho e) + \mathbf{u} \cdot \nabla (\rho e) &= 0, \\
\partial_t z + \mathbf{u} \cdot \nabla z &= 0.
\end{aligned} \tag{20}$$

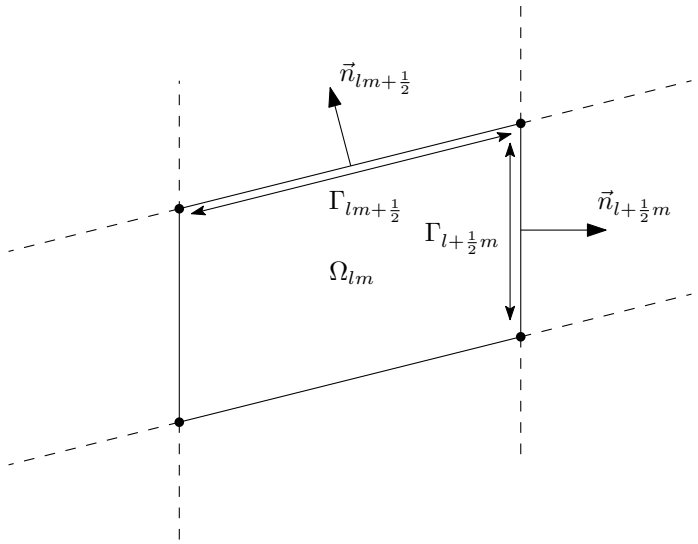


Figure 1: Parameters of a cell in two dimensions.

For any quantity x defined at a cell interface, let us introduce the following notations

$$[\![x_{.m}]\!]_l = x_{l+1/2,m} - x_{l-1/2,m},$$
$$[\![x_{l.}]\!]_m = x_{l,m+1/2} - x_{l,m-1/2}.$$

The volume of a cell $\Omega_{l,m}$ is denoted by $|\Omega_{l,m}|$. Likewise, the length of a face $\Gamma_{l+1/2,m}$ is denoted by $\left|\Gamma_{l+1/2,m}\right|$.

If a is a flow parameter, $a^n_{l,m}$ is an approximation of $\frac{1}{|\Omega_{l,m}|}\int_{\Omega_{l,m}} a(\mathbf{x}, t^n)\mathrm{d}\mathbf{x}$.

Acoustic step: Like in the one dimensional case, the scheme for the acoustic step reads

$$V^{n+1,L}_{l,m} = V^n_{l,m} - \frac{\Delta t}{\rho^n_{l,m}\,|\Omega_{l,m}|}\left([\![H_{.m}\,|\Gamma_{.m}|]\!]_l + [\![H_{l.}\,|\Gamma_{l.}|]\!]_m\right), \tag{21}$$

where $V = {}^t(\vartheta, y, \mathbf{u}, e, z)$ and $H_{l+1/2,m} = H(V_{l,m}, V_{l+1,m})$.

Transport step: For $\phi \in \{\rho, \rho_1 z, \rho\mathbf{u}, \rho e, z\}$, the scheme for the transport step reads:

$$\begin{aligned}\phi^{n+1}_{l,m} &= \phi^{n+1,L}_{l,m}\left(1 - \frac{\Delta t}{|\Omega_{l,m}|}\left([\![h^{(1)}_{.m}\,|\Gamma_{.m}|]\!]_l + [\![h^{(1)}_{l.}\,|\Gamma_{l.}|]\!]_m\right)\right) \\ &+ \frac{\Delta t}{|\Omega_{l,m}|}\left([\![h^{(1)}_{.m}\phi^{n+1,L}_{.m}\,|\Gamma_{.m}|]\!]_l + [\![h^{(1)}_{l.}\phi^{n+1,L}_{l.}\,|\Gamma_{l.}|]\!]_m\right)\end{aligned} \tag{22}$$

5 NUMERICAL FLUXES FOR THE ACOUSTIC STEP

In this section, different Lagrangian schemes used in the acoustic step are defined. The first one uses a classical five-point stencil while the second one is based on a nine-point stencil.

5.1 Scheme based on a five-point stencil

We consider the Lagrangian scheme developed in [3] for the gas dynamics. This Godunov-type scheme based on a simple Riemann solver, see Fig. 2, is positive, entropic and preserves contact discontinuities.

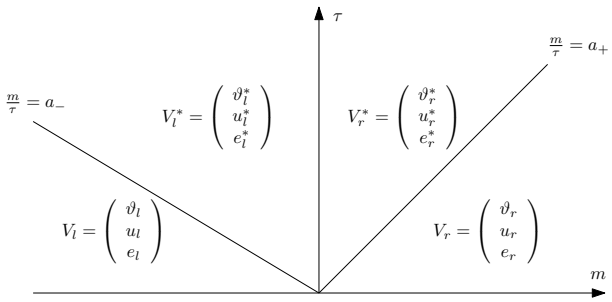


Figure 2: Five-point scheme. Riemann solver.

For gas dynamics, [3] shows that the numerical flux H is equal to the continuous flux G evaluated at an average state:

$$H\left(V_l, V_r\right) = {}^t(-\bar{u}, 0, \bar{p}\vec{n}, \bar{p}\bar{u}, 0)\,, \tag{23}$$

where the velocity and the pressure at the interface between the cells Ω_l and Ω_r are defined by:

$$\begin{aligned} \bar{u} &= \frac{a_-\mathbf{u}_l + a_+\mathbf{u}_r}{a_- + a_+} \cdot \vec{n} - \frac{p_r - p_l}{a_- + a_+}, \\ \bar{p} &= \frac{a_+p_l + a_-p_r}{a_- + a_+} - a_-a_+\frac{\mathbf{u}_r - \mathbf{u}_l}{a_- + a_+} \cdot \vec{n}. \end{aligned} \tag{24}$$

The Lagrangian sound velocities at the interface a_- and a_+ should be large enough in order to preserve the positivity of the scheme, see [3] for more details.

Remark: To compute an implicit scheme for the acoustic step, we use the method described in [2].

5.2 Scheme based on a nine-point stencil

Another choice for the Lagrangian part is to use the scheme developed in [5] for gas dynamics. In this case, fluxes computations are based on a nodal solver.

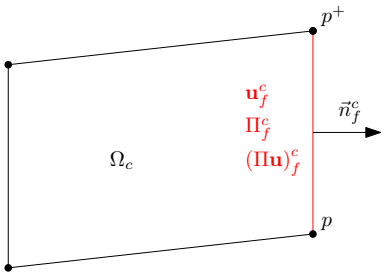


Figure 3: Nine-point scheme. Notations related to the cell Ω_c.

The numerical flux on the face $f = [p, p^+]$ of the cell Ω_c (see Fig. 3) is defined as:

$$H(f) = (\mathbf{u}_f^c \cdot \vec{n}_f^c, 0, \Pi_f^c\vec{n}_f^c, (\Pi\mathbf{u})_f^c \cdot \vec{n}_f^c, 0), \tag{25}$$

where

$$\mathbf{u}_f^c = \frac{1}{2}(\mathbf{u}_p + \mathbf{u}_{p^+}),$$

$$\Pi_f^c = \frac{1}{2}(\Pi_{\bar{p}}^c + \Pi_{\underline{p}^+}^c),$$
$$(\Pi \mathbf{u})_f^c = \frac{1}{2}(\Pi_{\bar{p}}^c \mathbf{u}_p + \Pi_{\underline{p}^+}^c \mathbf{u}_{p^+}).$$

The velocity and pressure on the face f are defined through the nodal velocities $\mathbf{u}_p$ and $\mathbf{u}_{p^+}$, and the nodal pressures, $\Pi_{\bar{p}}^c$ and $\Pi_{\underline{p}^+}^c$, see Fig. 4. The evaluation of these nodal quantities relies on an argument of conservation concerning both momentum and total energy. The nodal pressures are linked to the nodal velocity thanks to half-Riemann problems. Since the nodal velocity $\mathbf{u}_p$ depends of all the cells around the node p, the resulting scheme is based on a nine-point stencil.

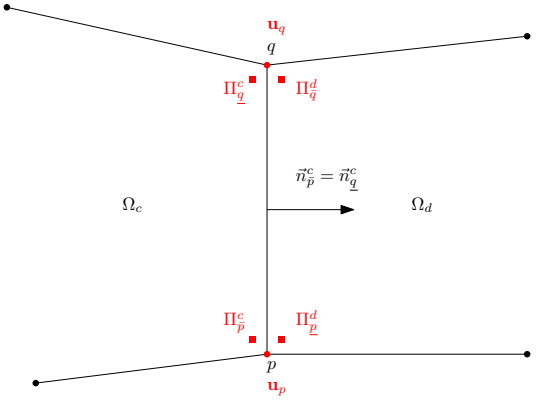


Figure 4: Nine-point scheme. Nodal velocities and pressures related to the face $[p, q]$

6 NUMERICAL RESULTS

In this section, several numerical experiments are presented in order to validate the splitting strategy and compare the different solvers used in the acoustic step.

6.1 Liquid-gas interaction

First, a classical two dimensional test case is considered. The interaction between a bubble of gas and a liquid has already been investigated in [7, 9].

The computational domain is a rectangle of size $L_x = 2$m and $L_y = 1$m described in Fig. 5. The gas is contained in a bubble of radius 0.4m located at the position $(0.5m, 0.5m)$. The shock area is delimited by $x < 0.04$m. Initial conditions are gathered in Table 1.

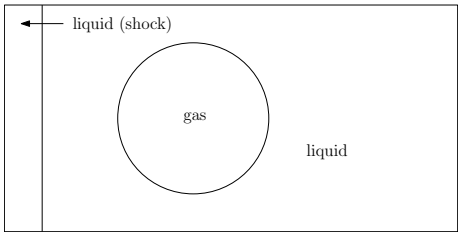


Figure 5: Liquid-gas interaction: description of the computational domain

	Density (kg m^{-3})	Pressure (Pa)	Velocity (m s^{-1})	γ	π (Pa)
Liquid (shock)	1030.9	3×10^9	$(300., 0.)$	4.4	6.8×10^8
Liquid	1000.	10^5	$(0., 0.)$	4.4	6.8×10^8
Gas	1.	10^5	$(0., 0.)$	1.4	0

Table 1: Liquid-gas interaction. Initial conditions.

A 600×300 space grid is used for the numerical computation. Wall boundary conditions are imposed on the upper and lower sides, and input/output boundary conditions on the other sides. The Courant number is 0.5.

The mapping of the volume fraction is plotted at several times in Fig. 6 and Fig. 7. Results show good agreement between the two solvers. We also note that our results are similar to those presented in [7] or [9].



Figure 6: Liquid-gas interaction. Mapping of the volume fraction z after 125 μs and 600 μs for the five-point stencil scheme [3]



Figure 7: Liquid-gas interaction. Mapping of the volume fraction z after 125 μs and 600 μs for the nine-point stencil scheme [5].

Remark: With the five-point-stencil scheme, the use of an implicit scheme in the acoustic part reduces the computational time by 34.

6.2 Steady supersonic flow over a cylinder

A Mach 10 steady flow over a cylinder is considered. The radius of the cylinder is equal to 1. The computational domain is covered by a 100×200 structured grid. We note that this case is monophasic.

A comparison between the two solvers for the Lagrangian step defined in section 5 is done, see Fig. 8. With the five-point stencil scheme, carbuncle phenomenon appears near the stagnation point. The numerical solution is not symmetrical. We note that the shape of the shock instabilities is not similar to those presented in the literature [10, 11, 12]. With the genuinely two-dimensional scheme, the solution is perfectly acceptable. Similar results have been presented in [11], where a nine-point scheme eliminates shock instabilities.

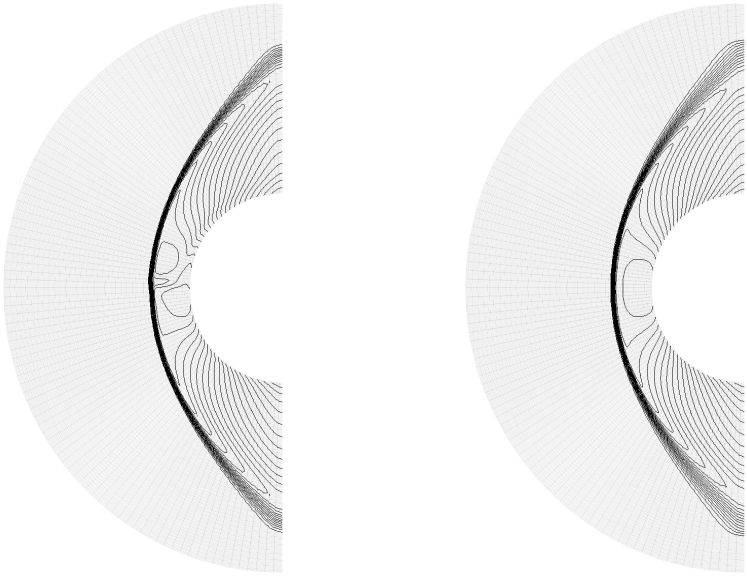


Figure 8: Mach 10 flow over a cylinder: density contours. Comparison between different solvers in the acoustic step. Left: five-point stencil from [3], Right: nine-point solver from [5].

Usually, such computations are done with direct Eulerian finite volume methods. In this case, a possible cure for the carbuncle phenomenon is to use of a dissipative scheme, like the HLL scheme [4]. Indeed these schemes are very robust for supersonic computation involving shocks but they are very dissipative in the boundary layer. Another approach is to use a combination of a dissipative flux function and a classical scheme [12].

In our splitting strategy, we used the HLL scheme for the acoustic step. One can see in Fig. 9 that carbuncle still appears.

Using the splitting strategy, carbuncle phenomenon seems to appear with all schemes based on a five-point stencil. However, the scheme based on a nine-point stencil in the Lagrangian part might offer a cure for carbuncle.

6.3 Interaction of supersonic gas flow with a liquid droplet

Finally, a unsteady supersonic two-phase flow test case is presented. We are interested in the interaction between a supersonic gas flow and a liquid droplet.

We consider a 5 mm$\times 3$ mm rectangular domain. A liquid droplet of radius $r = 0.25$ mm is initially located at the position $(1, 0)$ mm. The droplet is modelled by a stiffened gas with the coefficients $\gamma = 4.4$ and $\pi = 6.10^8$ Pa. The liquid is initially at rest, with a density of 1000 kg.m^{-3} at the atmospheric pressure. The gas flow is considered as a perfect gas with $\gamma = 1.4$. At the initial time, the gas density is equal to 1.293 kg.m^{-3}, $\mathbf{u} = (5000, 0)$ m.s^{-1}, and the pressure is 10^5 Pa.

The mapping of the fluid velocity magnitude for both solvers is plotted in Fig. 10. The position of the bow shock is the same with the two solvers. In Fig. 10, the black line represents the geometry of the liquid droplet. One can see that the location of droplet is the same but the shape is different.

The velocity contours are plotted in Fig. 11. One can see that carbuncle phenomenon also appears in this two-phase flow test case with the five-point stencil scheme. It seems to disturb the gas flow between the bow shock and the liquid droplet. With the genuinely two-dimensional

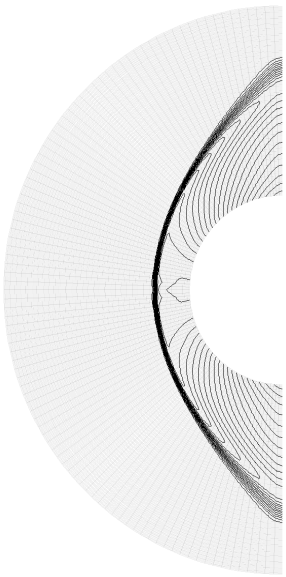


Figure 9: Mach 10 flow over a cylinder: density contours with HLL solver.

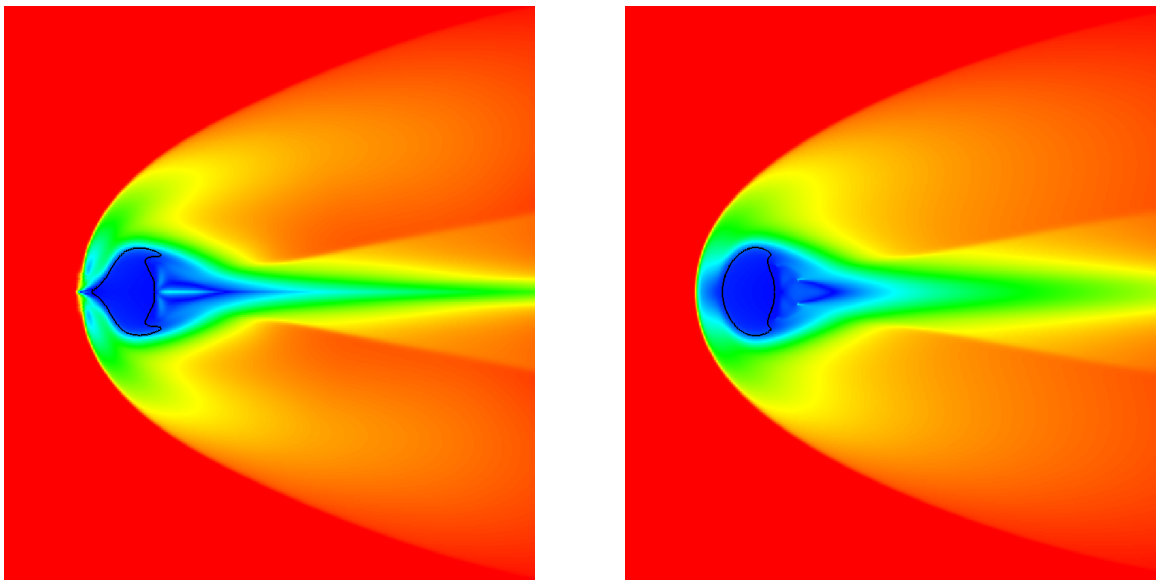


Figure 10: Interaction of supersonic gas flow with a liquid droplet. Mapping of the fluid velocity magnitude and contour of the liquid droplet (in black) after 1.57 μs. Left: five-point stencil from [3], Right: nine-point solver from [5].

scheme, there are no instabilities between the bow shock and the droplet. The scheme based on a nine-point stencil did not allow the carbuncle to appear even in this diphasic test case.

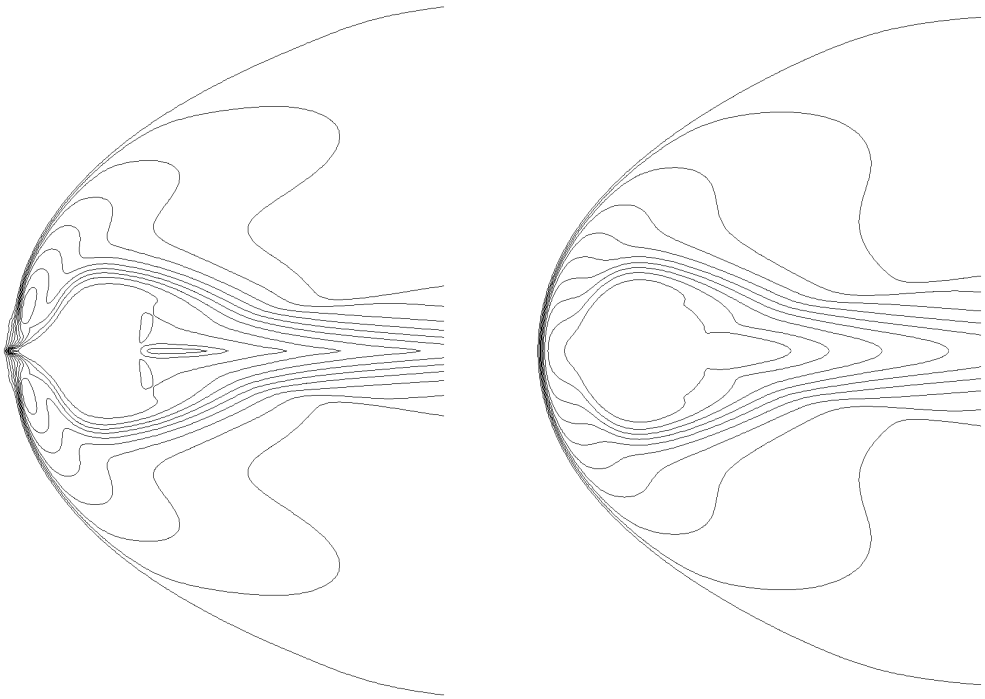


Figure 11: Interaction of supersonic gas flow with a liquid droplet. Velocity contour after 1.57 μs. Left: five-point stencil from [3], Right: nine-point solver from [5].

7 CONCLUSIONS

In this present work, an extension of the operator splitting strategy [2, 8] to two-phase flows using the five-equation model [1] have been presented. This strategy decouples acoustic and transport phenomena. Several schemes based on five-point or nine-point stencils are used in the acoustic step. In non-supersonic numerical simulations, the different schemes produce the same results. On supersonic velocity test case, carbuncle phenomenon appears with the five-point scheme. Using a very dissipative scheme like HLL in the Lagrangian part of the splitting does not prevent the apparition of the instabilities. However, carbuncle does not appear when the acoustic step is solved with a genuinely two-dimensional scheme based on a nine-point stencil.

In the future, the implicit scheme for the genuinely two-dimensional scheme based on a nine-point stencil has to be developed. The extension to Navier-Stokes equations is already in progress. Finally, the extension of the genuinely two-dimensional scheme based on an approximate Riemann solver located at the node in direct Eulerian approach will be considered.

REFERENCES

[1] G. Allaire, S. Clerc, S. Kokh, A Five-Equation Model for the Simulation of Interfaces between Compressible Fluids. *Journal of Computational Physics*, **181**, 577–616, 2002.

[2] C. Chalons, M. Girardin, S. Kokh, An all-regime Lagrange-Projection like scheme for gas dynamics equations on unstructured meshes. *submitted*, 2014.

[3] G. Gallice, Positive and Entropy Stable Godunov-type Schemes for Gas Dynamics and MHD Equations in Lagrangian or Eulerian Coordinates. *Numerische Mathematik*, **94**, 673–713, 2003.

[4] A. Harten, P.D. Lax, B. van Leer, On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM review*, **25**, 35–61, 1983

[5] P-H. Maire, A high-order cell-centered Lagrangian scheme for two-dimensional compressible fluid flows on unstructured meshes. *Journal of Computational Physics*, **228**, 2391–2425, 2009.

[6] MR. Baer, JW. Nunziato, A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. *International journal of multiphase flow*, **12**, 861–889, 1986.

[7] R. Saurel, R. Abgrall, A simple method for compressible multifluid flows. *SIAM J. Sci. Comput.*, **21**(3), 1115–1145, 1999.

[8] C. Chalons, M. Girardin, S. Kokh, An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes. *submitted*, 2014.

[9] S. Kokh, F. Lagoutière, An anti-diffusive numerical for the simulation of interfaces between compressible fluids by means of five-equation model. *Journal of Computational Physics*, **229**, 2773–2809, 2010.

[10] Y. Chauvat, J.-M. Moschetta and J. Gressier, Shock wave numerical structure and the carbuncle phenomenon. *International Journal for Numerical Methods in Fluids*, **47**, 903–909, 2005

[11] Z. Shen, W. Yan et G. Yuan, A robust and contact resolving Riemann solver on unstructured mesh, Part I, Euler method. *Journal of Computational Physics*, **268**, 432–455, 2014.

[12] H. Nishikawa, K. Kitamura, Very simple, carbuncle-free, boundary-layer-resolving, rotated-hybrid Riemann solvers. *Journal of Computational Physics*, **227**, 2560–2581, 2008.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

LAGRANGE-FLUX EULERIAN SCHEMES FOR COMPRESSIBLE MULTIMATERIAL FLOWS

Florian De VUYST[2], Thibault GASC[1,2,3], Renaud MOTTE[3], Mathieu PEYBERNES[4] and Raphaël PONCET[5]

[1] CMLA, ENS Cachan Université Paris-Saclay, CNRS,
94235 Cachan, France
e-mail: devuyst@cmla.ens-cachan.fr

[2]Maison de la Simulation USR 3441
CEA Saclay, F-91191 Gif-sur-Yvette
e-mail: thibault.gasc@cea.fr

[3]CEA DAM DIF
F-91297 Arpajon

[4] CEA, DEN, DM2S, STMF
CEA Saclay, F-91191 Gif-sur-Yvette
e-mail: mathieu.peybernes@cea.fr

[5] CGG
27 Avenue Carnot, F-91300 Massy
e-mail: raphael.poncet@cgg.com

Keywords: Compressible fluid, Lagrangian remapping, Lagrange-flux scheme, multimaterial flow, performance model

Abstract. *This paper is devoted to the redesign of the Lagrangian remapping process for compressible Hydrodynamics in order to achieve better computational performance. As an unintended outcome, the analysis has lead us to the discovery of a new family of solvers – the so-called Lagrange-flux schemes – that appear to be very promising for the CFD community. Results obtained with the Lagrange-flux approach are comparable with those obtained with the "legacy" Lagrange-remap solver using staggered variables, but provide a far better scalable computational performance, as proven using performance modeling. We also extend Lagrange-flux scheme to the case of multimaterial flows involving interface capturing advection schemes.*

1 Motivation and introduction

For complex compressible flows involving multiphysics phenomenons like e.g. high-speed elastoplasticity, multimaterial interaction, plasma, gas-particles etc., a Lagrangian description of the flow is generally preferred for easier physical coupling. To ensure robustness, some spatial remapping on a regular mesh may be added. A particular case is the family of the so-called Lagrange+remap schemes [2], also referred to as Lagrangian remapping that apply a remap step on a reference (say Eulerian) mesh after each Lagrangian time advance step. Acknowledged legacy codes implementing remapped Lagrange solvers usually define thermodynamical variables at cell centers and velocity variables at mesh nodes (see figure 1). In Poncet et al. [1],

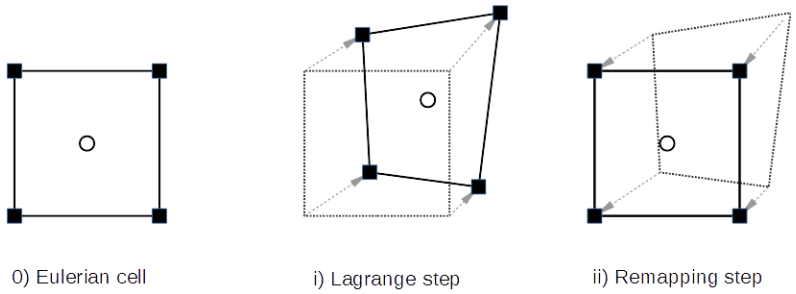


Figure 1: "Legacy" staggered Lagrange-remap scheme: thermodynamical variables are located at cell centers (circles) whereas velocity variables are located at cell nodes (squares).

we have achieved a node-based performance analysis of a reference legacy Lagrange-remap hydrodynamics solver used in industry. By analyzing each kernel of the whole algorithm, using roofline-type models [3] on one side and refined Execution Cache Memory (ECM) models [4], [5] on the other side, we have been able not only to quantitatively predict the performance of the whole algorithm — with relative errors in the single digit range — but also to identify a set of features that limit the whole global performance. This can be roughly summarized into three features:

1. Staggered velocity variables involve a rather big amount of communication to/from CPU caches and memory with low arithmetic intensity, thus lowering the whole performance;

2. Alternating direction (AD) strategies (see the appendix in [6]) or more specifically AD remapping procedures also generate too much communication with a loss of CPU occupancy and a rather poor multicore scalability.

3. For multimaterial hydrodynamics using VOF-based interface reconstruction methods, there is a strong loss of performance due to some array indirections and noncoalescent data in memory. Vectorization of such algorithms is also not trivial.

From these observations and as a result of the analysis, we decided to "rethink" Lagrange-remap schemes, with possibly modifying some aspects of the solver in order to improve node-based performance of the hydrocode solver. We have searched for alternative formulations that lower communication and improve both arithmetic intensity and SIMD property of the algorithm.

In this paper, we describe the process of redesign of Lagrange-remap schemes leading to higher performance solvers. Actually, this redesign methodology has also givn us ideas of

innovative Eulerian solvers. The so-called *Lagrange-flux schemes* appear to be very promising in the Computational Fluid Dynamics community.

The paper is organized as follows. In section 2, we first formulate the requirements for the design of higher-performance Lagrange-remap schemes. In section 3 we give a description of the Lagrange step and formulate it under a Finite Volume form. In section 4 we focus on the remap step which is reformulated as a finite volume scheme with pure convective fluxes. This interpretation is applied in section 5 to build the so-called Lagrange-flux schemes. We also discuss the important issue of achieving second order accuracy (in both space and time). In section 7 we deal with the extension to multimaterial flow with the use of low-diffusive and accurate interface-capturing methods.

2 Requirements

Starting from a "legacy" staggered Lagrange-remap solver and related observed performance measurements, we want to improve the performance by modifying the computational approach under the following constraints and requirements:

1. A Lagrangian solver must be used (for multiphysics coupling issue).
2. To reduce communication, we prefer to use collocated cell-centered variables rather than a staggered scheme.
3. To reduce communication, we prefer use a direct multidimensional remap solver rather than splitted alternating direction AD projections.
4. The method should be simply extended to second-order accuracy (in both space and time).
5. The solver must be able to be naturally extended to multimaterial flows.

Before going further, let us first comment the above requirements. The second requirement should imply the use of a cell-centered Lagrange solver. Fairly recently, Després and Mazeran in [7] and Maire and et al. [8] (with high-order extension in [9]) have proposed pure cell-centered Lagrangian solvers based on the reconstruction of nodal velocities. In our study, we will examine if it is possible to use approximate and simpler Lagrangian solvers in the Lagrange+remap context, in particular for the sake of performance. The fourth assertion requires a full multidimensional remapping step, probably taking into account geometric elements (deformation of cells and edges) if we want to ensure high-order accuracy remapping. We have to find a good trade-off between simplifications-approximations and accuracy (or properties) of the numerical solver.

3 Lagrangian time advance step

As example, let us consider the compressible Euler equations for two-dimensional problems. Denoting ρ, $\boldsymbol{u} = (u_i)_i$, $i \in \{1,2\}$, p and E the density, velocity, pressure and specific total energy respectively, the mass, momentum and energy conservation equations are

$$\partial_t U_\ell + \nabla \cdot (\boldsymbol{u}\, U_\ell) + \nabla \cdot \boldsymbol{\pi}_\ell = 0, \quad \ell = 1, \ldots, 4, \tag{1}$$

where $U = (\rho, (\rho u_i)_i, \rho E)$, $\boldsymbol{\pi}_1 = \vec{0}$, $\boldsymbol{\pi}_2 = (p, 0)^T$, $\boldsymbol{\pi}_3 = (0, p)^T$ and $\boldsymbol{\pi}_4 = p\boldsymbol{u}$. For the sake of simplicity, we will use a perfect gas equation of state $p = (\gamma - 1)\rho(E - \frac{1}{2}|\boldsymbol{u}|^2)$, $\gamma \in (1, 3]$. The speed of sound c is given by $c = \sqrt{\gamma p/\rho}$.

For any volume V_t that is advected by the fluid, from the Reynolds transport theorem we have

$$\frac{d}{dt}\int_{\mathscr{V}_t} U_\ell \, d\boldsymbol{x} = \int_{\partial\mathscr{V}_t} \{\partial_t U_\ell + \nabla\cdot(\boldsymbol{u}\, U_\ell)\}\, d\boldsymbol{x} = -\int_{\partial\mathscr{V}_t} \boldsymbol{\pi}_\ell\cdot\boldsymbol{\nu}\, d\sigma$$

where $\boldsymbol{\nu}$ is the normal unit vector exterior to $\mathscr{V}_t$. This leads to a natural explicit finite volume scheme in the form

$$|K^{n+1,L}|(U_\ell)_K^{n+1,L} = |K|(U_\ell)_K^n - \Delta t^n \sum_{A^{n+\frac{1}{2},L}\subset\partial K^{n+\frac{1}{2},L}} |A^{n+\frac{1}{2},L}|\, \boldsymbol{\pi}_A^{n+\frac{1}{2},L}\cdot\nu_A^{n+\frac{1}{2},L}. \tag{2}$$

In expression (2), the superscript "L" indicates the Lagrange evolution of the quantity. Any Eulerian cell K is deformed into the Lagrangian volume $K^{n+\frac{1}{2},L}$ at time $t^{n+\frac{1}{2}}$, and into the Lagrangian volume $K^{n+1,L}$ at time t^{n+1}. The pressure flux terms through the edges $A^{n+1/2,L}$ are evaluated at time $t^{n+\frac{1}{2}}$ in order to get second-order accuracy in time. Of course, that means that we need a predictor step for the velocity field $\boldsymbol{u}^{n+\frac{1}{2},L}$ at time $t^{n+\frac{1}{2}}$ (not written here for simplicity).

Notations. From now on, we will use the simplified notation $\boldsymbol{v}^{n+\frac{1}{2}} = \boldsymbol{u}^{n+\frac{1}{2},L}$.

4 Rethinking the remap step

The remapping step consists in projecting the fields U_ℓ defined at cell centers $K^{n+1,L}$ onto the initial (reference) Eulerian mesh with cells K. Starting from an interpolated vector-valued field $\mathscr{I}^{n+1,L}U^{n+1,L}$, we project the field on piecewise-constant function on the Eulerian mesh, according to the integral formula

$$U_K^{n+1} = \frac{1}{|K|}\int_K \mathscr{I}^{n+1,L}U^{n+1,L}(\boldsymbol{x})\, d\boldsymbol{x}. \tag{3}$$

Practically, they are many ways to deal with the projection operation (3). One can assemble elementary projection contributions by computing the volume intersections between the reference mesh and the deformed mesh. But this procedure requires the computation of all the geometrical elements. Moreover, the projection needs local tests of projection with conditional branching (think about the very different cases of compression, expansion, pure translation, etc). Thus the procedure is not SIMD and with potentially a loss of performance. The incremental remapping can also interpreted as a transport/advection process, as already emphasized by Dukowicz and Baumgardner [10].

Let us now write a different original formulation of the remapping process. In this step, there is no time evolution of any quantity, and in some sense we have $\partial_t U = 0$, that we rewrite

$$\partial_t U = \partial_t U + \nabla\cdot(-\boldsymbol{v}^{n+\frac{1}{2}}U) \;+\; \nabla\cdot(\boldsymbol{v}^{n+\frac{1}{2}}U) = 0.$$

We decide to split up this equation into two substeps, a backward convection and a forward one:

i) Backward convection:

$$\partial_t U + \nabla\cdot(-\boldsymbol{v}^{n+\frac{1}{2}}U) = 0. \tag{4}$$

ii) Forward convection:

$$\partial_t U + \nabla\cdot(\boldsymbol{v}^{n+\frac{1}{2}}U) = 0. \tag{5}$$

Each convection problem is well-posed on the time interval $[0, \Delta t^n]$ under a standard CFL condition. Let us now focus into these two steps and the way to solve them.

4.1 Backward convection in Lagrangian description

After the Lagrange step, if we solve the backward convection problem (4) over a time interval Δt^n using a Lagrangian description, we have

$$|K|(U_\ell)_K^{n,\star} = |K^{n+1,L}|(U_\ell)_K^{n+1,L}. \tag{6}$$

Actually, from the cell $K^{n+1,L}$ we go back to the original cell K with conservation of the conservative quantities. For $\ell = 1$ (conservation of mass), we have

$$|K|\,\rho_K^{n,\star} = |K^{n+1,L}|\,\rho_K^{n+1,L}$$

showing the variation of density by volume variation. For $\ell = 2, 3, 4$, it is easy to see that both velocity and specific total energy are kept unchanged is this step:

$$\boldsymbol{u}^{n,\star} = \boldsymbol{u}^{n+1,L}, \quad E^{n,\star} = E^{n+1,L}.$$

Thus, this step is clearly computationally inexpensive.

4.2 Forward convection in Eulerian description

From the discrete field $(U_K^{n,\star})_K$ defined on the Eulerian cells K, we then solve the forward convection problem over a time step Δt^n under an Eulerian description. A standard Finite Volume discretization of the problem will lead to the classical time advance scheme

$$U_K^{n+1} = U_K^{n,\star} - \frac{\Delta t^n}{|K|} \sum_{A \subset \partial_K} |A|\, U_A^{n+\frac{1}{2},\star}\, (\boldsymbol{v}_A^{n+\frac{1}{2}} \cdot \nu_A) \tag{7}$$

for some interface values $U_A^{n+\frac{1}{2},\star}$ defined from the local neighbor values $U_K^{n,\star}$. We finally get the expected Eulerian values U_K^{n+1} at time t^{n+1}.

Notice that from (6) and (7) we have also

$$|K|\, U_K^{n+1} = |K^{n+1,L}|\, U_K^{n+1,L} - \Delta t^n \sum_{A \subset \partial K} |A|\, U_A^{n+\frac{1}{2},\star}\, (\boldsymbol{v}_A^{n+\frac{1}{2}} \cdot \nu_A) \tag{8}$$

thus completely defining the remap step under the finite volume scheme form (8). Let us emphasize that we do not need any mesh intersection or geometric consideration to achieve the remapping process. The finite volume form (8) is now suitable for a straightforward vectorized SIMD treatment. From (8) it is easy to achieve second-order accuracy for the remapping step by usual finite volume tools (MUSCL reconstruction + second-order accurate time advance scheme for example).

4.3 Full Lagrange+remap time advance

Let us note that the Lagrange+remap scheme is actually a conservative finite volume scheme: putting (2) into (8) gives for all ℓ:

$$\begin{aligned} (U_\ell)_K^{n+1} = (U_\ell)_K^n \quad &- \quad \frac{\Delta t^n}{|K|} \sum_{A^{n+\frac{1}{2},L} \subset \partial K^{n+\frac{1}{2},L}} |A^{n+\frac{1}{2},L}|\, (\boldsymbol{\pi}_\ell)_A^{n+\frac{1}{2},L} \cdot \nu_A^{n+\frac{1}{2},L} \\ &- \quad \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A|\, (U_\ell)_A^{n+\frac{1}{2},\star}\, (\boldsymbol{v}_A^{n+\frac{1}{2}} \cdot \nu_A) \end{aligned} \tag{9}$$

that can also be written

$$(U_\ell)_K^{n+1} = (U_\ell)_K^n - \frac{\Delta t^n}{|K|} \sum_{A\subset\partial K} |A| \left(\frac{|A^{n+\frac{1}{2},L}|}{|A|} (\boldsymbol{\pi}_\ell)_A^{n+\frac{1}{2},L} \cdot \nu_A^{n+\frac{1}{2},L} \right) - \frac{\Delta t^n}{|K|} \sum_{A\subset\partial K} |A| \left((U_\ell)_A^{n+\frac{1}{2},\star} (\boldsymbol{v}_A^{n+\frac{1}{2}} \cdot \nu_A) \right). \quad (10)$$

We recognize into (10) pressure-related fluxes and convective numerical fluxes.

4.4 Comments

The finite volume formulation (10) is attractive and seems rather simple at first sight. But we should not forget that we have to compute a velocity Lagrange vector field $\boldsymbol{v}^{n+\frac{1}{2}} = \boldsymbol{u}^{n+\frac{1}{2},L}$ where the variables should be located at cell nodes to return a well-posed deformation. Moreover, expression (10) involves geometric elements like the length of the deformed edges $A^{n+\frac{1}{2},L}$. Among rigorous constructions of collocated Lagrangian solvers, let us mention the GLACE scheme by Després-Mazeran [7] and the cell-centered EUCCLHYD solver by Maire et al [8]. Both are rather computationally expensive and their second-order accurate extension is not easy to achieve.

Although it is possible to couple these Lagrangian solvers with the flux-balanced remapping formulation, it is also worth searching for simplified approximate Lagrange formulations. One of the difficulty in the analysis of Lagrange-remap schemes is that, in some sense, space and time are coupled by the deformation process.

Below, we derive a formulation that leads to a proper separation between space and time, in order to determine the order of accuracy in a simple way. The idea is to make the time step tend to zero in the Lagrange-remap scheme (method of lines [11]), then exhibit the instantaneous spatial numerical fluxes through the Eulerian cell edges that will serve for the construction of an explicit finite volume scheme. Because the method needs an approximate Riemann solver in Lagrangian form, we will call it a Lagrange-flux scheme.

5 Derivation of a simple second-order accurate Lagrange-flux scheme

From conclusions of the discussion above, we would like to be free from any "complex" collocated Lagrangian solver involving complex geometric elements. Another difficult point is to define the deformation velocity field $\boldsymbol{v}^{n+\frac{1}{2}}$ at time $t^{n+\frac{1}{2}}$, in an accurate manner.

In what follows, we are trying to deal with time accuracy in a different manner. Let us comme back to the Lagrange+remap formula (10). Let us consider a "small" time step $t > 0$ that fulfils the usual stability CFL condition. We have

$$(U_\ell)_K(t) = (U_\ell)_K^n - \frac{t}{|K|} \sum_{A\subset\partial K} |A| \left(\frac{|A^L(t/2)|}{|A|} (\boldsymbol{\pi}_\ell)_A^L(t/2) \cdot \nu_A^L(t/2) \right) - \frac{t}{|K|} \sum_{A\subset\partial K} |A| \, (U_\ell)_A^\star(t/2) \, \boldsymbol{v}_A(t/2) \cdot \nu_A.$$

By making t tend to zero, ($t > 0$), we have $A^L(t/2) \to A$, $(\boldsymbol{\pi}_\ell)^L(t/2) \to \boldsymbol{\pi}_\ell$, $\boldsymbol{v}(t/2) \to \boldsymbol{u}$, $(U_\ell)^\star \to U_\ell$, then we get a semi-discretization in space of the conservation laws. That can be seen as a method of lines ([11]):

$$\frac{d(U_\ell)_K}{dt} = -\frac{1}{|K|} \sum_{A\subset\partial K} |A| \, ((\boldsymbol{\pi}_\ell)_A \cdot \nu_A) - \frac{1}{|K|} \sum_{A\subset\partial K} |A| \, (U_\ell)_A \, (\boldsymbol{u}_A \cdot \nu_A). \quad (11)$$

We get a classical finite volume method in the form

$$\frac{dU_K}{dt} = -\frac{1}{|K|} \sum_{A \subset \partial K} |A| \, \Phi_A$$

with a numerical flux Φ_A whose components are

$$(\Phi_\ell)_A = (U_\ell)_A \, (\boldsymbol{u}_A \cdot \nu_A) + (\boldsymbol{\pi}_\ell)_A \cdot \nu_A. \tag{12}$$

In (11), pressure fluxes $(p_\ell)_A$ and interface velocities $\boldsymbol{u}_A$ can be computed from an approximate Riemann solver in Lagrangian coordinates (for example the Lagrangian HLL solver, see [12]). Then, the interface states $(U_\ell)_A$ can be computed from a upwind process according to the sign of the normal velocity $(\boldsymbol{u}_A \cdot \nu_A)$. To get higher-order accuracy in space, one can use a standard MUSCL reconstruction + slope limiting process. At this stage, because there is no time discretization, all acts on the Eulerian mesh and fluxes are defined at the edges the the Eulerian cells.

To get high-order accuracy in time, one can then apply a standard high-order time advance scheme (RK2, etc.). For the second-order Heun scheme for example, we have the following algorithm:

1. Compute the time step Δt^n subject to some CFL condition;
2. **Predictor step**. MUSCL reconstruction + slope limitation: from the discrete values U_K^n, compute a discrete gradient for each cell K.
3. Use a Lagrangian approximate Riemann solver to compute pressure fluxes $\boldsymbol{\pi}_A^n$ and interface velocities $\boldsymbol{u}_A^n$
4. Compute the upwind edge values $(U_\ell)_A^n$ according to the sign of $(\boldsymbol{u}_A^n \cdot \nu_A)$;
5. Compute the numerical flux Φ_A^n as defined in (12);
6. Compute the first order predicted states $U_K^{\star,n+1}$:
$$U_K^{\star,n+1} = U_K^n - \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \, \Phi_A^n$$
7. **Corrector step**. MUSCL reconstruction + slope limitation: from the discrete values $U_K^{\star,n+1}$, compute a discrete gradient for each cell K.
8. Use a Lagrangian approximate Riemann solver to compute pressure fluxes $\boldsymbol{\pi}_A^{\star,n+1}$ and interface velocities $\boldsymbol{u}_A^{\star,n+1}$
9. Compute the upwind edge values $(U_\ell)_A^{\star,n+1}$ according to the sign of $(\boldsymbol{u}_A^{\star,n+1} \cdot \nu_A)$;
10. Compute the numerical flux $\Phi_A^{\star,n+1}$ as defined in (12);
11. Compute the second-order accurate states U_K^{n+1} at time t^{n+1}:
$$U_K^{n+1} = U_K^n - \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \, \frac{\Phi_A^n + \Phi_A^{\star,n+1}}{2}.$$

One can appreciate the simplicity of the numerical solver.

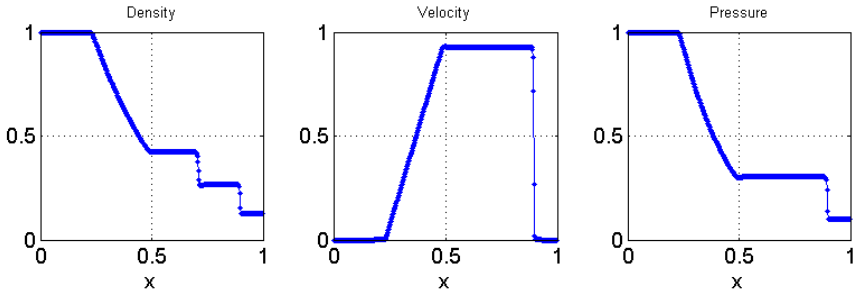


Figure 2: Second-order Lagrange-flux scheme on reference Sod's 1D shock tube problem. Time is $T = 0.23$, 384 mesh points. Here Sweby's slope limiter with coefficient 1.5 is used.

5.1 Details on the Lagrangian HLL approximate solver

A HLL approximate Riemann solver [12] in Lagrangian coordinates can be used to easily compute interface pressure and velocity. For a local Riemann problem made of a left state U_L and a right state U_R, the contact pressure $p^\star$ is given by the formula

$$p^\star = \frac{\rho_R p_L + \rho_L p_R}{\rho_L + \rho_R} - \frac{\rho_L \rho_R}{\rho_L + \rho_R} \max(c_l, c_R)\,(u_R - u_L), \tag{13}$$

and the normal contact velocity $u^\star$ by

$$u^\star = \frac{\rho_L u_L + \rho_R u_R}{\rho_L + \rho_R} - \frac{\rho_L \rho_R}{\rho_L + \rho_R} \frac{p_R - p_L}{\max(c_l, c_R)} \tag{14}$$

leading to very simple operations.

5.2 Numerical experiments

Shock tube problems. An as example, we test the Lagrange-flux scheme presented in section 5 on the reference one-dimensional Sod shock tube problem [13]. We use a Runge-Kutta 2 (RK2) time integrator and a MUSCL reconstruction using the second-order Sweby slope limiter [14]

$$\phi(a,b) = (ab > 0)\ \text{sign}(a)\ \max\big(\min(|a|, \beta|b|),\ \min(\beta|a|, |b|)\big)$$

with coefficient $\beta = 1.5$. We use a uniform grid made of 384 points. The final time is $T = 0.23$ and a CFL number equal to 0.25 . On figure 2, one can observe a good behaviour of the Eulerian solver, with rather sharp discontinuities and low numerical diffusion into rarefaction fans.

The second reference example is a case of two moving-away rarefaction fans under near-vacuum conditions (see Toro [12]). It is known that the Roe scheme breaks down for this case. The related Riemann problem is made of the left state $(\rho, u, p)_L = (1, -2, 0.4)$ and right state $(\rho, u, p)_R = (1, 2, 0.4)$. The final time of $T = 0.16$. We use 8192 mesh points to stress the behaviour on a refined mesh. Numerical results are given in figure 3. One can observe once again a nice behaviour of the method with a rather good accuracy.

Complex flow with high density ratios. In order to validate and compare both staggered Lagrange+remap and collocated Lagrange-flux, we propose the following original two-dimensional monofluid case. The computational domain is the rectangle $[0, L_x] \times [0, L_y]$ with $L_x = 3$ and $L_y = 2$. Boundary conditions are wall slip boundary conditions. The initial solution is a piecewise constant solution being constant into three different regions (see figure 4). Into

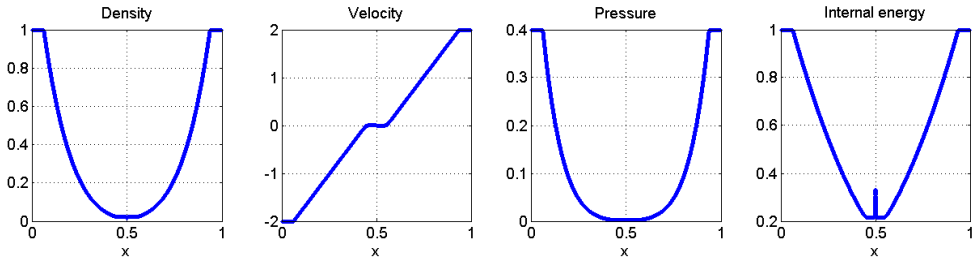


Figure 3: Second-order Lagrange-flux scheme on a double rarefaction near-vacuum case. Final time is $T = 0.16$, 8192 mesh points.

the region number 1 delimited by the rectangle $[0, 1] \times [1.1,\ 2]$, the solution is defined by $(\rho, u, v, p) = (1, 0, 0, 1)$. Into the region number 2 delimited by the rectangle $[0, 2] \times [0.9,\ 1.1]$, the solution is defined by $(\rho, u, v, p) = (100, 0, 0,\ 0.1)$. Into the last region number 3, we impose $(\rho, u, v, p) = (0.125, 0, 0,\ 0.1)$. A perfect gas law with $\gamma = 1.4$ is used. For spatial discretization, we use a rather fine Cartesian mesh made of 1200×400 points.

Into region nb 2, we have a dense cold fluid and a high density ratio between zone 2 and the two other zones. In region nb 1, we have a greater pressure that initiates a flow dynamics. The solution develops a complex structure made of several propagating shock waves, a low moving heavy fluid zone delimited by contact discontinuities, and a vortex of rarefied gas near the right border of the region 2. Numerical results between Lagrange-flux discretization versus staggered

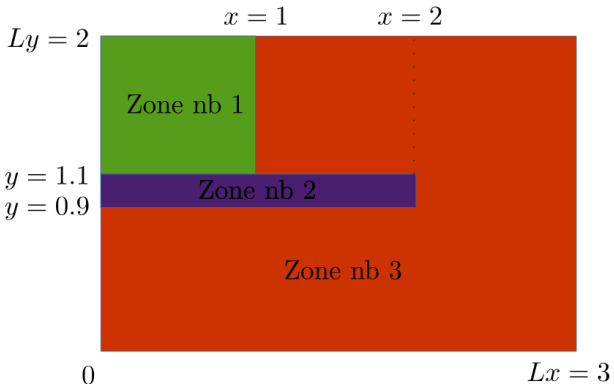
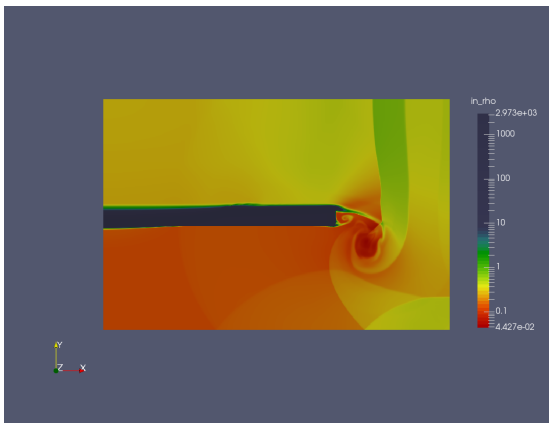


Figure 4: Geometry and initial conditions. All boundary conditions are wall slip conditions.

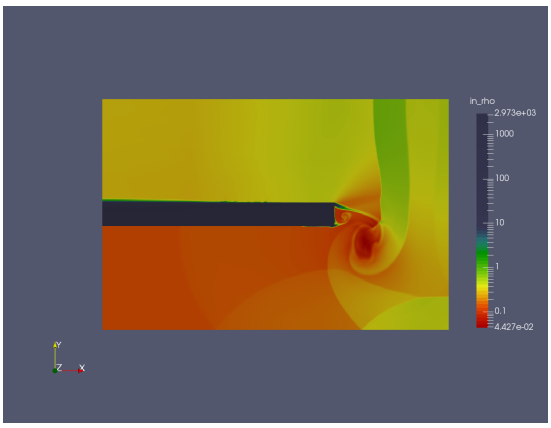
Lagrange+remap with splitted alternating directional remapping are shown in figure 5. One can observe that the density fields at snapshot time $t = 2$ sec are very comparable from the accuracy point of view. One can only notice a slight difference of capture of the contacts.

6 Performance results

In the "performance" companion paper [15], we propose a rigorous performance modeling approach to analyze the performance of both legacy Lagrange-remap solver and the Lagrange-flux scheme. It is indeed shown that performance results are far better using the Lagrange-flux scheme. The reformulated approach is also more suitable for SIMD and multi-core scalability. We refer to the companion paper [15] for interested readers.



(a) Lagrange-flux scheme, final time $t = 2$ sec. Density field, colormap in log scale.



(b) Legacy staggered Lagrange + splitted directional remapping, final time $t = 2$ sec. Density field, colormap in log scale.

Figure 5: Comparison of results between Lagrange-flux discretization vs staggered Lagrange+remap with splitted directional remapping.

7 Dealing with multimaterial flows

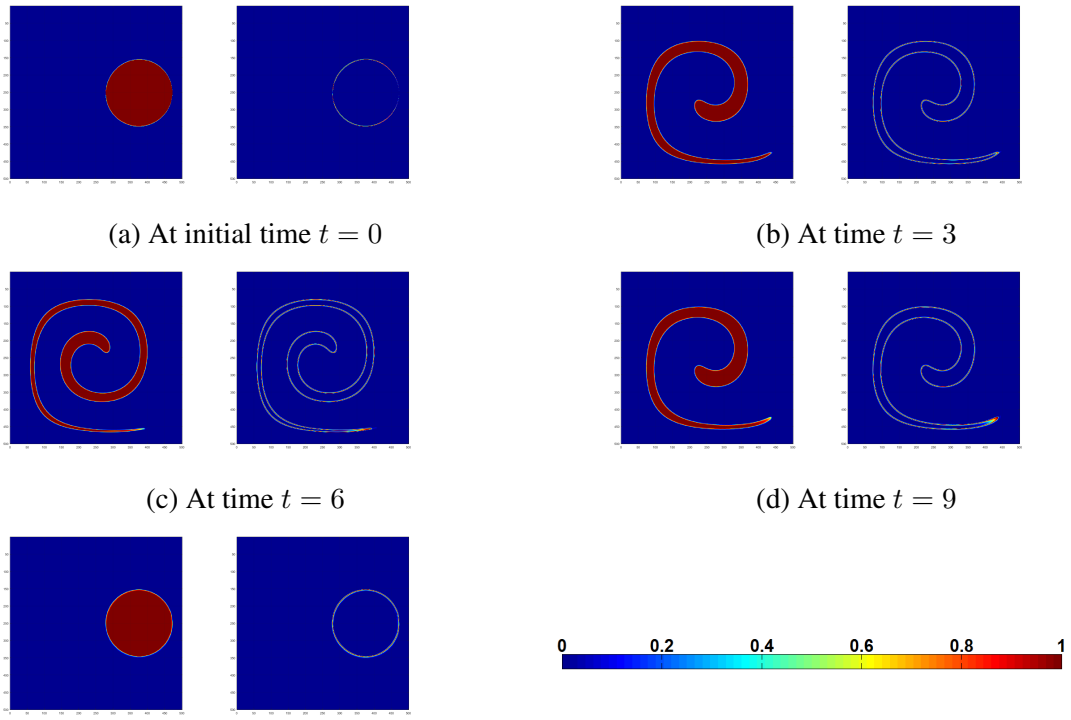
We would like to give an outline of the possible extension of Lagrange-flux schemes to compressible multimaterial/multifluid flows, i.e. flows that composed of different immiscible fluids and separated by free boundaries. Details of the multimaterial extension can be found into [18].

For pure Lagrange+remap schemes, usually VOF-based interface reconstruction (IR) algorithms are used (Young's PLIC, etc.). After the Lagrangian evolution, for the cells that host more than one fluid, fluid interfaces are reconstructed. During the remapping step, one has to evaluate the mass fluxes per material. From the computational point of view and computing performance, this process generally slows down the whole performance because of many array indirections in memory and specific treatment into mixed cells along with the material interfaces.

If the geometry of the Lagrangian cells is not completely known (as in the case of Lagrange-flux schemes), anyway we have to proceed differently. A possibility is to use interface capturing (IC) schemes, e.g. conservative Eulerian schemes that evaluate the convected mass fluxes through Eulerian cell edges. This can be achieved by the use of antidiffusive/low-diffusive advection solvers in the spirit of Després-Lagoutière's limited-downwind approach [16] or VoFire [17]. In a recent work [18], we have analysed the origin of known artefacts and numerical interface instabilities for this type of solvers and concluded that the reconstruction of fully multidimensional gradients with multidimensional gradient limiters was necessary. Thus, we decided to use low-diffusive advection schemes with a Multidimensional Limiting Process (MLP) in the spirit of [20]. The resulting method is quite accurate, shape-preserving and free from any artefact. We show some numerical results in the two next subsections. Let us emphasize that the interface capturing strategy perfectly fits with the Lagrange-flux flow description, and the resulting schemes are really suitable for vectorization (SIMD feature) with data coalescence into memory.

7.1 Interface capturing for pure advection problems

Let us first present numerical results on a pure scalar linear advection problem. The forward-backward advection case proposed by Rider and Kothe [21] is a hat-shaped function which is advected and stretched into a rotating vector field, leading to a filament structure. Then by applying the opposite velocity field, one have to retrieve the initial disk shape. In figure 6, we show the numerical solutions obtained on a grid 500^2 for both the scalar field (of variable $z \in [0, 1]$) and the quantity $z(1 - z)$ that indicates the spreads of the diffuse interface. One can conclude the very good behaviour of the method, providing both stability, accuracy and shape preservation.



(a) At initial time $t = 0$

(b) At time $t = 3$

(c) At time $t = 6$

(d) At time $t = 9$

(e) At final time $t = 12$

Figure 6: Validating the low-diffusive interface capturing scheme on the Kothe-Rider advection case, mesh 500^2.

7.2 Three-material hydrodynamics problem

We then consider our interface capturing method for multifluid hydrodynamics. Because material mass fractions are advected, that is

$$\partial_t y_k + \boldsymbol{u} \cdot \nabla y_k = 0, \quad k = 1, ..., N,$$

one can use the advection solver of these variables but we prefer a conservative form of the equations

$$\partial_t(\rho y_k) + \nabla \cdot (\rho y_k \boldsymbol{u}) = 0$$

in order to enforce mass conservation (see also [22]). The multimaterial Lagrange-flux scheme is tested on the reference "triple point" test case, found e.g. in Loubère et al. [23]. This problem is a three-state two-material 2D Riemann problem in a rectangular vessel. The simulation

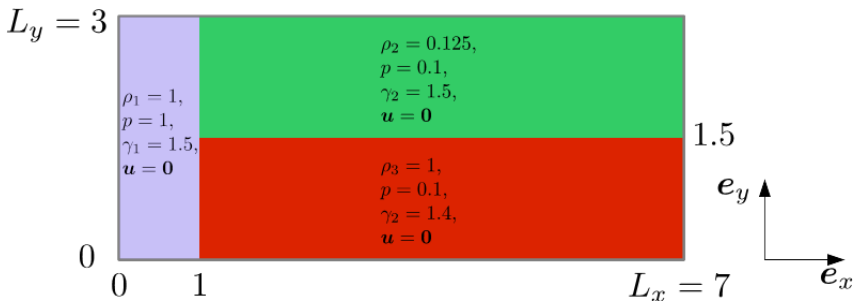
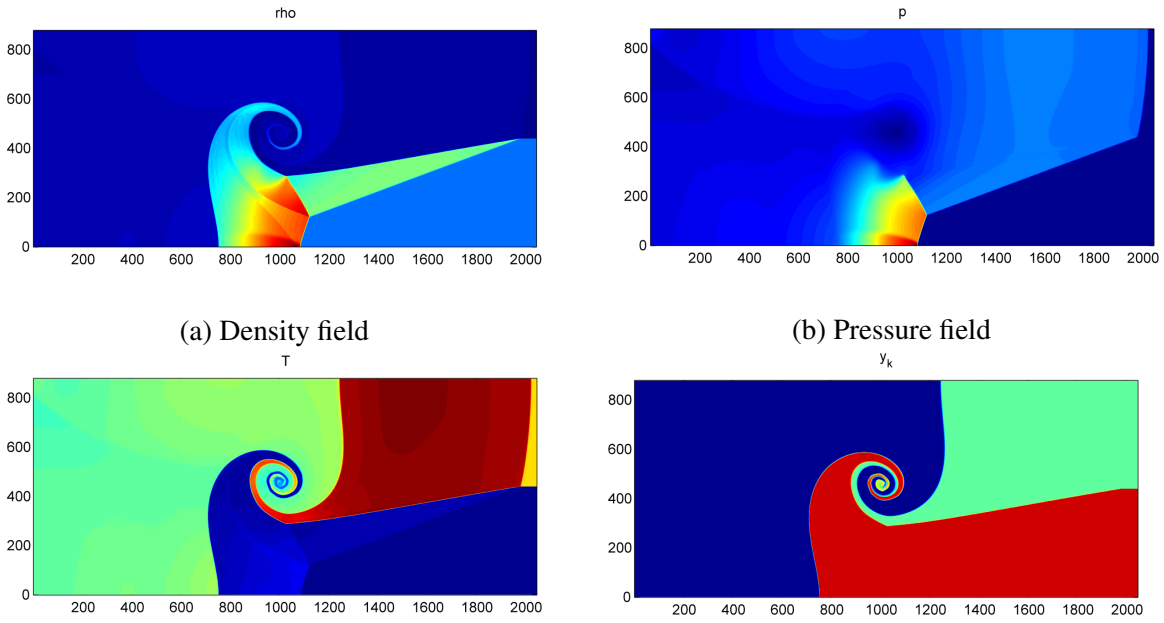


Figure 7: Geometry and initial configuration for the reference triple-point case.



(a) Density field

(b) Pressure field

(c) Temperature field

(d) Colored representation of material indicators

Figure 8: Results on the multimaterial "triple point" case (perfect gases) using a collocated Lagrange+remap solver + low-diffusive interface capturing advection scheme, mesh made of 2048×878 points. Final time is $T = 3.3530$.

domain is $\Omega = (0, 7) \times (0, 3)$ as described in figure 7. The domain is splitted up into three regions Ω_i, $i = 1, 2, 3$ filled with two perfect gases leading to a two-material problem. Perfect gas equations of state are used with $\gamma_1 = \gamma_3 = 1.5$ and $\gamma_2 = 1.4$. Due to the density differences, two shocks in sub-domains Ω_2 and Ω_3 propagate with different speeds. This create a shear along the initial contact discontinuity and the formation of a vorticity. Capturing the vorticity is of course the difficult part to compute. We use a rather fine mesh made of 2048×878 points (about 1.8M cells). One figure 8, we plot the density, pressure, temperature fields respectively and indicate the location of the three material zones. On can observe a nice capture of both shocks and contact discontinuities. The vortex is also captured accurately.

8 Concluding remarks and perspectives

This paper is firstly focused on the redesign of Lagrange-remap Hydrodynamics solvers in order to achieve better HPC node-based performance. We have reformulated the remapping step under a finite volume flux balance, allowing for a full SIMD algorithm. As an unin-

tended outcome, the analysis has lead us to the discovery of a new promising family of Eulerian solvers – the so-called Lagrange-flux solvers – that show simplicity, accuracy, and flexibility with a proven higher performance capability (see the "performance" ECCOMAS companion paper [15]). Interface capturing methods can be easily plugged-in for solving multimaterial flow problems.

REFERENCES

[1] R. Poncet, M. Peybernes, T. Gasc and F. De Vuyst, Performance modeling of a compressible hydrodnamics solver on multicore CPUs, Proceedings of the int. conf. on parallel computing PARCO2015, Edinburgh, 2015 (in press).

[2] C.W. Hirt, A.A. Amsden and J.L. Cook, An arbitrary Lagrangian-Eulerian computing method for all flow speeds. Journal of Computational Physics, 14:227–253 (1974).

[3] S. Williams, A. Waterman, and D. Patterson: Roofline: An Insightful Visual Performance Model for Multicore Architectures, Commun. ACM, 52, pp 6576 (2009).

[4] J. Treibig and G. Hager, Introducing a Performance Model for Bandwidth-Limited Loop Kernels. Proceedings of the Workshop Memory issues on Multi- and Manycore Platforms at PPAM 2009, Lecture Notes in Computer Science, 6067, pp. 615–624 (2010).

[5] H. Stengel, J. Treibig, G. Hager and G. Wellein, Quantifying performance bottlenecks of stencil computations using the Execution-Cache-Memory model. Proc. ICS15, the 29th Int. Conf. on Supercomputing, 2015, DOI: 10.1145/2751205.2751240.

[6] P. Colella and P.R. Woodward, The numerical simulation of two-dimensional fluid flow with strong shocks, J. Comput. Phys.,54:115–173 (1984).

[7] B. Després and C. Mazeran, Lagrangian gas dynamics in two dimensions and Lagrangian systems, Arch. Rational Mech. Anal. 178 (2005) 327–372.

[8] P.-H. Maire, R. Abgrall, J. Breil and J. Ovadia, A cell-centered Lagrangian scheme for compressible flow problems, SIAM J. Sci. Comput. 29 (4) (2007) 1781–1824.

[9] P.-H. Maire, A high-order cell-centered Lagrangian scheme for two-dimensional compressible fluid flows on unstructured meshes, J. Comput. Phys. 228 (2009) 2391–2425.

[10] J. K. Dukowicz and J. R. Baumgardner, Incremental remapping as a transport/advection algorithm. J. Comput. Phys., 160, 318335 (2000).

[11] W. E. Schiesser, The Numerical Method of Lines, Academic Press, ISBN 0-12-624130-9 (1991).

[12] E.F. Toro, Riemann solvers and numerical methods for fluid dynamics, 3rd Edition, Springer (2010).

[13] G.A. Sod, A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws" . J. Comput. Phys. 27: 1–31 (1971).

[14] P.K. Sweby, High resolution schemes using flux-limiters for hyperbolic conservation laws, SIAM J. Num. Anal. 21 (5): 995–1011 (1984).

[15] T. Gasc, F. De Vuyst, M. Peybernes, R. Poncet, R. Motte, Building a more efficient Lagrange-remap scheme thanks to performance modeling, this conference, Paper P12210, in Minisymposium "MS 414 - New trends in numerical methods for multi-material compressible fluid flows", Proc. of the Conference ECCOMAS 2016, submitted.

[16] B. Després and F. Lagoutière, Contact discontinuity capturing schemes for linear advection and compressible gas dynamics, J. Sci. Comp., 16(4), 479–524 (2001).

[17] B. Després, F. Lagoutière, E. Labourasse and I. Marmajou, An antidissipative transport scheme on unstructured meshes for multicomponent flows, IJFV, 30–65 (2010).

[18] F. De Vuyst, M. Béchereau, T. Gasc, R. Motte, M. Peybernes and R. Poncet, Stable and accurate low-diffusive interface capturing advection schemes, submitted to IJNMF, Special Issue of the MULTIMAT 2015 Conf. Würsburg, of the (2016).

[19] F. De Vuyst, T. Gasc, R. Motte, M. Peybernes and R. Poncet, Lagrange-flux schemes: reformulating second-order accurate Lagrange-remap schemes for better node-based HPC performance, submitted to Oil & Gas Science and Technologies OGST (2016).

[20] J.S. Park and C. Kim, Multi-dimensional Limiting Process for Discontinuous Galerkin Methods on Unstructured Grids, chapter in Computational Fluid Dynamics 2010, Springer, 179–184 (2011).

[21] A.J. Rider and D.B. Kothe, Reconstructing volume tracking, J. Comput. Phys., 141(2), 112–152 (1998).

[22] A. Bernard-Champmartin and F. De Vuyst, A low diffusive Lagrange-remap scheme for the simulation of violent air-water free-surface flows, J. of Comput. Physics, 274, 19–49 (2014).

[23] R. Loubère, P.-H. Maire, M. Shashkov, J. Breil and S. Galera, ReALE: A reconnection-based arbitrary-Lagrangian-Eulerian method, J. Comp. Phys., 229, 4724–4761 (2010).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ISOTROPY–PRESERVING SLOPE LIMITERS FOR FINITE VOLUME METHODS ON SQUARE MESHES

J. Velechovsky[1] **and M. M. Francois**[1]

[1]Los Alamos National Laboratory
Los Alamos, NM 87544, USA
e-mail: {jan,mmfran}@lanl.gov

Keywords: Slope limiter, isotropic numerical scheme, symmetry preserving numerical scheme, finite volume method, directional split, MUSCL–Hancock.

Abstract. *Directional split is a common technique used in Finite–Volume Methods (FVM) for solving multidimensional hyperbolic partial differential equations on regular grids. The directional split method is based on independent one–dimensional flux calculation for each Cartesian direction. Second order accuracy is achieved by doing a piecewise linear calculation of cell quantities in a Monotonic Upstream–Centered Scheme for Conservation Laws (MUSCL) scheme. This one–dimensional reconstruction results in artificial grid–orientation dependence of the numerical solution, i.e. some anisotropy are introduced even for originally isotropic problems. To mitigate the anisotropy of these schemes we propose to use extended two–dimensional stencils during the reconstruction step, while keeping the directional–split nature of the scheme.*

In this work, we introduce a novel multi–dimensional modification of the minmod slope limiter in the context of a MUSCL Hancock directional split FVM numerical scheme. Local direction of the unlimited gradient is utilized in the construction of our limiter. Improvements in the isotropy of the numerical solution are demonstrated for the linear advection as well as for standard hydrodynamics test cases on square meshes.

1 INTRODUCTION

In Finite–Volume Methods (FVM), second–order accuracy is achieved by using piecewise linear reconstruction of the variable quantities [1]. Near discontinuities, the most important part of the reconstruction is to reduce (limit) calculated slopes. This limitation is typically enforced by a slope limiter. FVM schemes traditionally employ directional split on regular grids. Consequently, the piecewise linear reconstruction is typically constructed independently for each Cartesian direction. Such directional one–dimensional reconstruction has for effect to introduce artificial anisotropy error in the numerical solution. The goal of this paper is to analyze and mitigate this numerical anisotropy error.

For hydrodynamic problems with a smooth solution, isotropic finite–difference scheme such as in [2] can be employed. The scheme is free of the directional (anisotropic) error in error terms of the lowest order. To approximate the first derivative, these isotropic finite–differences are based on six–point stencil in two spatial dimensions contrary to two–point (one–dimensional) stencil for the standard central difference approximation. Further theoretical analysis and the derivation of an isotropic finite volume discretization was presented in [3].

For a discontinuous solution, necessary constraints to avoid a numerical oscillation during the application of slope limiters on square meshes were derived in [4]. These additional constraints for one–dimensional Total Variation Disminishing (TVD) limiters are the main part of this Multi–dimensional Limiting Process (MLP). On unstructured grids, this MLP procedure reduces [5] to the classical multi–dimensional Barth–Jespersen limiter [6] with an extended stencil. A multi–dimensional application of slope limiters on unstructured grids with particular interest to definition of maximum principle regions was reviewed in [7]. An alternative approach to direct reconstruction limiting using Vector Image Polygon for scalar variables was presented in [8].

In this paper, we examine the standard one–dimensional minmod limiter [9] for a piecewise linear reconstruction. On a set of numerical examples, we demonstrate the development of an artificial anisotropy, which is introduced by this limiter. We propose a novel multi–dimensional extension of the minmod limiter to mitigate this anisotropy. This is in contrast to MLP and another approaches, where the original minmod limiter satisfies all multi–dimensional requirements.

The paper is organized as follows. A review of a directional split MUSCL Hancock numerical discretization [1] of the Euler equations is provided in next section. The third section is devoted to the piecewise linear reconstruction and to the presentation of our novel multi–dimensional variations of the minmod limiter. These limiters are numerically investigated in section 4. At first, an analysis of convergence and symmetry preservation is provided for the linear advection. Performance of our multi–dimensional Principal limiter is further demonstrated for Sedov and Noh hydrodynamic test cases.

2 NUMERICAL DISCRETIZATION

In this work, we use a directional split MUSCL Hancock numerical scheme [1] to solve the Euler equations. The scheme is based on a sequential application of one–dimensional numerical method along main axes x and y. For every time–step Δt, we apply both updates in the x– and y–direction independently, whereas the order of the direction is alternated.

Here we provide a short review of the directional split MUSCL Hancock scheme employing a piece–wise linear reconstruction of primitive variables inside a computational domain. One–

dimensional piecewise linear reconstruction in each computational cell i has the form

$$\mathbf{W}^{\pm}{}_i = \mathbf{W}_i \pm \frac{\mathbf{\Delta}_i}{2}\,, \tag{1}$$

for $\mathbf{W} = (w, v, p, \rho)^T$ in the case of the two–dimensional Euler equations.

$$\mathbf{\Delta_i} = \mathbf{W}_i^x \Delta x\,, \qquad\qquad \mathbf{W}_i^x = (w_i^x, v_i^x, p_i^x, \rho_i^x)^T\,. \tag{2}$$

These slope approximations $\mathbf{W}_i^x$ are determined by a slope limiter. Traditionally, the slope limiter is calculated independently using one–dimensional information. Different piecewise linear reconstructions using slope limiters are reviewed in section 3.

The first part of the scheme is the predictor step to the time–level $n + 1/2$. In the x–sweep, the system

$$\mathbf{W}_t + \mathbb{A}\mathbf{W}_x = \mathbf{0} \qquad \mathbb{A} = \begin{pmatrix} w & 0 & 1/\rho & 0 \\ 0 & w & 0 & 0 \\ \rho c^2 & 0 & w & 0 \\ \rho & 0 & 0 & w \end{pmatrix} \qquad \mathbf{W} = \begin{pmatrix} w \\ v \\ p \\ \rho \end{pmatrix} \tag{3}$$

with the sound speed c is solved explicitly in each computational cell i for the face extrapolated values

$$\mathbf{W}^{\pm}{}_i^{\,n+1/2} = \mathbf{W}^{\pm}{}_i^{\,n} - \frac{1}{2}\frac{\Delta t}{\Delta x}\mathbb{A}(\mathbf{W}^{\pm}{}_i^{\,n})\mathbf{\Delta}_i^n\,. \tag{4}$$

For these face extrapolated values, we apply the HLL Riemann solver [10] with wave speed approximations

$$S_L = \min(u_L - c_L, u_R - c_R) \tag{5}$$

$$S_R = \max(u_L + c_L, u_R + c_R)\,. \tag{6}$$

3 TWO–DIMENSIONAL PIECEWISE LINEAR RECONSTRUCTION

The general form of a piecewise linear reconstruction (preserving a mean quantity) inside a computational cell i, j is

$$u(x, y) = u_{i,j} + u_{i,j}^x(x - x_{i,j}) + u_{i,j}^y(y - y_{i,j})\,, \tag{7}$$

where $(x_{i,j}, y_{i,j})$ is the cell center coordinate and u is the quantity. On a square computational mesh, this reconstruction can be calculated independently for each orthogonal direction x and y assuming the location of inter–cell fluxes at centers of mesh faces as indicated in Figure 1. For faces aligned with the x axis, we can write

$$u_{i,j+1/2,L} = u_{i,j} + u_{i,j}^x(x_{i,j} - x_{i,j}) + u_{i,j}^y(y_{i,j} + \frac{\Delta x}{2} - y_{i,j}) = u_{i,j} + u_{i,j}^y\frac{\Delta x}{2} \tag{8}$$

and the application of one–dimensional slope limiters seems to be appropriate.

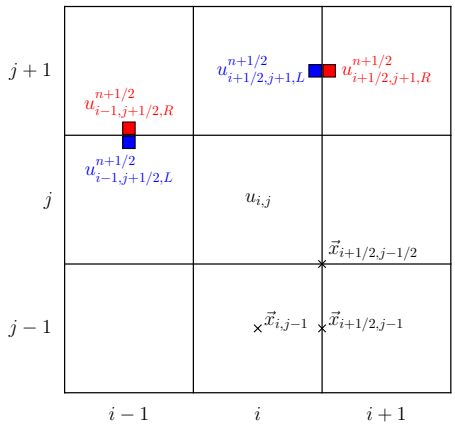
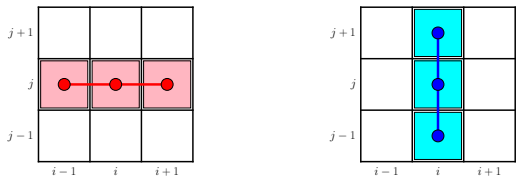


Figure 1: Computational mesh schematic. Cell coordinate and mean cell quantity $u_{i,j}$ shown in black. Left and right face–based reconstructed states shown in blue/red.



(a) (b)

Figure 2: Schematic representation of the MinMod reconstruction along split directions inside the central cell i, j. x–direction in (a) and y–direction in (b). Potential slopes indicated by "barbells". Effective cells emphasized in color.

3.1 One–dimensional mimmod limiter

The standard one–dimensional mimmod (1D MinMod) limiter for a piecewise linear reconstruction uses only the information along one direction. Here, we present the formulas only for x–direction. Assuming the mean value $u_i = u_{i,j}$ (for fixed j) in the cell i, we compute limited slopes as a combination of the left $u_i^{x,L} = \frac{u_i - u_{i-1}}{\Delta x}$ and the right $u_i^{x,R} = \frac{u_{i+1} - u_i}{\Delta x}$ finite difference approximations using the minmod function to get the final slope

$$u_i^{x,\text{minmod}} = \begin{cases} 0 & \text{if} \quad u_i^{x,L} u_i^{x,R} < 0 \\ \text{sign}(u_i^{x,L}) \min(|u_i^{x,L}|, |u_i^{x,R}|) & \text{else.} \end{cases} \tag{9}$$

Although 1D MinMod slopes are appropriate for second–order methods, we have observed that these slopes violate radial symmetry of radially symmetric problems, i.e. the anisotropy of a numerical solution is increased. Numerical examples of this artificial anisotropy are provided in section 4.

3.2 Two–dimensional limiters

The main motivation for genuinely multi–dimensional limiters is to reduce the anisotropy caused by the standard 1D MinMod limiter. At first, we derive an artificial limiter aligned with

a diagonal direction to show that most of the numerically observed anisotropy is caused by the limiter itself. Then, we construct novel limiters using different convex combination of these limiters.

3.2.1 Diagonal MinMod limiter

Here we construct a natural analogue of the 1D MinMod limiter rotated by 45 degrees. We start with the calculation of slopes along diagonals, i.e. along the lines

$$e(x,y) = \frac{x+y}{\sqrt{2}} \qquad\qquad f(x,y) = \frac{y-x}{\sqrt{2}}\,. \tag{10}$$

In this case, the natural choice of the left and right difference are

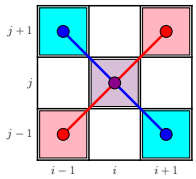


Figure 3: Schematic representation of the limited reconstruction inside the central cell i, j. Potential slopes indicated by "barbells". Effective cells emphasized in color.

$$u_{i,j}^{e,L} = \frac{u_{i,j} - u_{i-1,j-1}}{\sqrt{2}\Delta x} \qquad\qquad u_{i,j}^{f,L} = \frac{u_{i-1,j+1} - u_{i,j}}{\sqrt{2}\Delta x} \tag{11}$$

$$u_{i,j}^{e,R} = \frac{u_{i+1,j+1} - u_{i,j}}{\sqrt{2}\Delta x} \qquad\qquad u_{i,j}^{f,R} = \frac{u_{i,j} - u_{i+1,j-1}}{\sqrt{2}\Delta x}\,. \tag{12}$$

Having these differences, the minmod function (9) is applied to obtain limited diagonal slopes $u_{i,j}^{e}$ and $u_{i,j}^{f}$. Finally, Cartesian slopes are provided by the transformation

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial e}\frac{\partial e}{\partial x} + \frac{\partial u}{\partial f}\frac{\partial f}{\partial x} = \frac{1}{\sqrt{2}}\left(\frac{\partial u}{\partial e} + \frac{\partial u}{\partial f}\right) \tag{13}$$

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial e}\frac{\partial e}{\partial y} + \frac{\partial u}{\partial f}\frac{\partial f}{\partial y} = \frac{1}{\sqrt{2}}\left(\frac{\partial u}{\partial f} - \frac{\partial u}{\partial e}\right) \tag{14}$$

$$u_{i,j}^{x,\text{MinMod Diag}} = \frac{1}{\sqrt{2}}\left(u_{i,j}^{e} + u_{i,j}^{f}\right) \tag{15}$$

$$u_{i,j}^{y,\text{MinMod Diag}} = \frac{1}{\sqrt{2}}\left(u_{i,j}^{f} - u_{i,j}^{e}\right)\,. \tag{16}$$

3.2.2 Principal and Minor MinMod limiters

The Principal and the Minor MinMod limiters are constructed as a convex combination of the 1D MinMod limiter with the Diagonal MinMod limiter according to the local direction of the unlimited gradient inside each computational cell.

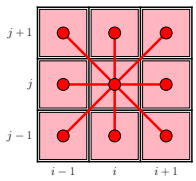


Figure 4: Schematic representation of the limited reconstruction inside the central cell i, j. Potential slopes indicated by "barbells". Effective cells emphasized in color.

Assuming that the direction of the unlimited gradient in a computational cell i, j is given by the polar angle $\Psi_{i,j}$, e.g. $\Psi_{i,j}$ = arctan($u_{i,j}^{y,\text{unlim}}/u_{i,j}^{x,\text{unlim}}$) for $x > 0$, we define

$$u_{i,j}^{x,\text{Principal}} = u_{i,j}^{x,\text{1D MinMod}} + \sin^2(2\Psi_{i,j}) \left(u_{i,j}^{x,\text{MinMod Diag}} - u_{i,j}^{x,\text{1D MinMod}} \right) \tag{17}$$

$$u_{i,j}^{x,\text{Minor}} = u_{i,j}^{x,\text{1D MinMod}} + \cos^2(2\Psi_{i,j}) \left(u_{i,j}^{x,\text{MinMod Diag}} - u_{i,j}^{x,\text{1D MinMod}} \right) \tag{18}$$

and analogically for the y direction.

The function $\sin^2(2\Psi_{i,j})$ reaches local maximum of one along directions of Cartesian axes. Therefore, the 1D MinMod limiter is preferred for the unlimited gradient aligned with Cartesian axes. The function $\cos^2(2\Psi_{i,j})$ in the Minor limiter is shifted by $\pi/4$. The Minor MinMod limiter represents less diffusive alternative to the Principal limiter.

4 NUMERICAL RESULTS

Here we present numerical results using the directional split numerical scheme from section 2 and the various limiters introduced in section 3. We call these results: 1D MinMod, Diagonal MinMod, Principal MinMod and Minor MinMod. Two more results are presented for comparison, namely results of an unlimited method using central differences (Unlimited) for a local gradient approximation and a low–order method with a piecewise constant reconstruction (Donor).

To compare the numerical results, we use the standard $L1$ error

$$E_{L1} = \sum_i \sum_j |u_{i,j}^{t_f} - u_{i,j}^e| \Delta x^2 \,,$$

where $u_{i,j}^{t_f} = u^{t_f}(x_{i,j}, y_{i,j})$ denotes numerical solution at the final time t_f and $u_{i,j}^e = u^e(x_{i,j}, y_{i,j})$ stands for the exact solution and Δx is the mesh spacing.

The $L1$ error is a global measure of a numerical error. It is not a representative measure of radial symmetry (isotropy) violation for the initially radial–symmetric problem. For discontinuous initial data, it is natural to expect some amount of numerical diffusion in the calculated data. In particular, there is no physical reason to introduce any anisotropy for isotropic initial data during linear advection. Standard limiters reduce this diffusion along diagonal directions with respect to square computational mesh. Whereas the $L1$ error is improved through this reduction, we prefer to obtain a mesh–alignment independent solution (even if it is more diffusive). To measure the radial symmetry violation, we replace the exact solution in the $L1$ error by the final numerical solution averaged over angular direction $u_{i,j}^{\text{mean}}$.

The symmetry violation is given by

$$E_s = \sum_i \sum_j |u_{i,j}^{t_f} - u_{i,j}^{\text{mean}}| \Delta x^2 \,. \tag{19}$$

All our test cases in the following are in two dimensions.

4.1 Smooth function advection

The initial smooth Gaussian function $u_{i,j}^0 = \exp(-(5r_{i,j})^2)$ is discretized on the computational domain $(-0.5, 0.5) \times (-0.5, 0.5)$. The advection speed is constant with the advection velocity $(a_1, a_2) = (1, 1)$ for the governing equation

$$\frac{\partial u}{\partial t} + a_1 \frac{\partial u}{\partial x} + a_2 \frac{\partial u}{\partial y} = 0 \,. \tag{20}$$

The $L1$ error of the numerical solution at final time $t_f = 1$ for four different mesh resolutions is presented in Table 1, where the "Donor" method stands for a piecewise constant reconstruction. All limiters show similar error and as well as the order of convergence, which is less than two, as indicated in Table 2.

Grid	Donor	Unlimited	1D MinMod	Diagonal MM	Principal MM	Minor MM
20^2	8.74e-02	9.32e-03	2.86e-02	3.67e-02	3.09e-02	2.93e-02
40^2	5.53e-02	1.58e-03	9.53e-03	1.25e-02	9.78e-03	8.27e-03
80^2	3.21e-02	3.15e-04	3.34e-03	4.32e-03	3.74e-03	2.41e-03
160^2	1.75e-02	7.50e-05	1.01e-03	1.38e-03	1.38e-03	6.85e-04

Table 1: L1 error for four mesh resolutions for the smooth Gaussian function advection.

Grid N/M	Donor	Unlimited	1D MinMod	Diagonal	Principal	Minor
$20^2/40^2$	0.7	2.6	1.6	1.6	1.7	1.8
$40^2/80^2$	0.8	2.3	1.5	1.5	1.4	1.8
$80^2/160^2$	0.9	2.1	1.7	1.6	1.4	1.8

Table 2: Numerical order of convergence, i.e. $\log_2(L1_N/L1_M)$, for the Gaussian function advection.

4.2 Circle advection

Grid	1D MinMod	Diagonal MM	Principal MM	Minor MM
20^2	1.03e-01	1.16e-01	1.11e-01	1.05e-01
40^2	6.62e-02	7.46e-02	7.28e-02	6.81e-02
80^2	4.29e-02	4.83e-02	4.70e-02	4.42e-02
160^2	2.78e-02	3.13e-02	3.04e-02	2.87e-02

Table 3: L1 error for four mesh resolutions for the circle advection.

The circle shape is represented by the initial function $u(r) = 1$ for $r < 0.25$ and $u(r) = 0$ elsewhere, discretized on the computational domain $(-0.5, 0.5)\times(-0.5, 0.5)$ with the advection velocity $(a_1, a_2) = (1, 1)$. The error at final time $t_f = 1$ in L1 norm is presented in Table 3. Initial and advected shapes on a coarse grid are plotted in Figure 5. We can see the artificial

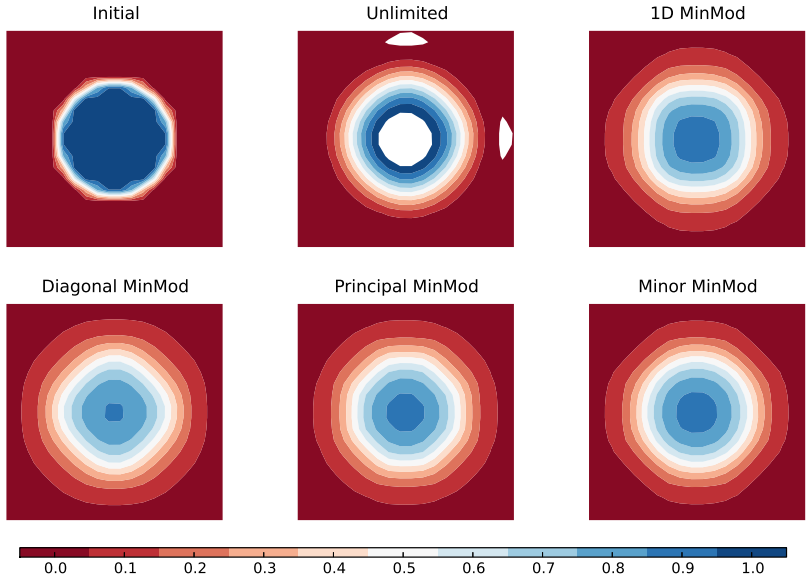


Figure 5: Circle shape contours at the final time 1 with a constant advection velocity (1,1) on a grid size of 20×20 cells for different limiters.

Grid	1D MinMod	Diagonal MM	Principal MM	Minor MM
20	7.31e-03	7.00e-03	4.52e-03	4.37e-03
40	4.36e-03	4.52e-03	3.81e-03	3.52e-03
80	2.84e-03	3.05e-03	2.41e-03	2.78e-03
160	1.55e-03	1.73e-03	1.53e-03	1.62e-03

Table 4: Symmetry violation E_s for four mesh resolutions for the circle shape advection.

anisotropy introduced by the 1D MinMod limiter. As a consequence, initially circular contours become more squared. The same "squaring" is presented for the Diagonal limiter, but the square is now rotated by 45 degrees. It suggests that this anisotropy is caused mainly by the limiter itself, regardless to the error of the directional split update of numerical fluxes. Our proposed Principal and Minor limiters provide more isotropic result.

In Table 4, the error E_s given by Equation (19) is presented. This error gives a measure of the anisotropy of the limiters. For both the Principal and the Minor limiters, the anisotropy is about two thirds of the 1D limiter for a coarse mesh. This difference is becoming less important with mesh refinement. However, the Principal limiter shows better convergence than the Minor limiter. The convergence with respect to the symmetry violation is shown in Figure 6.

4.3 Hydrodynamic tests

We start the hydrodynamic section with 2D Sedov blast wave followed by 2D Noh implosion test case, showing the importance of an isotropic piecewise linear reconstruction. We compare

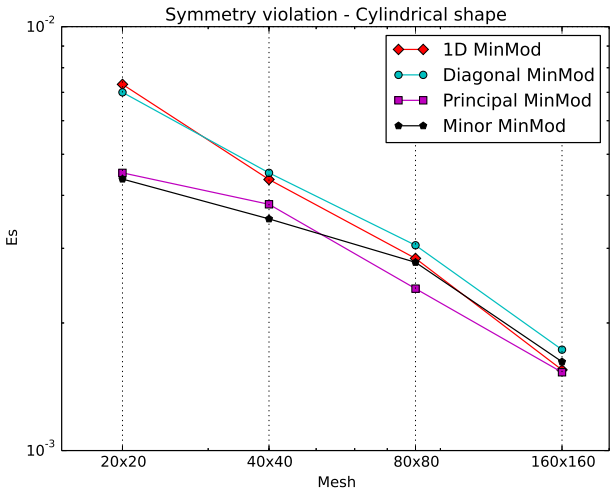


Figure 6: Symmetry violation convergence plots for linear advection of the circle shape for novel limiters.

three different limiters for a piecewise linear reconstruction, which is used in both the predictor step and the Riemann solver state definition.

4.3.1 2D Sedov blast wave

The Cartesian computational domain $(-1.2, 1.2) \times (-1.2, 1.2)$ is covered by a square mesh with 200×200 cells. At the initial time $t = 0$, the fluid (ideal gas with $\gamma = 1.4$) is static and has uniform density $\rho = 1$. Fluid pressure is $p = 10^{-6}$ everywhere except four computational cells

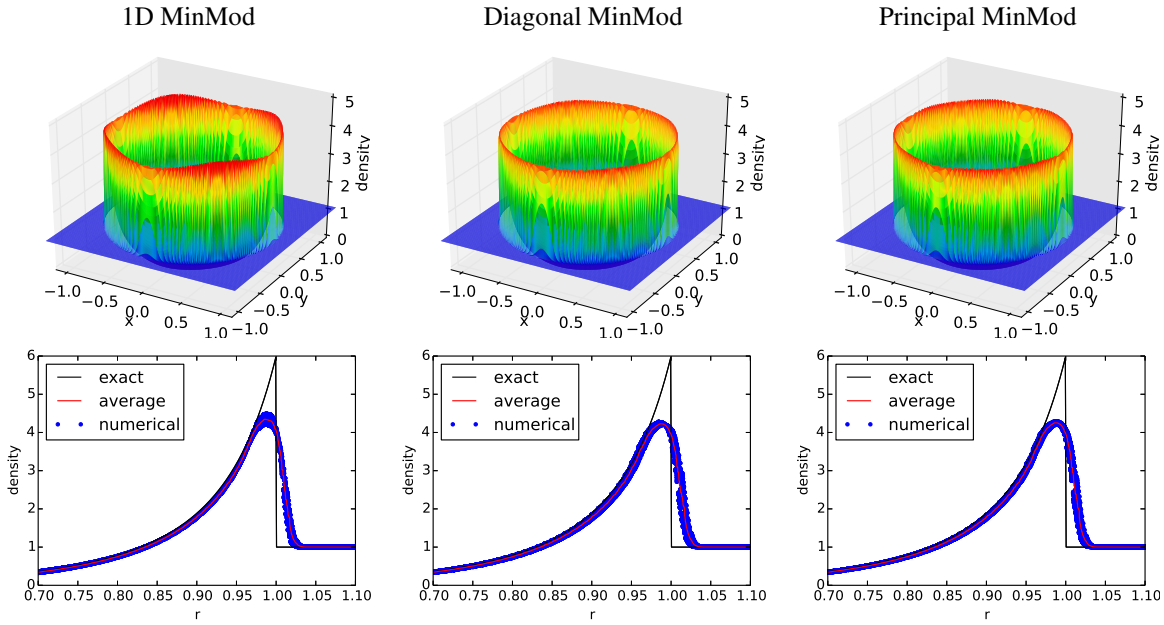


Figure 7: Density surface and scatter plots for Sedov explosion test case. 200×200 cells at final time 1.

in the center of the domain, where $p = 0.0979264/\Delta x^2$, which corresponds to the amount of energy in the definition of Sedov problem [11]. These high–energy cells represent a point initial

energy generating a circular shock wave spreading from the center. In the final time $t = 1$, the shock wave reaches the radius $r = 1$. Numerical results are shown in Figure 7.

For the standard 1D MinMod limiter, there is a clear angular dependency of the shock maximal density. The Diagonal limiter is more diffusive along diagonal directions, however it provides a more isotropic solution compared to the 1D MinMod limiter. Overall, the results obtained with the Principal and Diagonal limiters are comparable to each other for this test case and are much more isotropic than the results obtained with the 1D MinMod limiter. The lower part of Figure 7 provides a scatter plot of the final density, where the average solution (in red) is used in the calculation of the symmetry violation (19).

4.3.2 2D Noh implosion

Reduced anisotropy is further demonstrated by results of Noh implosion test case [12] shown in Figure 8 at the final time $t = 0.6$. The computation is initialized with the constant initial

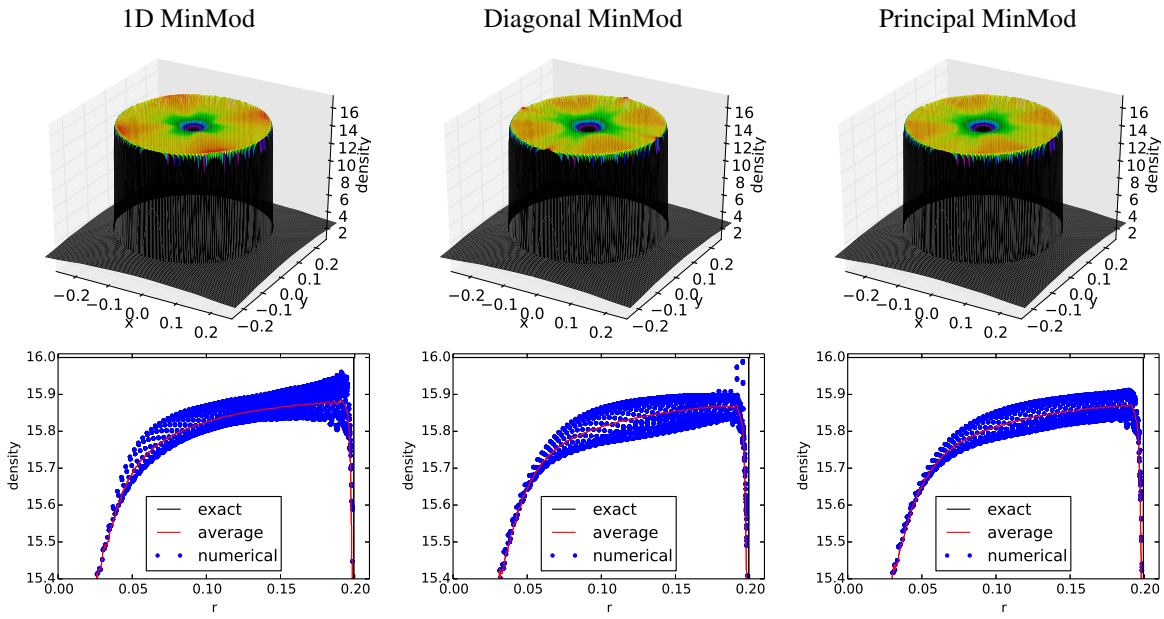


Figure 8: Density surface and scatter plots for Noh implosion test case. 200 × 200 cells at final time 0.6.

density $\rho = 1$ and the radial velocity $u^r = -1$ directed towards the origin on the square domain $(-0.8, 0.8) \times (-0.8, 0.8)$ covered by the equidistant rectangular grid with 200 × 200 cells. The initial and boundary pressure of the fluid with $\gamma = 5/3$ are set to $p = 1 \cdot 10^{-6}$ and we apply the exact density condition on the computational domain boundary $\rho(r, t) = 1 + t/r$ for $r > t/3$.

According to the construction of the standard 1D MinMod limiter, the numerical diffusion is again lower along the diagonal directions. Consequently, the numerical result is more accurate in regions along the diagonal directions. In Figure 8, these regions are located close to the boundary of the dense central spot with the radius $r < 0.2$, where $\rho^e = 16$. With the Diagonal MinMod limiter the highest–density regions are produced along the directions of the Cartesian axes, whereas with the 1D MinMod limiter, highest–density regions are produced along the diagonal axes, i.e. the regions are shifted by an angle of 45 degrees. In regions of highest–density, results with the Principal MinMod limiter are more diffusive but more isotropic compared to

the results with the 1D MinMod limiter. Reduced spread of the density along the central spot is further demonstrated by a zoom in scatter plots in the lower part of Figure 8.

5 CONCLUSIONS

We have demonstrated improved isotropy of the standard one–dimensional slope limiter for a piecewise linear reconstruction on square meshes for a directional split numerical scheme. The anisotropy is reduced by the application of a genuinely multi–dimensional slope limiter. The biggest improvement in the numerical solution calculated by our multi–dimensional minmod limiter (Principal MinMod) is observed on coarse meshes. This multi–dimensional approach could also be extended to other slope limiters. In future work, we will consider estimation of the numerical diffusion in conjunction with estimation of isotropic solution improvements.

ACKNOWLEDGEMENT

This work was performed under the auspices of the National Nuclear Security Administration of the US Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396. This work is supported by the Advanced Simulation and Computing program under the Hydrodynamics project. LA-UR-16-20072.

REFERENCES

[1] E. F. Toro. *Riemann solvers and numerical methods for fluid dynamics. A practical introduction.* Springer, Berlin, 1999.

[2] A. Kumar. Isotropic finite-differences. *Journal of Computational Physics*, 201:109–118, 2004.

[3] R. K. Shukla and P. Giri. Isotropic finite volume discretization. *Journal of Computational Physics*, 276:252–290, 2014.

[4] K. H. Kim and C. Kim. Accurate, efficient and monotonic numerical methods for multi-dimensional compressible flows Part II: Multi-dimensional limiting process. *Journal of Computational Physics*, 208:570–615, 2005.

[5] J. S. Park and C. Kim. Multi-dimensional limiting process for finite volume methods on unstructured grids. *Computers & Fluids*, 65:8–24, 2012.

[6] T. J. Barth and D. Jespersen. The design and application of upwind schemes on unstructured meshes. Technical Report AIAA-89-0366, NASA Ames Research Center, 1989.

[7] M. E. Hubbard. Multidimensional slope limiters for muscl-type finite volume schemes on unstructured grids. *Journal of Computational Physics*, 155:54–74, 1999.

[8] G. Luttwak and J. Falcovitz. VIP (vector image polygon) multi-dimensional slope limiters for scalar variables. *Computers & Fluids*, 83:90–97, 2013.

[9] B. van Leer. Towards the ultimate conservative difference scheme I. The quest of monotonicity. *Springer lecture notes in Physics*, 18:163–168, 1973.

[10] A. Harten, P.D. Lax, and B. van Leer. On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws. *SIAM Rev.*, 25:35–61, 1983.

[11] L. I. Sedov. *Similarity and Dimensional Methods in Mechanics*. Academic Press, 1959.

[12] W. F. Noh. Errors for calculations of strong shocks using an artificial viscosity and artificial heat flux. *Journal of Computational Physics*, 72:78–120, 1987.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

BUILDING A MORE EFFICIENT LAGRANGE-REMAP SCHEME THANKS TO PERFORMANCE MODELING

Thibault GASC[1,2,3], Florian De VUYST[2], Mathieu PEYBERNES[4], Raphaël PONCET[5] and Renaud MOTTE[3]

[1]Maison de la Simulation USR 3441
CEA Saclay, F-91191 Gif-sur-Yvette
e-mail: thibault.gasc@cea.fr

[2] CMLA, ENS Cachan, CNRS, Université Paris-Saclay,
94235 Cachan, France
e-mail: devuyst@cmla.ens-cachan.fr

[3]CEA DAM DIF
F-91297 Arpajon

[4] CEA, DEN, DM2S, STMF
CEA Saclay, F-91191 Gif-sur-Yvette
e-mail: mathieu.peybernes@cea.fr

[5] CGG
27 Avenue Carnot, F-91300 Massy
e-mail: raphael.poncet@cgg.com

Keywords: Performance Modeling, Co-design, Lagrange-Remap, Roofline model, Execution Cache Memory Model

Abstract. *This paper is a practical example of co-design between numerical analysis and high performance computing, applied to compressible fluid mechanics. We consider a legacy numerical method, based on a Lagrange-Remap solver for the compressible Euler equations, use tools from analytical performance modeling to quantitatively understand its behavior on recent multicore CPUs, and extract its computational bottlenecks. This analysis inspires us to propose a new numerical method, called Lagrange-Flux. Experimental results show that this new method yields an algorithm that is more computationally efficient, and at the same time retains the good numerical properties of the original Lagrange-Remap solver.*

1 Introduction

This paper is motivated by the adaptation and modification of a legacy numerical method arising in compressible fluid mechanics — the Lagrange-Remap solver —, in order to devise an alternative algorithm that is numerically equivalent, but more computationally efficient on current typical computing hardware: multicore CPUs. This kind of solver is used to approximate the compressible Euler equations, and is of particular interest for industrial applications involving complex compressible flows. It is for instance used in numerical simulation of plasma, high-speed flows of elastoplastic materials, and multi-material flows [1].

Computational efficiency of a given piece of code on a given hardware can be readily evaluated by measuring the runtime of this code. It is then straightforward to evaluate relative performance of two algorithms solving the same problem by comparing their results on the same benchmark, for both numerical quality and runtime. However, the validity of these results become tied to the particular hardware used for the tests. It is a problem, since in general, computing hardware evolves much more quickly other time than numerical methods (especially in the case of algorithms used in legacy industrial codes).

Hence, we propose to use a more future-proof, quantitative and general methodology to assess algorithm computational efficiency: we use a theoretical performance model to predict algorithm performance. This model depends on the algorithm itself, and on a simplified hardware description using a few parameters (typically less than 10). Hence, similarly to the scientific method applied to physics or engineering, if our performance model is rich enough to adequately capture the behaviour of our algorithm, it can be used not only for quantitative prediction, but also for precise qualitative understanding. In a previous work [2], we established that the so-called ECM model [3] has the capability to accurately predict Lagrange-Remap solver performance on multicore CPUs.

The main contribution of this paper is to show how performance modeling using the ECM model leads us to extract the computational bottlenecks of the Lagrange-Remap algorithm, link them to properties of the numerical method, and naturally propose the Lagrange-Flux algorithm (originally introduced in [4] from a numerical analysis point view) as an alternative. Superior computational performance of this new algorithm — from both a scalability and absolute performance point of view — is then inferred from theoretical analysis, and verified by numerical experiments. Up to our knowledge, this is the first published work in computational fluid dynamics using analytical performance models as a quantitative methodology for carrying out co-design between numerical analysis and high performance computing.

This paper is organized as follows: in section 2, we detail the analytical performance models used in this study for predicting and understanding algorithm runtimes. In section 3, we briefly recall the formulation of the legacy Lagrange-Remap solver, sum up the results of its performance analysis carried out in our previous work [2], and deduce its performance bottlenecks. In section 4 we briefly introduce the Lagrange-Flux solver, conduct its performance analysis, theoretically show its computational advantages over the Lagrange-Remap solver, and verify the better performance and scalability of this new algorithm. Finally we discuss the interest of our methodology and some perspectives brought by this work in section 5.

2 Performance modeling

2.1 Introduction

Performance modeling aims at quantitatively predicting and understanding code performance — in terms of runtime — using simple analytical models for both the algorithm itself, and the

hardware on which the algorithm is executed. Because of the complexity of recent computing hardware, it is not a simple task.

Indeed, let us consider the simplest performance model one could think of for predicting a given algorithm runtime on a given machine: count the number of arithmetic operations (such as additions, multiplications or divisions) the algorithm performs, and compare it with the maximum — or *peak* — number of operations the hardware can execute per unit of time (which can be readily obtained using information from hardware vendor specification sheets, e.g. frequency and peak number of arithmetic instructions per hardware cycle). In most of the case, it will yield a very inaccurate number, for several reasons:

- the algorithm runtime is determined not only by the arithmetic operations it performs, but also the rate at which it can fetch data to/from memory. The bottlenecks of most algorithms on recent hardware correspond to the latter case (we talk of *memory bound* algorithms, compared to *compute bound* algorithms whose performance is bounded by the rate at which the hardware can process arithmetic instructions).

- the memory subsystem of any recent architecture is complex: it consists in several hierarchical layers of memory, such as caches for CPUs, with various capacities and bandwidths.

- all arithmetic instructions are not equal performance wise: for instance, divisions and square roots typically have a throughput at least one order of magnitude lower than additions or multiplications.

- last, but not least, any recent piece of hardware has several layers of parallelism. For instance, common multicore CPUs such as the one considered in this paper have three levels of parallelism: they consist in several cores, each core having single instruction multiple data — SIMD — vectorized units (such as AVX for recent Intel CPUs). Finally, each core has several pipelines executing instructions in parallel (this is called instruction level parallelism — ILP). Moreover, this parallelism is tied to the memory subsystem: for instance, some memory layers (usually the L1 and L2 caches) are private to CPU cores, but some (usually the L3 cache and the DRAM) are shared between all cores.

This underlying hardware complexity implies that precisely predicting and understanding the performance of even deceptively simple algorithms such as the Schönauer Vector Triad $A = B + C * D$ (where A, B, C and D are arrays) on a typical multicore CPU is not trivial [3].

However, some successful analytical performance models have been recently popularized in order to assess algorithmic performance on recent hardware. The most known model is the so-called Roofline model ([5]), which has been initially proposed as a conceptual qualitative model characterizing the behavior of an algorithm as compute-bound (CB) or memory-bound (MB).

2.2 The Roofline model

In the following, we consider elementary algorithms called *kernels*, acting on floating point data. Hence, we use the *billion floating point operations per second* (GFlop/s) metric for reporting Roofline performance, as is commonly done. The advantage of this metric is that it is easy to directly compare algorithm performance with peak CPU arithmetic throughput which is also

given in GFlop/s, and thus get an insight on kernel efficiency on a given machine. The Roofline model approximates the performance of an algorithm with the following single formula:

$$P = \min(\mathrm{P_{max}}(\mathrm{P_{peak}}), \mathrm{AI} * \mathrm{B_{max}}), \tag{1}$$

where $\mathrm{P_{peak}}$ is the machine peak arithmetic throughput (in GFlop/s), $\mathrm{P_{max}}$ is the kernel maximum arithmetic throughput (in GFlop/s), AI is the kernel arithmetic intensity (in Flop/Byte), and $\mathrm{B_{max}}$ is the maximum machine bandwidth (in GByte/s). Parameter AI only depends on the algorithm (and can be readily obtained by counting the ratio between the *number* of arithmetic instructions and transferred data), whereas parameter $\mathrm{P_{peak}}$ only depends on the machine. Several choices are possible for parameter $\mathrm{B_{max}}$ and for the function $\mathrm{P_{max}}(\mathrm{P_{peak}})$. The basic version of Roofline takes for $\mathrm{B_{max}}$ the peak CPU bandwidth provided by the vendor, and determine $\mathrm{P_{max}}(\mathrm{P_{peak}})$ by manually counting algorithm arithmetic instructions. However, this yields an overly optimistic model which is rarely quantitatively accurate. Hence, we use a refined version of the Roofline model (sometimes called *Roofline with ceilings*), where $\mathrm{B_{max}}$ is chosen to be the effective bandwidth of the full CPU system measured using a STREAM benchmark [6]. Moreover, $\mathrm{P_{max}}(\mathrm{P_{peak}})$ is determined using static analysis, using the `Intel IACA` tool [7] in throughput mode (thus ignoring any latency effects). In particular, the obtained maximum arithmetic throughput takes into account the heterogeneous throughput of arithmetic instructions, and their concurrent scheduling on the core pipelines. In this refined version of the model, the function $\mathrm{P_{max}}(\mathrm{P_{peak}})$ is part of the model. We must emphasize that it does not come from execution of the kernel, but from *static inspection* of the binary by `Intel IACA`.

This Roofline has a simple interpretation: it is an optimistic model that simply expresses the fact that, for a kernel to be executed, data has to be fetched to/from memory, which corresponds to the $\mathrm{AI} * \mathrm{B_{max}}$ term, and kernel computations have to be performed, which corresponds to the $\mathrm{P_{max}}(\mathrm{P_{peak}})$ term. Hence, kernel performance is necessary less than the minimum of these two terms. Roofline assumes that kernel performance can be approximated by this asymptotic value.

This model has enjoyed wide success [5], nevertheless it has some limitations. First, it does not accurately model caches. The consequence is that single core performance is not quantitatively accurate for memory-bound kernels (it is too optimistic). Hence, in particular, multicore scalability can not be inferred from this model [3, 8]. Moreover, the GFlop/s metric which is generally used with Roofline models has some drawbacks from a practical methodological point of view. First, it is not directly correlated with a meaningful performance metric for algorithm developers (who are generally interested in the number of loop iterations done per unit of time). Moreover, in some cases, the Flop/s hardware counters on recent multi-core CPUs — such as the Sandy Bridge processor — are known to be unreliable. The ECM — Execution Cache Memory — model has been recently introduced to overcome these drawbacks.

2.3 The ECM model

The ECM model ([3], [8]) is a refinement of the Roofline model for multicore CPUs that still neglects any latency effects, but takes into account the cache hierarchy. It uses the *cycles per cacheline worth of data* (cy/CL) performance metric. A *cacheline worth of data* corresponds, in a loop-based algorithm, to the number of loop elements that fit in a cacheline (64 Bytes in most modern CPU architectures, such as the one we used in this paper). For instance, for algorithms using double precision — corresponding to 8 byte storage —, a cacheline worth of data is 8 elements for scalar non-vectorized code, but 2 elements for AVX vectorized code (because AVX

registers are 4-Byte wide). This metric is interesting because, on the one hand, it is directly correlated with the number of loop iterations per unit of time, and on the other hand, it is a reliable metric that can be measured on any CPU (because it can be derived directly from CPU runtime), and does not depend on hardware counters.

As in [3], [8], [9], we assume that STORE and floating point arithmetic instructions are overlapped with the data transfers between different levels of the memory system, whereas LOAD operations do not overlap. The data transfers between the L1 and L2 caches, and between the L2 and L3 caches, depends on the architecture and are given in table 1. The data transfer transfer rate $B_{\mathrm{L3}\to\mathrm{mem}}$ (in cy/CL) between the L3 cache and memory subsystem depends on the CPU frequency f (in GCycles/s) and the sustainable bandwidth $B_{\max}$ (in GBytes/sec), and is given by the following formula:

$$B_{\mathrm{L3}\to\mathrm{mem}} = \frac{64f}{B_{\max}}. \tag{2}$$

In practice, ECM model delivers a prediction of the number of CPU cycles required to execute a certain number of iterations of a given loop on a single core. Then, the prediction time T_{ECM} is given by $T_{ECM} = \max(T_{OL}, T_{nOL} + T_{data})$ ([3], [8]), where T_{data} is the transfer time through the memory hierarchy (L1 upwards), T_{nOL} the time taken by the load instructions (we assume that these instructions can not overlap with any memory transfer) and T_{OL} the time for all other instructions (store and arithmetic instructions). As for the refined Roofline model, the times T_{OL} and T_{nOL} are determined using `Intel IACA`. For data in memory, the time T_{data} is given by $T_{data} = T_{L1L2} + T_{L2L3} + T_{L3Mem}$, where T_{L1L2}, T_{L2L3} and T_{L3Mem} are respectively the transfer times between L1 and L2, L2 and L3 and L3 and memory. Transfer times for streaming loads and stores can be computed readily using micro architecture knowledge (see table 1 for the parameters used in this study). We refer to [9], [2] for the estimation of performance for stencil-like access patterns.

3 Identifying the legacy Lagrange-Remap solver computational bottlenecks using performance modeling

In this section, we first recall the main findings of the performance analysis we carried out for the Lagrange-Remap algorithm in [2]. We then use these results to identify the bottlenecks of this algorithm.

3.1 Description of the legacy Lagrange-Remap solver

We consider a cartesian Lagrange-Remap solver with the following features:

- discrete thermodynamic variables (mass, internal energy, pressure) are cell-centered variables;
- discrete velocity are node-base variables (this helps for physics coupling, and mesh distortion management);
- the lagrangian step is performed using a 2^{nd} order in time integration scheme of leap-frog type;
- the remap step is performed using an alternating direction strategy;
- second order in space accuracy is obtained thanks to a MUSCL-type reconstruction during the remap step;

- for shock capturing and entropy consistency, pseudo-viscosity is used.

This solver has been described in [1, 10, 11]. Its implementation is organized as follows: at the start of each time step, the mass (or density), the energy (or pressure) and the velocity fields are known. The lagrangian step is performed to obtain intermediate updated lagrangian fields of these variables. Then to go back to the fixed cartesian grid, a remap step is performed in each direction. Each step consists of several kernels. Other secondary variables are also introduced. Figure 1 is a dataflow diagram for the idealized Lagrange-Remap solver studied in [2] that shows dependencies between kernels and data for the lagrangian step and the remap step in one direction. Part of the complexity of these graphs are due to the use of staggered variables, as we explain in section 3.2.

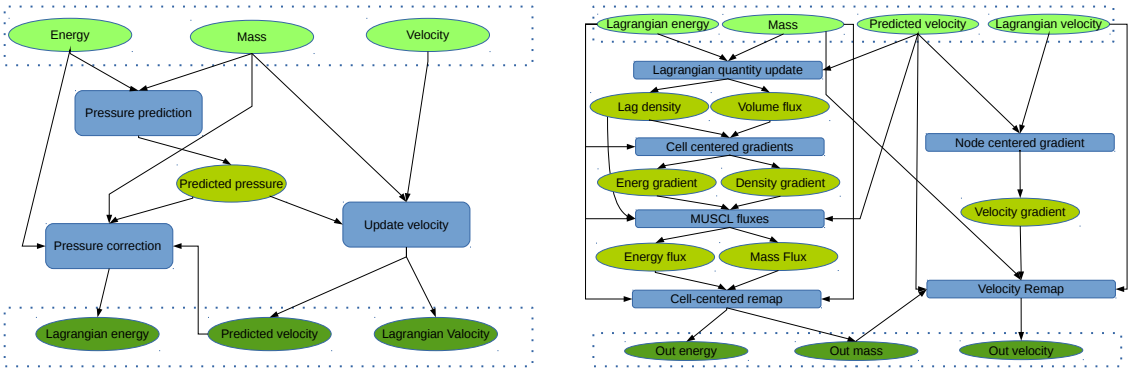


Figure 1: Dataflow diagrams of the lagrange (left) and remap (right parts of the algorithm. Kernels are represented in blue, input data in light green, output data in dark green, and temporary data in khaki green. Courtesy of [2]

3.2 Performance modeling of the Lagrange-Remap solver

The main result of paper [2] is that the performance of all individual kernels of our solver can be theoretically predicted with single digit accuracy on multicore CPUs, if the ECM model is used. The reported mean and median errors between model prediction and measurement lie between 3% and 8%. Tests have been conducted using Haswell and SandyBridge micro-architecture, for input data in L3 cache and main memory, and for scalar, vectorized, and multithreaded & vectorized versions of each kernel.

This validates the relevance of the ECM model for finely assessing performance of the Lagrange-Remap solver algorithm (and by extension, of most explicit solvers for fluid mechanics on cartesian meshes). Moreover, a closer look at the results allows us to identify three performance bottlenecks, and, more importantly, to link them with numerical properties of the numerical method:

1. using **staggered variables**;
2. using an **alternate direction** strategy for the remap;
3. taking into account **geometry** of the lagrangian mesh.

Staggered variables Most kernels are stencil type kernels. This means output data at a given grid location is obtained from input data at this location and its nearest neighbours. In

the Lagrange+Remap solver, staggered variables are used (see figure 2). Hence, most kernels alternate reading and writing data between the primal cell-centered grid, and the dual node-centered grid. For this reason, successive kernels can not be easily merged. Thus, we have multiple small kernels in which arithmetical intensity is low, and data reuse is limited. As a consequence, most kernels are memory-bound, which yields suboptimal performance and low scalability.

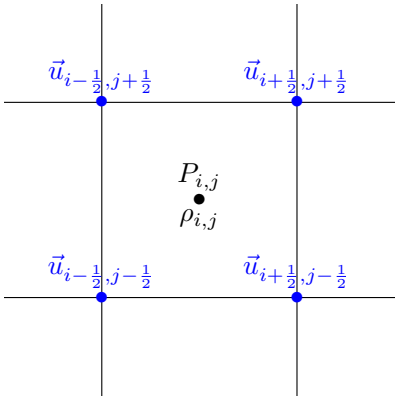


Figure 2: Variable positions for the legacy staggered Lagrange-Remap solver. Pressure and density are stored at cell centers, velocity is stored at node centers.

Alternate directions Using an alternate direction strategy for the remap step, e.g. a succession of 1D remap steps, introduces an extra intermediaty states, which adds stress on memory transfers compared to a true multidimensional remap step, which is already the bottleneck of most of our kernels [2]. Since memory bandwidth is shared among multiple cores, it also lowers multi-core scalability.

Lagrangian geometry Since the mesh is moving, lengths of the edges of the mesh are changing. Thus it is not possible to pre-compute useful geometrical data and this can lead to costly computations. The most significant example in the numerical scheme is the computation of gradients during the remap step. Indeed computing $\frac{q_i - q_{i-1}}{x_i - x_{i-1}}$ is more expensive if we cannot precompute once and for all $\frac{1}{x_i - x_{i-1}}$ because it is not constant. If it was, we could save the result and perform a multiplication instead of a division, which is typically an order of magnitude faster.

4 The Lagrange-Flux solver

4.1 Lagrange-Flux numerical scheme

The detailed construction of the Lagrange-Flux algorithm is given in [4]. Hereafter, we give key ideas of its construction and provide a schematic description. From the numerical properties of the original Lagrange-Remap solver and the conclusions of its performance analysis presented in section 3, we extract the following desirable properties constraining the design of the Lagrange-Flux solver:

1. A lagrangian solver is used (for multi-material flows or physics coupling);

2. A cell centered description is used rather than a staggered one, this will allow kernel fusion optimization and reduce data communication;

3. A direct multidimensional remap strategy is used rather than alternating direction strategy, this should also reduce data communications;

4. The method can be extended to second order accuracy in space and time, matching the accuracy of the legacy Lagrange-Remap solver;

5. Complex geometry related computations should be limited;

6. 3D extension of the solver should remain simple.

Let us now briefly describe the proposed scheme. We consider the compressible Euler equations for two-dimensional problems. Denoting ρ, $\boldsymbol{u} = (u_i)_i$, $i \in \{1, 2\}$, p and E the density, velocity, pressure and specific total energy respectively, the mass, momentum and energy conservation equations can be written in the compact following form:

$$\partial_t U_\ell + \nabla \cdot (\boldsymbol{u} U_\ell) + \nabla \cdot \boldsymbol{\pi}_\ell = 0, \quad \ell = 1, \dots, 4, \tag{3}$$

where $U = (\rho, (\rho u_i)_i, \rho E)$, $\boldsymbol{\pi}_1 = \vec{0}$, $\boldsymbol{\pi}_2 = (p, 0)^T$, $\boldsymbol{\pi}_3 = (0, p)^T$ and $\boldsymbol{\pi}_4 = p\boldsymbol{u}$. For the sake of simplicity, we will close this system using a perfect gas equation of state $p = (\gamma - 1)\rho(E - \frac{1}{2}|\boldsymbol{u}|^2)$, $\gamma \in (1, 3]$.

We first split the usually considered 2 steps Lagrange-Remap scheme in 3 steps. The lagrange step is taken identical, the remap step is split in a backward lagrangian motion followed by a forward advection step described in the eulerian frame.

This leads to the following scheme:

$$\begin{aligned} (U_\ell)_K^{n+1} = (U_\ell)_K^n \quad &- \quad \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \left(\frac{|A^{n+\frac{1}{2},L}|}{|A|} \left(\boldsymbol{\pi}_\ell\right)_A^{n+\frac{1}{2},L} \cdot \nu_A^{n+\frac{1}{2},L} \right) \\ &- \quad \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \left(\left(U_\ell\right)_A^{n+\frac{1}{2},\star} \left(\boldsymbol{v}_A^{n+\frac{1}{2}} \cdot \nu_A\right) \right). \end{aligned} \tag{4}$$

where K is the cell area, A an edge of the cells, ν_A the unit normal vector of the edge A pointing outwards, $\cdot^L$ superscript denotes elements of the lagrangian mesh, $\cdot^{n+\frac{1}{2}}$ superscript denotes elements at time $t^{n+\frac{1}{2}}$ used for targeted second order in time accuracy, and $\boldsymbol{v}_A^{n+\frac{1}{2}}$ denotes the velocity of the grid used in the lagrangian step.

From this intermediate numerical scheme scheme we would like to suppress the geometrical computation of lagrangian geometry. By making Δt go to zero, ($t > 0$), we have $A^{n+\Delta t/2,L} \to A$, $(\boldsymbol{\pi}_\ell)^{n+\Delta t/2,L} \to \boldsymbol{\pi}_\ell$, $\boldsymbol{v}^{n+\Delta t/2} \to \boldsymbol{u}$, $(U_\ell)^\star \to U_\ell$; we obtain a partial discretization in space of the conservation laws which can be seen as a method of lines:

$$\frac{d(U_\ell)_K}{dt} = -\frac{1}{|K|} \sum_{A \subset \partial K} |A| \left((\boldsymbol{\pi}_\ell)_A \cdot \nu_A \right) - \frac{1}{|K|} \sum_{A \subset \partial K} |A| \left(U_\ell\right)_A \left(\boldsymbol{u}_A \cdot \nu_A\right). \tag{5}$$

This can be also be written as a finite volume method:

$$\frac{dU_K}{dt} = -\frac{1}{|K|} \sum_{A \subset \partial K} |A| \, \Phi_A,$$

with a numerical flux Φ_A whose components are

$$(\Phi_\ell)_A = \left(U_\ell\right)_A \left(\boldsymbol{u}_A \cdot \nu_A\right) + (\boldsymbol{\pi}_\ell)_A \cdot \nu_A. \tag{6}$$

In (5), pressure fluxes $(p_\ell)_A$ and interface velocities $\boldsymbol{u}_A$ are computed from an approximate Riemann solver in lagrangian coordinates (for example the lagrangian HLL solver, see [12]). Then, the interface states $(U_\ell)_A$ are computed from a upwind process according to the sign of the normal velocity $(\boldsymbol{u}_A \cdot \nu_A)$. Higher-order accuracy in space can be achieved using a standard MUSCL reconstruction with a slope limiting process. In this semi-discrete formalism, there is no time discretization, thus all kernel act on the eulerian mesh and fluxes are defined at the edges the eulerian cells. To achieve second-order accuracy in time, one can apply any second-order time advance scheme. The second-order Heun scheme for example leads to the following algorithm:

1. Compute the time step Δt^n subject to the CFL stability condition;
2. Predictor step:
   (a) Reconstruct higher order interpolation from the discrete values U_K^n (MUSCL + slope limitation): compute a discrete gradient for each cell K ;
   (b) Compute interface velocities $\boldsymbol{u}_A^n$ and pressure fluxes $\boldsymbol{\pi}_A^n$ using a lagrangian approximate Riemann solver;
   (c) Select the upwind edge values $(U_\ell)_A^n$ according to the sign of $(\boldsymbol{u}_A^n \cdot \nu_A)$;
   (d) Compute the numerical flux Φ_A^n as defined in (6);
   (e) Compute the first order predicted states $U_K^{\star,n+1}$:
$$U_K^{\star,n+1} = U_K^n - \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \, \Phi_A^n;$$
3. Corrector step:
   (a) Repeat steps (2a-2d) to compute the numerical flux $\Phi_A^{\star,n+1}$ from the predicted state $U_K^{\star,n+1}$;
   (e) Compute the second-order accurate states U_K^{n+1} at time t^{n+1}:
$$U_K^{n+1} = U_K^n - \frac{\Delta t^n}{|K|} \sum_{A \subset \partial K} |A| \, \frac{\Phi_A^n + \Phi_A^{\star,n+1}}{2}.$$

This numerical scheme fullfills the design specifications presented at the beginning of this section. It has been validated using several test cases. We present the results of such a test in figure 3: numerical quality of the method is assessed by comparing it to the original Lagrange-Remap solver on a triple point type computation.

4.2 Implementation details of the Lagrange-Flux solver

The implementation of the Lagrange-Flux solver is quite simple compared to the legacy Lagrange-Remap solver. Excluding the computation of the time step common to both solvers, we first derive a naive implementation by simply following the algorithm description given at the end of the previous section. At this point, we start optimizing the code. Since the Lagrange-Flux solver uses collocated cell-centered variables, we are able to merge multiple small kernels into bigger ones. In the most simple case, instead of doing computations for kernel K_1, writing data to memory, reloading it immediately after, and doing the computations for kernel K_2, we

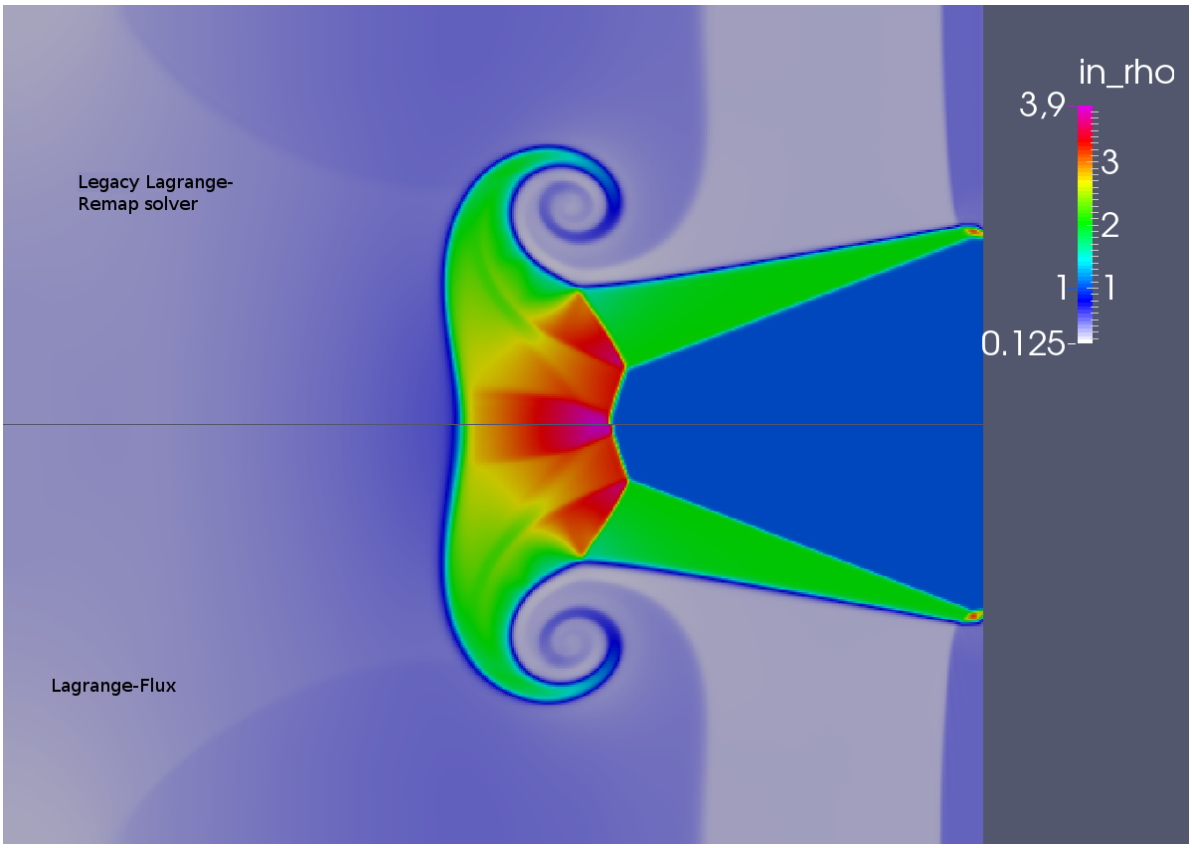


Figure 3: Comparison of the Lagrange-Remap solver and the Lagrange-Flux solver on a triple point test case. In this two-dimensional Riemann problem, we check that the shock wave and the vortex are correctly captured. Density maps at time $t = 3.5s$ are presented for the legacy Lagrange-Remap solver (upper part) and the Lagrange-Flux solver (lower part) on a 560×240 mesh. One can notice that both solvers capture the same physical phenomenons.

load the data once and do the computations for both kernels K_1 and K_2. This mechanically increases arithmetic intensity. In more complex cases, there is a compromise to be made: if the result of a given kernel K is reused by N other kernels, merging can be done, at the cost of doing kernel K computations N times. In this case, merging is beneficial only when the time needed for extra computation is shorter that the time saved from increased memory reuse. For the Lagrange Flux algorithm, it turns out that merging aggressively is the best option.

At the end of the optimization process, we are left with only two large kernels which are almost identical: one for the prediction step, and one for the correction step (see figure 5). In these kernels, density, velocity and energy are updated at the same time using the same stencil (see figure 4). Comparing figures 5 and 1, we can expect that the Lagrange-Flux solver is less memory-bound than the original Lagrange-Remap solver. The following section confirms that this is indeed the case.

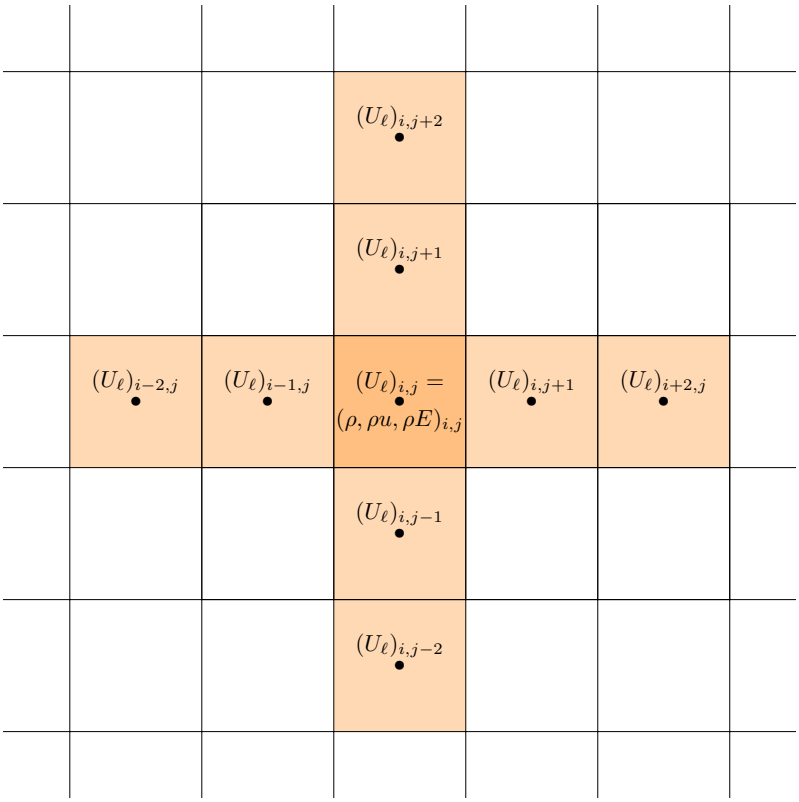


Figure 4: Stencil of the Lagrange-Flux kernels. The target updated cell is in dark orange, the full neighborhood is in lighter orange. All variables (density, energy and velocities) are loaded from the same stencil. Computations for their update are performed in the same time and many common intermediary results can be reused without being stored in the main memory.

4.3 Performance analysis and measurements

In this section, we prove that the ECM model has predictive capabilities for both the full Lagrange-Remap solver, and its Lagrange-Flux alternative. The machine used for the following tests is a dual socket octo-core SandyBridge processor (see 1).

First, in table 2, we present a comparison between model prediction and measurements on a single core, without vectorization. We observe quantitative agreement for both solvers. A closer look at the Lagrange-Flux performance model output shows that the divide port is the bottle-

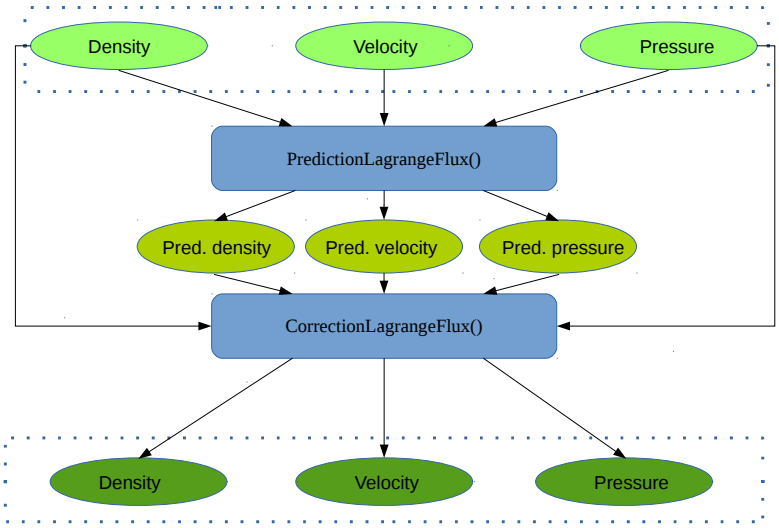


Figure 5: Dataflow diagrams of the Lagrange-Flux solver. Kernels are represented in blue, input data in light green, output data in dark green, and temporary data in khaki green

Table 1: Characteristics of the CPU hardware used.

Micro Architecture	**Intel SandyBridge**
Model	E5-2670
Number of cores	2x8
Clock (GHz)	2.6
Peak Flops DP (GFlop/s) (full socket)	168
Cache sizes: L1	8x32kB
L2	8x256kB
L3	20MB
Theoretical Bandwidth (GBytes/s)	51.2
Measured triad Bandwidth $B_{\max}$ (GBytes/s)	38.4
L1 $\leftrightarrow$ L2 bandwidth (cy/CL)	2
L2 $\leftrightarrow$ L3 bandwidth (cy/CL)	2

Table 2: Comparison between single core non-vectorized ECM prediction and measurements for the intel SandyBridge architecture. Performance is expressed in cycle per cache line worth of data (cy/CL), and kernel type can be CB — Compute-Bound — or MB — Memory-Bound. The analytical ECM model gives an accurate performance prediction.

Kernel name	**type**	**prediction**	**measure**	**error**
Full Lagrange-Remap	MB	10697 cy/CL	10092 cy/CL	6%
Prediction Lagrange-Flux	CB	5712 cy/CL	5548 cy/CL	2.8%
Correction Lagrange-Flux	CB	5712 cy/CL	5690 cy/CL	0.3%

Table 3: Speed up on single core Intel SandyBridge with AVX for both Lagrange-Flux kernels. Even if the kernels are compute bound, speedup is not perfect, because the bottleneck is the DIV port throughput.

Kernel name	**type**	**predicted**	**measured**
Prediction Lagrange-Flux	CB	2.0	1.9
Correction Lagrange-Flux	CB	2.0	1.9

neck for both kernels, making them compute-bound. Note that in this case, the Roofline and ECM models are equivalent, and give the same prediction. With AVX SIMD vectorization, the bottleneck remains the divide port. The consequence is that the ECM model predict a speedup of 2 with respect to the baseline scalar version: on the one hand, elements are processed 4 by 4 thanks to AVX, but on the other hand, the divide port throughput is only doubled. Since the Lagrange-Flux is compute-bound, our performance model also predicts perfect scalability (e.g. x16 speedup) for multithreaded execution. It is close to the observed speedup (x13.3).

Finally, in order to compare absolute performance and scalability of Lagrange-Flux and Lagrange-Remap solvers, we report in 4.3 a measured runtime comparison for the same test case. The results confirm what our performance model is able to predict: the Lagrange-Flux solver has superior absolute performance. It is interesting to note that this is achieved by trading single core unvectorized performance (which is worse than the corresponding Lagrange-Remap solver) for much better scalability.

5 Conclusions

In this paper, we have proposed a practical and quantitative way of modifying a fluid mechanics solver, the Lagrange-Remap algorithm, for better performance on multicore CPUs. The resulting algorithm, the Lagrange-Flux solver, has comparable numerical quality, but better runtime, and is more compact (two kernels instead of a dozen). The main tools used in our

Scheme	**1 core**	**1 core AVX**	**16 cores AVX**	**scalability**
Lagrange-Flux	1.9	3.9	52.0	27.1
Lagrange-Remap	2.4	3.7	36.5	14.6

Table 4: Performance comparison between the reference Lagrange-Remap solver and the Lagrange-Flux solver in millions of cell updates per second (MCUPs) on SandyBridge. Different machine configurations are presented. Scalability (last column) is computed as the speed up of the multi-threaded vectorized version compared to the baseline purely sequential version. Tests are performed for fine meshes, such that kernel data lies in DRAM memory. The Lagrange-Flux solver exhibits superior scalability, because it has — by design — better arithmetic intensity.

methodology are analytical performance models, allowing us to quantitatively assess and qualitatively precisely understand algorithmic performance. They explain the main reason behind the superior performance of our new solver: it is a compute-bound algorithm, whereas the original is memory-bound.

We believe that our methodology is quite general, and can be extended to other computational physics algorithms, for fluid mechanics but also for other domains of application. An other interesting perspective is to extend the use of performance models to many-core architectures such as graphical processing units (GPUs), or the Intel Many Integrated Core (MIC).

References

[1] David J Benson. "Computational methods in Lagrangian and Eulerian hydrocodes". *Computer methods in Applied mechanics and Engineering* 99.2 (1992), pp. 235–394.

[2] Raphael Poncet, Mathieu Peybernes, Thibault Gasc, and Florian De Vuyst. *Performance modeling of a compressible hydrodynamics solver on multicore CPUs*. accepted at ParCo2015 conference, Edinburgh, Scotland, UK 1-4 sept. 2015. 2015.

[3] Jan Treibig and Georg Hager. "Introducing a Performance Model for Bandwidth-Limited Loop Kernels". *Parallel Processing and Applied Mathematics*. Ed. by Roman Wyrzykowski, Jack Dongarra, Konrad Karczewski, and Jerzy Wasniewski. Vol. 6067. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, pp. 615–624.

[4] Florian De Vuyst, Thibault Gasc, Renaud Motte, Mathieu Peybernes, and Raphael Poncet. *Lagrange-flux schemes: reformulating second-order accurate Lagrange-remap schemes for better node-based HPC performance*. Proccedings of the SimRace2015 conference in OGST. 2016.

[5] Samuel Williams, Andrew Waterman, and David Patterson. "Roofline: An Insightful Visual Performance Model for Multicore Architectures". *Commun. ACM* 52.4 (Apr. 2009), pp. 65–76.

[6] JD McCalpin. "STREAM: Sustainable memory bandwidth in high performance computers". *Technical Report, University of Virginia* (continuously updated).

[7] *Intel architecture code analyzer*. Available: http://software.intel.com/en-us/articles/intel-architecture-code-analyzer/.

[8] Georg Hager, Jan Treibig, Johannes Habich, and Gerhard Wellein. "Exploring performance and power properties of modern multi-core chips via simple machine models". *Concurrency and Computation: Practice and Experience* 28.2 (2016), pp. 189–210.

[9] Holger Stengel, Jan Treibig, Georg Hager, and Gerhard Wellein. "Quantifying Performance Bottlenecks of Stencil Computations Using the Execution-Cache-Memory Model". *Proceedings of the 29th ACM on International Conference on Supercomputing*. ICS '15. Newport Beach, California, USA: ACM, 2015, pp. 207–216.

[10] Mark L Wilkins. "Use of artificial viscosity in multidimensional fluid dynamic calculations". *Journal of computational physics* 36.3 (1980), pp. 281–303.

[11] Paul Woodward and Phillip Colella. "The numerical simulation of two-dimensional fluid flow with strong shocks". *Journal of computational physics* 54.1 (1984), pp. 115–173.

[12] Eleuterio F Toro. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction. Third edition*. Springer-Verlag, 2009.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DEVELOPMENT OF CONFINED VISCOPLASTIC FLOWS WITH HETEROGENEOUS WALL SLIP

Pandelitsa Panaseti[1], **Maria Philippou**[1], **Zacharias Kountouriotis**[1], **Georgios Georgiou**[1]

[1] Department of Mathematics and Statistics, University of Cyprus
PO Box 20537, 1678 Nicosia, Cyprus
{panaseti.pandelitsa, filippou.maria, kountouriotis.zacharias, georgios}@ucy.ac.cy

Keywords: Plane Poiseuille Flow, Bingham plastic, Wall Slip, Navier Slip, Flow Development, Finite Elements.

Abstract. *The steady, pressure-driven flow of a Herschel-Bulkley fluid in a channel is considered assuming that slip occurs on one wall only due to slip heterogeneities. Hence, the velocity profile is allowed to be asymmetric. The fully-developed solutions are derived and the different flow regimes are identified. The development of the flow is investigated numerically using the Papanastasiou regularization for the constitutive equation, a power-law slip equation, and finite element simulations. The combined effects of slip and the Bingham number are discussed.*

.

1 INTRODUCTION

Wall slip is important in many industrial applications, such as the extrusion of complex fluids, ink jet processes, oil migration in porous media, and in microfluidics. Viscoplastic or yield stress materials, such as polymeric solutions, suspensions, and gels, are known to exhibit wall slip [1-3]. While wall slip with polymer melts is observed at large rates of strains, it appears within a range of rather small strains in the case of pasty materials [4]. Based on the analysis of apparent slip flows of Herschel-Bulkley fluids in various geometries, Kalyon [5] proposed a power-law slip equation, relating the wall shear stress, τ_w, to the slip (or sliding) velocity, u_w, defined as the relative velocity of the fluid with respect to that of the wall,

$$\tau_w = \beta u_w^s \tag{1}$$

where s is the exponent, and β is the slip coefficient. The latter incorporates the effects of all other factors affecting slip, such as the temperature, the normal stress, the molecular parameters, and the properties of the fluid/wall interface [4]. Setting s=1 in Eq. (1) leads to the classical Navier slip condition [6]

$$\tau_w = \beta u_w \tag{2}$$

in which case β is related to the so-called slip (or extrapolation) length b, i.e. $\beta \equiv \mu / b$, where μ denotes the viscosity. The no-slip and full-slip limiting cases are recovered when $\beta \to \infty$ and $\beta = 0$, respectively.

The present work is motivated by the recent findings of Vayssade et al. [7] who imaged the motion of well characterized softy glassy suspensions in microchannels whose walls impose different slip velocities. These materials exhibit yield stress and are modeled as Herschel-Bulkley fluids. The Herschel-Bulkley constitutive equation, which involves three material parameters, the yield stress τ_0, the consistency index k, and the power-law exponent, n. The tensor form of the model is as follows:

$$\begin{cases} \dot{\boldsymbol{\gamma}} = \mathbf{0}, & \tau \le \tau_0 \\ \boldsymbol{\tau} = \left(\dfrac{\tau_0}{\dot{\gamma}} + \mathrm{k}\dot{\gamma}^{\mathrm{n}-1} \right) \dot{\boldsymbol{\gamma}}, & \tau > \tau_0 \end{cases} \tag{3}$$

where $\boldsymbol{\tau}$ is the viscous stress tensor, $\dot{\boldsymbol{\gamma}} \equiv \nabla \mathbf{u} + (\nabla \mathbf{u})^T$ is the rate of strain tensor, $\mathbf{u}$ is the velocity vector, and the superscript T denotes the transpose. The magnitudes of $\dot{\boldsymbol{\gamma}}$ and $\boldsymbol{\tau}$, denoted respectively by $\dot{\gamma}$ and τ, are defined by $\dot{\gamma} \equiv \sqrt{\dot{\boldsymbol{\gamma}} : \dot{\boldsymbol{\gamma}} / 2}$ and $\tau \equiv \sqrt{\boldsymbol{\tau} : \boldsymbol{\tau} / 2}$, where the symbol II stands for the second invariant of a tensor.

The experiments showed that as the channel height decreases the flow ceases to be symmetric and slip heterogeneities effects become important [7]. Some of the experimental velocity profiles are characterized by overshoots similar to those encountered in entry flows. We thus revisited the classical flow development problem of a Bingham plastic with asymmetric wall slip, i.e. with different Navier slip conditions at the two walls. If slip occurs along both walls and slip at the upper wall is stronger, there are three flow regimes depending on the pressure gradient, as illustrated in Fig. 1. Below a certain critical value G_1 (Regime I) the fluid simply slides and the velocity is plug. Above G_1 (Regime II), the fluid yields only near the lower plate and the unyielded region extends up to the upper wall. Finally, above a second critical pressure gradient G_2 (Regime III), the fluid yields near both the walls and the velocity profile is asymmetric with a plug core. For simplicity, in the present work we assume that

along the lower wall there is no slip, as in Fig. 2. In such a case only Regimes II and III are relevant.

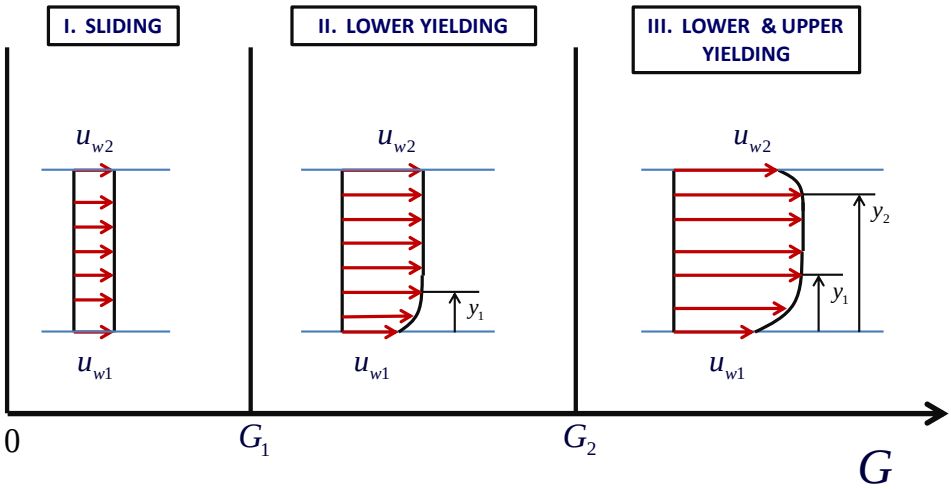


Figure 1: Different flow regimes in the case of plane viscoplastic Poiseuille flow with asymmetric slip.

In the literature, the development length is usually defined as the length required for the maximum velocity to attain 99% of its fully-developed value scaled either by the pipe diameter or the channel width [9]. This definition implies that maximum velocity develops more slowly than its counterparts at any other vertical distance which may not be the case in all geometries and for all fluids, especially viscoplastic ones which are characterized by a maximum flat velocity. For the asymmetric flow under study, we define as development length the smallest length at which the velocity at any vertical distance differs from its fully developed value by 0.01%. Hence, the velocity is required to attain either 99% or 1.01% of its fully developed value, depending on whether the latter is greater or less than unity.

The governing equations are presented in Section 2 where the solutions corresponding to fully-developed flow are also derived and the various flow regimes are identified. In Section 3, the numerical method is briefly presented and preliminary results are discussed. Finally, the conclusions are summarized in Section 4.

2 GOVERNING EQUATIONS

The governing equations are de-dimensionalized scaling lengths by the gap height H of the channel, the velocity vector by the uniform inlet velocity U, and the pressure and the stress tensor components by kU^n/H^n. If we denote de-dimensionalized variables with stars, then the continuity and momentum equations for steady incompressible flow with zero gravity read:

$$\nabla^* \cdot \mathbf{u}^* = 0 \tag{4}$$

and

$$Re\,\mathbf{u}^* \cdot \nabla^* \mathbf{u}^* = -\nabla^* p + \nabla^* \cdot \boldsymbol{\tau}^* \tag{5}$$

respectively, where $Re \equiv \rho U^{2-n} H^n / k$ is the Reynolds number, ρ being the constant density of the material.

The standard finite element method and the Pananastasiou regularization [8] are used in order to solve the flow numerically. The dimensionless form of the Papanastasiou-regularized constitutive equation is written as follows:

$$\boldsymbol{\tau}^* = \left[Bn \frac{1-\exp(-M\dot{\gamma}^*)}{\dot{\gamma}^*} + \dot{\gamma}^{*n-1} \right] \dot{\boldsymbol{\gamma}}^* \tag{6}$$

where $Bn \equiv \tau_0 H^n / (kU^n)$ is the Bingham number and $M \equiv mU/H$ is the dimensionless growth exponent, which has to be sufficiently high so that the flow of the ideal discontinuous Herschel-Bulkley fluid is approximated satisfactorily [8].

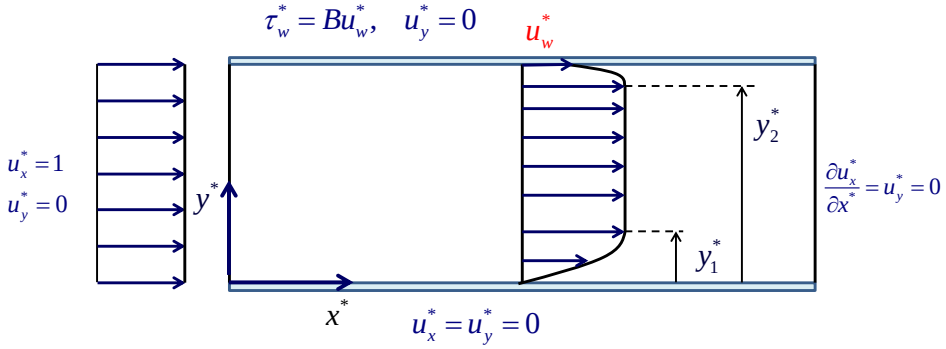


Figure 2: Geometry and boundary conditions of the flow development of a Bingham plastic in a channel with slip along the upper wall.

The geometry and the boundary conditions of the flow are illustrated in Fig. 2. At the inlet plane, the velocity component in the direction of the flow is uniform ($u_x^* = 1$) and the transversal one vanishes. At the lower wall there is no slip and no penetration so that both velocity components are zero. At the upper wall the vertical velocity is again zero slip is assumed to occur following a power-law slip equation,

$$\tau_w^* = Bu_w^{*s} \tag{7}$$

where $B \equiv \beta H^n / (kU^{n-s})$ is the (dimensionless) slip number. Finally, the exit plane is taken sufficiently far downstream so that the flow can be assumed fully developed. For the low Reynolds number considered here, we took L_{mesh}=20.

2.1 Fully-developed solutions

The non-dimensionalization introduced above is based on the mean velocity, which implies that there is flow, i.e. Regime I is excluded. Due to the definition of the slip number, the no-slip case, which corresponds to a symmetric velocity profile with respect to the midplane of the channel, is recovered for $B \to \infty$. Obviously, this solution, which involves two symmetric yield points at y_1^* and $y_2^* = 1 - y_1^*$ falls into Regime III. Keeping the Bingham number constant and decreasing the slip number, enhances slip at the upper wall and the velocity becomes asymmetric: the two yield points move towards the upper wall (that is both y_1^* and y_2^* increase) and the size of the plug region ($y_2^* - y_1^*$) increases while its velocity is reduced. This trend continues till a critical slip number, B_c, at which the upper yield point reaches the wall (the dimensionless upper wall shear stress is equal to *Bn*) signaling the transition from Regime III to Regime II. The general dimensionless solution for $B_c \le B < \infty$ is given by

$$u_x^*(y^*)=\begin{cases}\dfrac{1}{A_{III}}\left[y_1^{*1/n+1}-(y_1^*-y^*)^{1/n+1}\right], & 0\le y^*\le y_1^*\\ \dfrac{y_1^{*1/n+1}}{A_{III}}, & y_1^*\le y^*\le y_2^*\\ u_w^*+\dfrac{1}{A_{III}}\left[(1-y_2^*)^{1/n+1}-(y^*-y_2^*)^{1/n+1}\right], & y_2^*\le y^*\le 1\end{cases} \tag{8}$$

where

$$u_w^*=\frac{1}{A_{III}}\left[y_1^{*1/n+1}-(1-y_2^*)^{1/n+1}\right] \tag{9}$$

and

$$A_{III}=y_1^{*1/n+1}\left(1-\frac{n}{1+2n}y_1^*\right)-\frac{n}{1+2n}(1-y_2^*)^{1/n+2} \tag{10}$$

The positions of the two yield points can be found by solving the following system of equations:

$$(2-y_1^*-y_2^*)Bn-(y_2^*-y_1^*)Bu_w^{*s}=0 \tag{11}$$

and

$$(1+1/n)^n(y_2^*-y_1^*)-2Bn\,A_{III}^n=0 \tag{12}$$

No-slip case

In the no-slip case ($u_w^*=0$) Eq. (9) yields

$$y_2^*=1-y_1^* \tag{13}$$

which indicates that the flow is symmetric with respect to the midplane of the channel. Substituting into Eq. (10) gives

$$A_{III}=y_1^{*1/n+1}\left(1-\frac{2n}{1+2n}y_1^*\right) \tag{14}$$

and Eq. (12) then becomes:

$$(1+1/n)^n(1-2y_1^*)-2Bn\,y_1^{*n+1}\left(1-\frac{2n}{1+2n}y_1^*\right)^n=0 \tag{15}$$

Critical value of the slip number

The critical value B_c of the slip number can be found by setting $y_2^*=1$. Let us denote the corresponding critical values of y_1^* and u_w^* by y_{1c}^* and u_{wc}^*, respectively. From Eq. (11) we get

$$B_c u_{wc}^{*s}=Bn \tag{16}$$

which simply says that the upper wall shear stress is equal to *Bn*. The critical slip velocity is given by

$$u_{wc}^{*} = \frac{1}{1-\frac{n}{1+2n} y_{1c}^{*}} \tag{17}$$

and, therefore,

$$B_c = \left(1-\frac{n}{1+2n} y_{1c}^{*}\right)^{s} Bn \tag{18}$$

Finally, from Eq. (12) we get

$$(1+1/n)^{n}(1-y_{1c}^{*}) - 2Bn\, A_{IIIc}^{n} = 0 \tag{19}$$

or

$$(1+1/n)^{n}(1-y_{1c}^{*}) - 2Bn\, y_{1c}^{*n+1} / u_{w}^{*n} = 0 \tag{20}$$

which is used to calculate y_{1c}^{*}. It should be noted that the value of y_{1c}^{*} is independent of the slip equation parameters. For example, in the Bingham plastic case (n=1), y_{1c}^{*} is a root of

$$Bn\, y_{1c}^{*3} - 3Bn\, y_{1c}^{*2} - 3y_{1c}^{*} + 3 = 0 \tag{21}$$

while the value of B_c can then be calculated from Eq. (18) for any value of s.

Solution in Regime II

If the slip number is further reduced, the yield point keeps moving towards the upper wall and the size of the plug is thus reduced while its velocity increases. Finally, in the limit $B = 0$ (full slip), the velocity profile corresponds to the no-slip solution in a channel of double width (2*H)*, i.e. to the no-slip solution corresponding to the Bingham number

$$Bn' = \frac{\tau_0 (2H)^{n}}{kU^{n}} = 2^{n} Bn \tag{22}$$

Hence, when $0 < B \le B_c$, the flow corresponds to Regime II and the dimensionless velocity is given by

$$u_x^{*}(y^{*}) = \begin{cases} \left[y_1^{*1/n+1} - (y_1^{*} - y^{*})^{1/n+1} \right] / \left[y_1^{*1/n+1} \left(1-\frac{n}{1+2n} y_1^{*}\right) \right], & 0 \le y^{*} \le y_1^{*} \\ u_w^{*}, & y_1^{*} < y^{*} \le 1 \end{cases} \tag{23}$$

where

$$u_w^{*} = \frac{1}{1-\frac{n}{1+2n} y_1^{*}} \tag{24}$$

and y_1^{*} is the root of

$$(1+1/n)^{n}(1-y_1^{*}) - \left(Bn + Bu_w^{*s}\right) y_1^{*n+1} / u_w^{*n} = 0 \tag{25}$$

Substituting Eq. (16) into the above equation yields Eq. (20) for y_{1c}^*. For n=1 (Bingham plastic) and s=1 (Navier slip) Eq. (25) is simplified to

$$Bn\, y_1^{*3} - 3(Bn + B)\, y_1^{*2} - 6y_1^* + 6 = 0 \tag{26}$$

Figure 3 illustrates the two flow regimes on the (Bn,B) plane in the case of a Bingham plastic (n=1), separated by the curve $B_c = (1 - y_{1c}^*/3)Bn$, which is slightly below the straight line $B = Bn$. Four representative velocity profiles, obtained taking Bn=1 and assuming Navier slip (s=1), are also shown. Two of them are in Regime III. The first profile corresponds to no-slip at both walls ($B \to \infty$) and it is thus symmetric. As slip at the upper wall is enhanced (e.g., for B=5), symmetry is destroyed and the two yield points move upwards and the maximum velocity decreases. The upper yield point moves faster than the lower one reaching the wall when $B = B_c$. The velocity profile for this critical case is also shown in Fig. 3. Below that number, i.e., in Regime II, the yield point continues moving upwards as slip is increased, but the maximum velocity is now increasing. In the limit of B=0, the maximum velocity is lower than that for $B \to \infty$, since it corresponds to the no-slip flow for a Bingham number equal to $Bn' = 2^n Bn = 2Bn$.

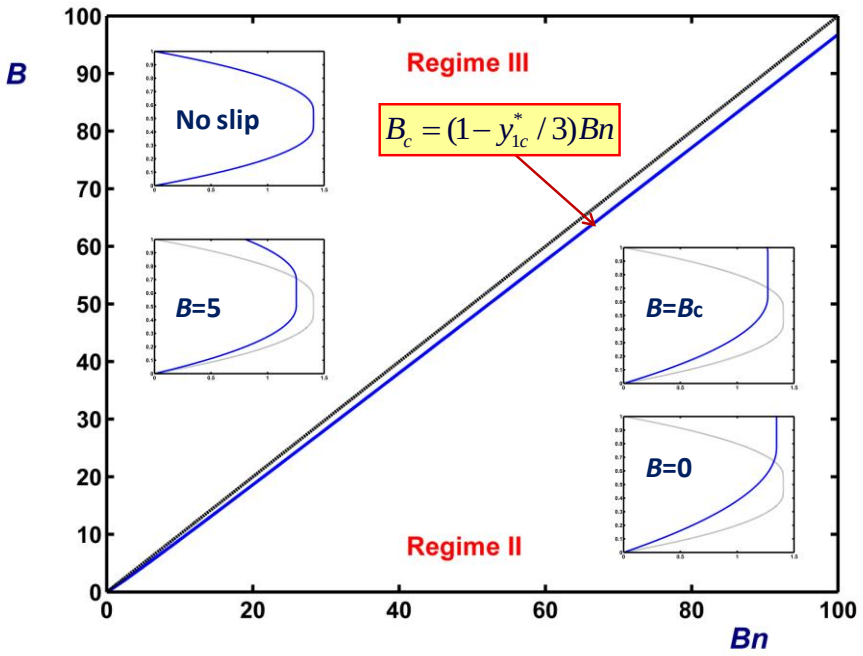
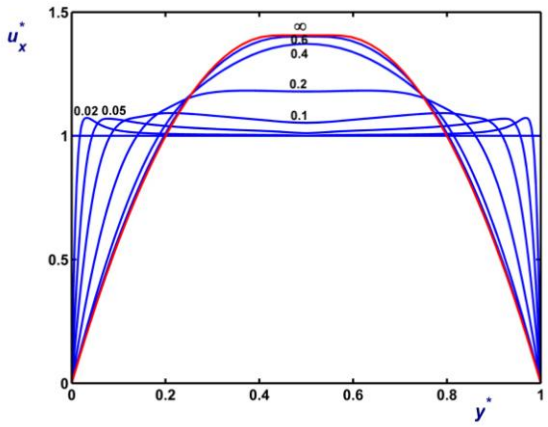


Figure 3: Flow regimes and representative velocity profiles in Bingham plastic (n=1) flow with Navier slip (s=1). The velocity profiles have been obtained for Bn=1 and various slip numbers.

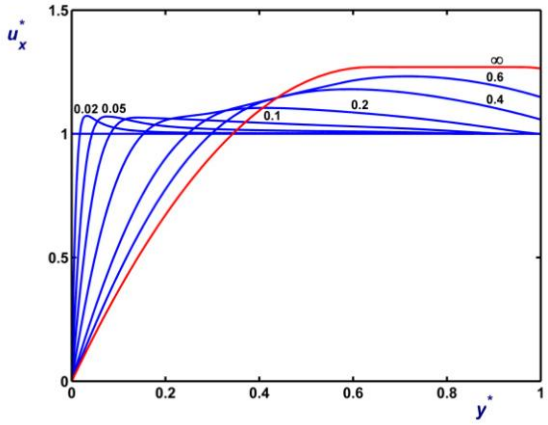
3 PRELIMINARY NUMERICAL RESULTS

The system of the governing equations and the boundary conditions presented in Section 2 was solved numerically using the finite element method (*u-v-p* formulation) with standard bi-quadratic basis functions for the two velocity components and bilinear ones for the pressure field. The Galerkin forms of the continuity and the momentum equations were used. The resulting nonlinear system of the discretized equations was solved with a Newton-Raphson iterative scheme with a convergence tolerance equal to 10^{-4}. Results have been obtained first for a Bingham plastic (n=1) with Bn=1 and various values of the slip number B. Figure 4 shows

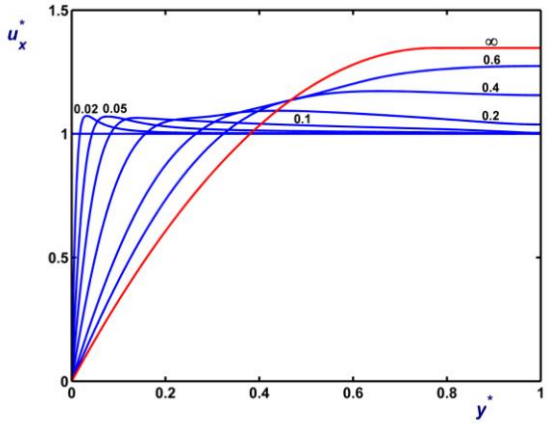
how the velocity component in the flow direction develops downstream attaining the fully developed profile. When $B=\infty$ (no slip), the velocity profiles are symmetric exhibiting a central unyielded region. As slip is increased, asymmetry is enhanced and the unyielded region moves towards upper wall and increases in size. If slip becomes even stronger then Regime II is eventually reached, i.e. the unyielded region reaches the upper wall. This is almost achieved for $B=1$, as revealed in Fig. 5, where the velocity contours are plotted.



(a)



(b)



(c)

Figure 4: Development of the velocity in creeping flow ($Re=0$) of a Bingham plastic ($n=1$) with $Bn=1$ and various slip numbers: (a) $B=\infty$ (no-slip); (b) $B=1$; (c) $B=0.01$.

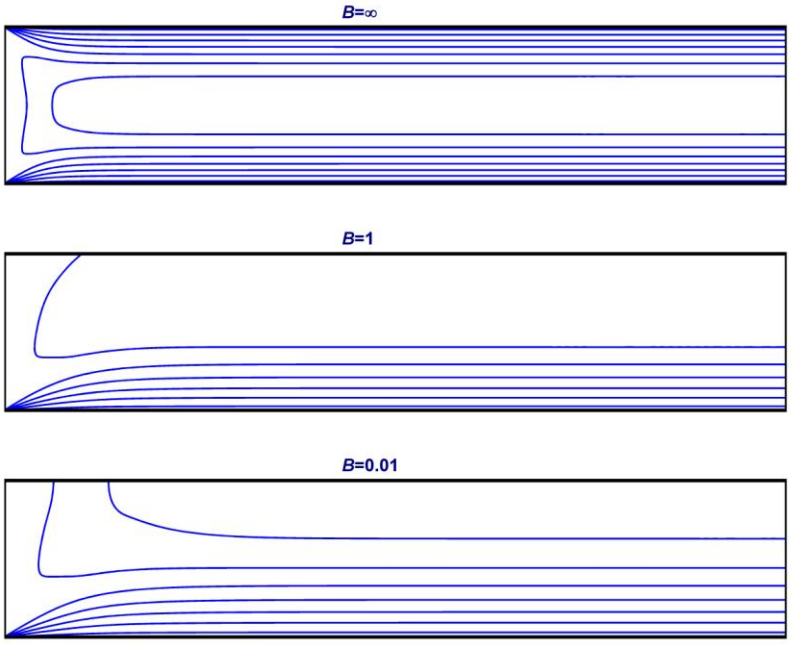


Figure 5: Velocity contours in flow development of a Bingham plastic (n=1) with Bn=1 and B=∞ (no-slip), 1 and 0.01; Re=0, s=1.

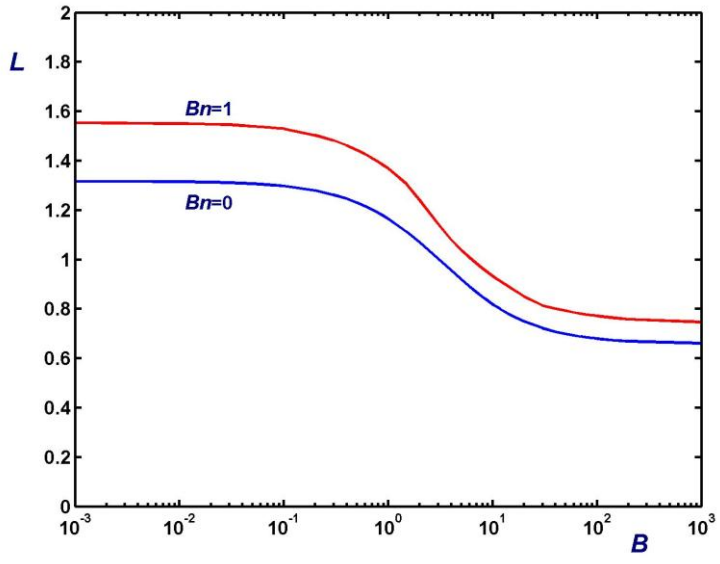


Figure 6: Development length for Bn=0 (Newtonian) and 1 versus the slip number; Re=0, n=1, s=1.

In Fig. 6, the development lengths computed for Bn=0 (Newtonian flow) and 1 are plotted versus the slip number B. Both curves are sigmoidal and the development length increases with the Bingham number. It should be noted that the development length for Newtonian flow (Bn=0) and full slip (i.e. B=0) is two times the development length for the no-slip case (infi-

nite *B*), since it corresponds to the flow development in a channel with no slip and with a gap width equal to 2. This is not the case for Bingham flow; the development length in the full slip case is two times the no-slip development length corresponding not to *Bn*=1 but to *Bn*=2.

4 CONCLUSIONS

We have solved numerically the entry flow of a Herschel-Bulkley fluid in a channel with asymmetric wall slip using finite elements and a regularized constitutive equation. We have considered in particular the case where there is slip only on the upper wall, identified the different flow regimes, and derived the fully-developed solutions. The effects of the Bingham and slip numbers on the development of the velocity and the development length have been discussed.

REFERENCES

[1] U. Yilmazer, D.M. Kalyon, Slip effects in capillary and parallel disk torsional flows of highly filled suspensions. *Journal of Rheology*, **33**, 1197-1212, 1989.

[2] H.A. Barnes, A review of the slip (wall depletion) of polymer solutions, emulsions and particle suspensions in viscometers: Its cause, character, and cure. *Journal of Non-Newtonian Fluid Mechanics*, **56**, 221-251, 1995.

[3] P. Ballesta, G. Petekidis, L. Isa, W.C.K. Poon, R. Besseling, Wall slip and flow of concentrated hard-sphere colloidal suspensions. *Journal of Rheology*, **56**, 1005-1037, 2012.

[4] M.M. Denn, Extrusion instabilities and wall slip. *Annual Reviews in Fluid Mechanics*, **33**, 265-287, 2001.

[5] D.M. Kalyon, Apparent slip and viscoplasticity of concentrated suspensions, *Journal of Rheology*, **49**, 621-640, 2005.

[6] C.L.M.H. Navier, Sur les lois du mouvement des fluides. Mémoires de l'Academie des Sciences de l'Institute de France, **6**, 389-440, 1827.

[7] A.L. Vayssade, C. Lee, E. Terriac, F. Monti, M. Cloitre, P. Tabeling, Dynamical role of slip heterogeneities in confined flows. *Physics Reviews*, **89**, 052309, 2014.

[8] T.C. Papanastasiou, Flows of materials with yield, *Journal of Rheology*, **31**, 385-404, 1987.

[9] R. Shah, A. London, *Laminar flow forced convection in ducts: a source book for compact heat exchanger analytical data*, Academic Press, 1978.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SQUEEZE FLOW OF VISCOPLASTIC BINGHAM MATERIAL

Larisa V. Muravleva[1]

[1] Lomonosov Moscow State University
Leninskie gory, 1, Main Building 119991 Moscow Russia
e-mail: lvmurav@gmail.com

Keywords: Viscoplastic Fluid, Squeeze Flow, Augmented Lagrangian Method.

Abstract. *We consider both planar and axisymmetric squeeze flows of a viscoplastic medium. Firstly we deal with no-slip boundary conditions. The asymptotic and the numerical solutions are developed. Previous theoretical analysis of this problem, using the standard lubrication approximation, has led to conflicting results, whereby the material around the plane of symmetry must both behaves as unyielded solid and translates in the main direction with a nonuniform velocity. This variation of the velocity implies that the plug region cannot be truly unyielded. Our solutions show that this region is a pseudo-plug region in which the leading order equation predicts a plug, but really it is weakly yielded at higher order. We follow the asymptotic technique suggested earlier by Balmforth, Craster (1999) and Frigaad, Ryan (2004). The obtained analytical expressions and numerical results are in a very good agreement with the earlier works. For numerical simulations we apply Augmented Lagrangian method (ALM) that yields superior results regarding the location of the yieldsurface. Finite-difference method on staggered grids is used as a discretization technique.*

Secondly we consider squeeze flow of a Bingham fluid subject to wall slip. If the wall shear-stress is smaller than the threshold value (the slip yield stress), the fluid adherence to the boundary is imposed and we have no-slip condition. When the wall shear stress reaches the slip yield stress, the fluid slips along the boundary. We solve numerically this problem also by ALM. The different flow regimes can be observed depending on the relative values of the yield stress and the slip yield stress. More precisely, for the case when the yield stress is smaller than the slip yield stress, there exists a particular value T_{crit} such that when the slip yield stress is greater or equal to T_{crit},the material fully sticks at the wall. When the slip yield stress is less than T_{crit} and bigger than the yield stress, the fluid sticks in the central region of the wall and slips close to the outer edges of discsorplates. For the case when the yield stress is bigger than the slip yield stress the material slips on the whole boundary.

1 INTRODUCTION

Squeeze flow is of considerable interest for fluids with a yield stress. The squeeze flow problem is well-studied, detailed review can be found in [1]. The analysis of viscoplastic materials started with the work of Scott [2] and Peek [3] who used the lubrication approximation, respectively assumed with different rigid zones in the flow region. Later, Covey and Stanmore [4] gave the approximate solutions for Bingham and HerschelBulkley fluids using lubrication theory.The lubrication solutions were developed using a no-slip wall boundary condition. An interesting controversy was initiated by Lipscomb and Denn [5] who showed that the unyielded (rigid) zone obtained by using the lubrication approximation for squeeze flow of Bingham materials includes the central plane of the material which however cannot move as a rigid plug for simple kinematic reasons. Gartling and Phan-Thien [6], ODonovan and Tanner [7], and Wilson [8] used instead of the original Bingham model the biviscosity fluid model both for analytical [8] and numerical [6],[7] studies. ODonovan and Tanner [7] solved the squeeze flow problem numerically and showed that unyielded material arises only adjacent to the center of the two disks.

The excellent work by Smyrnaios and Tsamopoulos [9] provided the locations of rigid zones either qualitatively or quantitatively. They employed both the original Bingham constitutive equation and the the Papanastasiou model and showed that unyielded material could only exist around the two stagnation points of flow extending in this way the work by ODonovan and Tanner [7]. Matsoukas and Mitsoulis [10], [11] also solved numerically the squeeze flow of viscoplastic materials, for both planar and axisymmetric flow, confirming the earlier results by Smyrnaios and Tsamopoulos.

Roussel at al.[12] studied squeeze flow of a Bingham fluid by means of a variational approach [13] They divided the zone of deformation into two kinematically admissible velocity fields,central region of pure extensional flow and sheared regions adjacent to the plates. Minimizing the energy dissipation functional with respect to a parameter defining the horizontal boundary of the two zones, they obtained lower-bound solutions for Bingham materials. The work of Adams et al. [14] provided experimental evidence that the yield surfaces predicted by the lubrication solution for viscoplastic fluids are erroneous. Besides, Meeten [15] carried out both theoretical and experimental investigations of the squeeze flow of Herschel Bulkley fluid.

It has been well established that the application of the lubrication solution incorporating the Bingham and HerschelBulkley relationships leads to a profound kinematic inconsistency in the calculated velocity fields.

Balmforth and Craster [16] and Frigaard and Ryan [17] suggested the asymptotic technique that resolves the lubrication paradox and builds the consistent solution for thin layer problems. In a recent paper by Muravleva [18] the planar squeeze flow of a Bingham fluid is studied exploiting the asymptotic technique introduced in [16],[17],[19]. Besides, the results of computations in [18] show presence of unyielded regions near the two stagnation points of flow close to centers of plates (that was found earlier by many researchers) and new additional unyielded regions at the outer edge of the material (both for short and long plates). More recently Fusi at all [20] investigated these unyielded regions, adjacent to the outer edge exploiting the integral formulation of the linear momentum balance. In many instances, the boundary condition on the wall is not perfectly no-slip as has been assumed above. It is frequently observed that some viscoplastic fluids slip along boundaries under some conditions which depend on the solid-fluid interaction [21], [22]. A number of slip conditions with variable degrees of complexity have been employed in the past for modeling squeeze flows of viscoplastic materials. For

viscoplastic materials,the shear stress at the plates are usually assumed to be a constant fraction of the yield stress of the material and was consequently independent of position. Sherwood and Durban analyzed squeeze flow of a Bingham [23] and Herschel-Bulkley [24] materials under this slip wall boundary conditions by using an asymptotic expansion of the velocities in h/r, where r is a radial coordinate, and h the instantaneous distance between the plates. Adams et al. [14] presented the results of squeeze-flow experiments of a HerschelBulkley material between two rigid plates, investigated both experimentally and computationally. Stick and slip boundary conditions were simulated and, in the case of the later, a Coulombic boundary condition was adopted. Lawal and Kalyon [] considered Herschel-Bulkley materials with Navier slip boundary conditions following the lubrication approximation,and obtained numerical solutions. Karapetsas and Tsamopoulos [25] examined the transient squeeze flow of a viscoplastic material between two parallel coaxial disks. The Papanastasiou model was used. On the surface of the disk, the sip conditions are imposed. This slip model divides the wall boundary into a slip region and a no-slip region. To this end, the slip coefficient is an exponential function of the radial distance from the contact point and in this way it achieves a continuous transition between the slip and the no-slip region. Yang and Zhu [26] and later Ayadi [27] analyzed the squeeze flow of biviscous fluids with the Navier slip condition, based on lubrication analysis.

The experimental data show that slip often occurs only when a critical value of the shear-stress is reached [28].If the flow field is such that the wall shear-stress is smaller than the critical value, the fluid adherence to the boundary and we have no-slip condition. When the wall shear stress reaches the threshold value, the fluid slips along the boundary with a given friction coefficient. The boundary condition can be written as follows :

$$\begin{aligned} \tau_w = \tau_c + C_f u_w, \quad & \text{if } \tau_w > \tau_c, \\ u_w = 0, \quad & \text{if } \tau_w \leq \tau_c \end{aligned} \tag{1}$$

where u_w denote the tangential velocity, τ_w is the wall shear stress and C_f is the friction coefficient, τ_c is the critical value of the shear stress under which no slip will occur,known as the slip yield stress. If τ_c is zero, we have the usual Navier slip condition:

$$\tau_w = -C_f u_w \tag{2}$$

The no-slip and the perfect-slip cases are obtained for C_f tends to ∞ and $C_f = 0$, respectively.

The numerical implementation of slip boundary condition 1 Fortin et al. [29] applied the augmented Lagrangian method to impose a slip boundary condition on walls 1 for the round Poiseuille flow of a Bingham plastic. Huilgol [31] analyzed the variational principle of a yield-stress fluid together with a slip yield condition. Roquet and Saramito [30] investigated in detail the Poiseuille flow of a Bingham fluid in a square duct with slip yield boundary condition at the wall.They used the augmented Lagrangian algorithm. Damianou et al. [32] solved numerically the cessation of axisymmetric Poiseuille flow of a HerschelBulkley fluid under the assumption that slip occurs along the wall with slip yield stress. Damianou and Georgiou used regularized versions of both the constitutive and the slip equations along with finite elements in order to solve the steady-state flow of a Herschel–Bulkley fluid in a rectangular duct with wall slip and non-zero slip yield stress [33].

Here we shall apply the asymptotic technique introduced in [16], [17] to obtain a consistent thin-layer solution for the squeeze problem of viscoplastic material. Further we compare our analytical solution with the numerical one. The main difficulty in the numerical simulation of viscoplastic fluid flow is related to the non-differentiable form of constitutive law and inability

to evaluate the stresses in regions where the material has not yielded. There are two principal approaches that have been proposed in the literature to overcome the mathematical problem of viscoplastic fluid flow. The first one, known as regularization method, consists in approximating the constitutive equation by a smoother one. Regularization replaces the rigid zones by very viscous fluid. The yield surface is computed a posteriori and the determination of the unyielded regions relies on the Von Mises stress criterion. The criterion based on the strain rate tensor cannot be used since in the unyielded regions is never truly zero, but only close to zero. The second approach is based on the theory of variational inequalities [34]. In the latter case, the problem reduces to the minimization of a functional and a further solution of the equivalent saddle-point problem which is solved iteratively (see for a detailed review [35]). Augmented Lagrangian method (ALM) introduced by Fortin and Glowinski [36]has become the most popular in this family of methods. The main advantage of these methods is that they involve the true constitutive relation and the algorithm gives truly unyielded regions with exactly zero strain rate. ALM is generally slower than codes with regularization methods. Regularization methods produce smoother solutions for the velocity field. This smoothness allows the application of more sophisticated flow solvers and leads to faster computations. Another important computation issue is the accurate tracking of the yield surfaces. Compared to the regularization method, the augmented Lagrangian method yields superior results regarding the location of the yield surface and explorations of the plastic limit, for example, in simulation of the flow cessation or no-flow limit. According to [37], pragmatically, the choice between augmented Lagrangian and regularization is related to whether one needs to determine the position of the yield surface, or whether a reasonable approximation to the velocity field is sufficient. Since the question of the existence and the position of the unyielded regions is of particular interest in the squeeze flow, we choose ALM for validation of our asymptotic solution.

2 PROBLEM STATEMENT

The goal of this study is to obtain a consistent solution for a squeeze flow of an incompressible Bingham viscoplastic fluid for axisymmetric geometry. and $2H$ and R are the gap and radius of the plates, respectively.

The flow is governed by the dimensionless conservation equations of momentum and mass:

$$\begin{aligned} Re\Big(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial r}+w\frac{\partial u}{\partial z}\Big) &= -\frac{\partial p}{\partial r}+\varepsilon^2\frac{\partial \tau_{rr}}{\partial r}+\frac{\partial \tau_{rz}}{\partial z}+\varepsilon^2\frac{\tau_{rr}-\tau_{\theta\theta}}{r},\\ \varepsilon^2 Re\Big(\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial r}+w\frac{\partial w}{\partial z}\Big) &= -\frac{\partial p}{\partial z}+\varepsilon^2\frac{\partial \tau_{rz}}{\partial r}-\varepsilon^2\frac{\partial \tau_{rr}}{\partial z}-\varepsilon^2\frac{\partial \tau_{\theta\theta}}{\partial z}+\varepsilon^2\frac{\tau_{rz}}{r},\\ \frac{\partial u}{\partial r}+\frac{u}{r}+\frac{\partial w}{\partial z} &= 0 \end{aligned} \tag{3}$$

The z-coordinate is aligned with the layer axis, p is the pressure, τ_{ij} is the deviatoric stress tensor. The components of velocity in the(r,θ,z) directions are $(u,0,w)$. We have scaled lengths in the r and z directions differently, with the disk radius $\hat{R}$ as the horizontal length-scale and with the half the distance between the disks $\hat{H}$ as a characteristic thickness in the z-direction, respectively, $\hat{W}$ is taken as the charasteristic velocity in the transverse direction, and the velocity component in the r-direction is scaled with $\hat{U}$ ($\hat{U}\hat{H}/\hat{R}=\hat{W}$). The pressure is scaled with $\hat{\mu}\hat{U}\hat{L}/\hat{H}^2$, and time with $\hat{L}/\hat{U}$. The shear-stress components are scaled with $\hat{\mu}\hat{U}/\hat{H}$. The extensional stresses are scaled with $\hat{\mu}\hat{U}/\hat{R}$. We denote dimensional variable with a hat symbol. In the above equations, the small aspect ratio ε and Reynolds number Re are

defined by

$$\varepsilon = \frac{\hat{H}}{\hat{R}}, \quad Re = \frac{\hat{\rho}\hat{U}\hat{H}}{\hat{\mu}} \tag{4}$$

where $\hat{\rho}$ and $\hat{\mu}$ are the density and plastic viscosity(both dimensional).

The constitutive laws for a Bingham fluid are

where $\dot{\boldsymbol{\gamma}}$ is strain rate tensor

$$\dot{\gamma}_{rr} = 2\frac{\partial u}{\partial r}, \quad \dot{\gamma}_{rz} = \Big(\frac{\partial u}{\partial z} + \varepsilon^2\frac{\partial w}{\partial r}\Big), \quad \dot{\gamma}_{zz} = 2\frac{\partial w}{\partial z}, \quad \dot{\gamma}_{\theta\theta} = 2\frac{u}{r} \tag{5}$$

and τ, $\dot{\gamma}$ are the second invariants of $\boldsymbol{\tau}$ and $\dot{\boldsymbol{\gamma}}$, i.e.

$$\text{with } \tau = \sqrt{\tau_{rz}^2 + \varepsilon^2(\tau_{rr}^2 + \tau_{\theta\theta}^2 + \tau_{rr}\tau_{\theta\theta})} \tag{6}$$

$$\text{with } |\dot{\gamma}| = \sqrt{(\frac{\partial u}{\partial z})^2 + 4\varepsilon^2\Big((\frac{\partial u}{\partial r})^2 + (\frac{u}{r})^2 + \frac{\partial u}{\partial r}\frac{u}{r}\Big) + 2\varepsilon^2\frac{\partial u}{\partial z}\frac{\partial w}{\partial r} + \varepsilon^4(\frac{\partial w}{\partial z})^4} \tag{7}$$

The dimensionless number appearing above is the Bingham number, B:

$$B = \frac{\hat{\tau_0}\hat{H}}{\hat{\mu}\hat{U}} \tag{8}$$

where $\hat{\tau_0}$ is yield stress. The Bingham number represents the ratio of yield stress to viscous stress.

We consider $Re \ll 1$. Neglecting the inertial terms, 3 are replaced by:

$$\begin{aligned} -\frac{\partial p}{\partial r} + \varepsilon^2\frac{\partial \tau_{rr}}{\partial r} + \frac{\partial \tau_{rz}}{\partial z} + \varepsilon^2\frac{\tau_{rr} - \tau_{\theta\theta}}{r} = 0, \\ -\frac{\partial p}{\partial z} + \varepsilon^2\frac{\partial \tau_{rz}}{\partial r} - \varepsilon^2\frac{\partial \tau_{rr}}{\partial z} - \varepsilon^2\frac{\partial \tau_{\theta\theta}}{\partial z} + \varepsilon^2\frac{\tau_{rz}}{r} = 0, \\ \frac{\partial u}{\partial r} + \frac{u}{r} + \frac{\partial w}{\partial z} = 0 \end{aligned} \tag{9}$$

The above equations are solved together with appropriate boundary conditions.

The studied domain now becomes a the upper-right quarter of the space occupied by the material and the following on the plane conditions are imposed on the plane of a symmetry Exploiting the double symmetry of the axisymmetric squeeze flow we consider the boundary condition for part of the material contained between the r and z axes and the upper plate. On the plane of symmetry $z = 0$ (central plane) :

$$\tau_{rz} = 0, \quad w = 0 \tag{10}$$

no slip condition at the contact with the upper plate ($z = 1$):

$$u = 0, \quad w = -1 \tag{11}$$

on the axis of symmetry $r = 0$:

$$u = 0, \quad \tau_{rz} = 0 \tag{12}$$

zero tangential and normal stress condition on the free surface $r = 1$:

$$\sigma_{rr} = -p + \varepsilon^2\tau_{rr} = 0, \quad \tau_{rz} = 0 \tag{13}$$

3 ASYMPTOTIC EXPANSIONS

We now solve the equations by introducing an asymptotic expansion. First we consider shear flow near the plate for which we may find a solution through a straightforward expansion of the equations. We consider regular expansions in ε of form:

$$\begin{aligned} u &= u^0 + \varepsilon u^1 + \varepsilon^2 u^2 \ldots, \\ w &= w^0 + \varepsilon w^1 + \varepsilon^2 w^2 \ldots, \\ p &= p^0 + \varepsilon p^1 + \varepsilon^2 p^2 \ldots, \\ \tau_{ij} &= \tau_{ij}^0 + \varepsilon \tau_{ij}^1 + \varepsilon^2 \tau_{ij}^2 \ldots \end{aligned} \tag{14}$$

We substitute these expansions into the governing equations 9, and collect together terms of the same order.

The lubrication equations for the first two orders are:

$$\begin{aligned} \mathcal{O}(1) \qquad & -\frac{\partial p^0}{\partial r} + \frac{\partial \tau_{rz}^0}{\partial z} = 0, \\ & -\frac{\partial p^0}{\partial z} = 0, \\ & \frac{\partial u^0}{\partial r} + \frac{u^0}{r} + \frac{\partial w^0}{\partial z} = 0 \end{aligned} \tag{15}$$

$$\begin{aligned} \mathcal{O}(\varepsilon) \qquad & -\frac{\partial p^1}{\partial r} + \frac{\partial \tau_{rz}^1}{\partial z} = 0, \\ & -\frac{\partial p^1}{\partial z} = 0, \\ & \frac{u^1}{r} + \frac{\partial u^1}{\partial r} + \frac{\partial w^1}{\partial z} = 0 \end{aligned} \tag{16}$$

3.1 Shear region

After the solution of the two first equations of 15 we have

$$p^0 = p_0(r), \quad \tau_{rz}^0 = z p_0'(r) \tag{17}$$

where p_0 is a function only of r and the prime $()'$ represents derivative $\frac{d}{dr}$. The leading order second invariants of strain rate and stress are given by

$$\tau^0 = |\tau_{rz}^0|, \quad \dot{\gamma}^0 = \left|\frac{\partial u^0}{\partial z}\right| \tag{18}$$

Provided the yield stress is exceeded, the stress tensor components become

$$\tau_{rz}^0 = \frac{\partial u^0}{\partial z} + Bsgn(\frac{\partial u^0}{\partial z}), \quad \tau_{rz}^1 = \frac{\partial u^1}{\partial z}, \tau_{rr}^0 = 2\left(\frac{\partial u^0}{\partial r} + \frac{B\frac{\partial u^0}{\partial r}}{\left|\frac{\partial u^0}{\partial z}\right|}\right) \tag{19}$$

Since we are looking for a solution with $\frac{\partial u^0}{\partial r} < 0$ in the domain $r > 0$, $s > 0$, we get

$$\tau_{rz}^0 = \frac{\partial u^0}{\partial z} - B \tag{20}$$

Integrating the last expression and using no-slip boundary conditions 11 we have

$$u^0 = \frac{p_0'(r)}{2}(z^2 - 1) + B(z - 1) \tag{21}$$

We utilize the continuity equation and obtain

$$w^0 = -\frac{z^3 - 3z + 2}{6}\Big(p_0''(r) + \frac{p_0'(r)}{r}\Big) - \frac{B}{2r}(z - 1)^2 - 1 \tag{22}$$

We now focus on the first-order approximation. Integrating two first equations of 16 and 19, using $u^1 \mid_{z=1} = 0$, we receive:

$$p^1 = p_1(r), \quad \tau_{rz}^1 = z \cdot p_1'(r) + g(r), \quad u^1 = \frac{p_1'(r)}{2}(z^2 - 1) + g(r)(z - 1) \tag{23}$$

where p_1 is a function only of r and $g(r)$ is an unknown function of integration.

3.2 Plastic region.

At leading order, at each $r \in [0, 1]$ we have $\tau_{rz}^0 = z p_0'(r)$. The $|\tau_{rz}^0|$ exceeds its maximum value at point $z = 1$ and vanish at point $z = 0$. Hence, there exists the point $z = z_0$ at which $|\tau_{rz}^0| = B$, that means $\frac{\partial u^0}{\partial z} = 0$. But according to 18 $\tau^0 = B$ and $\dot\gamma^0 = 0$. Hence, at leading order, the yield condition holds at this point and $z_0(r) = \frac{B}{|p_0'(r)|}$ is the position of pseudo-yield surface. For $z \in [0,\, z_0]$ we have $\tau^0 < B$ and $\dot\gamma^0 = 0$.

The expression for the velocity can be written as follows:

$$u^0(r, z) = \begin{array}{ll} \frac{B}{2z_0}(1 - z_0)^2, & z \in [0,\, z_0], \\ \frac{B}{2z_0}\Big((1 - z_0)^2 - (z - z_0)^2\Big), & z \in (z_0,\, 1] \end{array} \tag{24}$$

We denote the pseudo-plug velocity by $u_0(r)$:

$$u_0(r) = u^0(z_0(r)) = \frac{B}{2z_0}(1 - z_0)^2 \tag{25}$$

To determine pseudo-yield surface $z = z_0(r)$, we use the flow rate for the channel.

$$\int_0^1 u^0(r, z)\, dz = Q(r) \equiv \frac{r}{2} \tag{26}$$

Substituting 24 into the last equation leads to the cubic equation (rewritten the Buckingham equation), where the unknown is the pseudo-yield surface $z_0(r)$

$$z_0^3 - z_0\Big(3 + \frac{3r}{B}\Big) + 2 = 0 \tag{27}$$

This equation can be solved by the method proposed in [38]. There is single root of this equation for which $0 < z_0 \le 1$ and can be written as

$$z_0(r) = -2\sqrt{1 + \frac{r}{B}}\cos\Big(\frac{1}{3}\arccos\frac{1}{\sqrt{(1 + \frac{r}{B})^3}} - \frac{2\pi}{3}\Big) \tag{28}$$

It can be seen that this zeroth order solution has the characteristic Bingham-Poiseuille profile of velocity in sheared layer. But it's evident that the plug region isn't a true plug region as the leading order velocity varies in the r-direction. This is the essence of the lubrication paradox for yield stress fluids.

The source of the problem is revealed by the diagonal components of the stress. From the above expression for u_0 we can easily see that $\frac{\partial u^0}{\partial r}$ will not in general vanish while $\frac{\partial u^0}{\partial z}$ does. In the domain $|z| \leq z_0$ we should consider the higher-order equations and pay attention to the diagonal stress components.

We assume that the domain occupied by the medium can be separated into two subdomains:

1. The external regions are situated near the plates where the yield criterion is overcome.There the shear stress is dominant. We name them "shear regions".

2. The inner region includes the centerplane, where yield criterion is not reached, the shear stress is smaller and in the centerplane equal to zero. The flow is close to extensional flow. We call this region "pseudoplug" or "plastic region".

These regions are separated by an interface represented by the smooth surface pseudo-plug $z = z_0(r)$.

Let us consider the domain near the centerplane of thickness $0 < z < z_0$. Below the fake yield surface, $z = z_0(r)$, the asymptotic expansion described above breaks down. To find a solution appropriate to this region we look for a slightly different set of asymptotic sequences. The principal difference is in the expansion of the horizontal velocity component:

$$\begin{aligned} u &= u^{p,0}(r) + \varepsilon u^{p,1}(r,z) + \varepsilon^{p,2} u^2(r,z) + \dots, \\ w &= w^{p,0}(r,z) + \varepsilon w^{p,1}(r,z) + \varepsilon^2 w^{p,2}(r,z) + \dots, \\ p &= p^{p,0}(r,z) + \varepsilon p^{p,1}(r,z) + \varepsilon^2 p^{p,2}(r,z) + \dots \end{aligned} \tag{29}$$

where the property that $\frac{\partial u^{p,0}}{\partial z} = 0$ at z_0 is explicitly built in. We use the same symbol for the pseudo-plug solution. When we match the two solutions, we add a subscript p to explicitly distinguish the pseudo-plug solution.

Thus, by introducing an asymptotic expansion, we find

$$\begin{aligned} \tau_{rr}^{p,-1} &= \frac{2B}{\dot\gamma^{p,0}} \frac{\partial u^{p,0}}{\partial r}, \\ \tau_{rz}^{p,0} &= \frac{B}{\dot\gamma^{p,0}} \frac{\partial u^{p,1}}{\partial z}, \\ \tau_{\theta\theta}^{p,-1} &= \frac{2B}{\dot\gamma^{p,0}} \frac{u^{p,0}}{r} \end{aligned} \tag{30}$$

where the leading order term in the rate of strain tensor is

$$\dot\gamma^{p,0} = \sqrt{\Big(\frac{\partial u^{p,1}}{\partial z}\Big)^2 + 4\Big(\Big(\frac{\partial u^{p,0}}{\partial r}\Big)^2 + \Big(\frac{u^{p,0}}{r}\Big)^2 + \frac{\partial u^{p,0}}{\partial r}\frac{u^{p,0}}{r}\Big)} \tag{31}$$

The lubrication equations for the first two orders are:

$$\begin{aligned} \mathcal{O}(1) \qquad & -\frac{\partial p^{p,0}}{\partial r} + \frac{\partial \tau_{rz}^{p,0}}{\partial z} = 0, \\ & -\frac{\partial p^{p,0}}{\partial z} = 0, \end{aligned} \tag{32}$$

$$\mathcal{O}(\varepsilon) \qquad \begin{aligned} &\frac{\partial u^{p,0}}{\partial r}+\frac{u^{p,0}}{r}+\frac{\partial w^{p,0}}{\partial z}=0\\ &-\frac{\partial p^{p,1}}{\partial r}+\frac{\partial \tau_{rr}^{p,-1}}{\partial r}+\frac{\tau_{rr}^{p,-1}-\tau_{\theta\theta}^{p,-1}}{r}+\frac{\partial \tau_{rz}^{p,1}}{\partial z}=0,\\ &-\frac{\partial p^{p,1}}{\partial z}-\frac{\partial(\tau_{rr}^{p,-1}+\tau_{\theta\theta}^{p,-1})}{\partial z}=0,\\ &\frac{\partial u^{p,1}}{\partial r}+\frac{u^{p,1}}{r}+\frac{\partial w^{p,1}}{\partial z}=0 \end{aligned} \tag{33}$$

Using symmetry about the centerplane, we have $w^p\mid_{z=0}=0$, $\tau_{rz}^p\mid_{z=0}=0$, $\frac{\partial u^p}{\partial z}\mid_{z=0}=0$. Integrating the first two equations of 32 and enforcing continuity p^0, τ_{rz}^0, u^0 at $z=z_0(r)$ leads to

$$p^{p,0}(r)=p_0(r),\quad p_0'=-\frac{B}{z_0},\quad \tau_{rz}^{p,0}=-\frac{Bz}{z_0},\quad u^{p,0}=u_0(r)=\frac{B}{2z_0}(1-z_0)^2 \tag{34}$$

From the last equation of 32 we obtain

$$w^{p,0}=-z\Big(u_0'+\frac{u_0}{r}\Big) \tag{35}$$

Differentiation of the expressions 27,34 after minor manipulations gives

$$z_0'=\frac{3z_0^2}{2B(z_0^3-1)}\quad u_0'=\frac{3(1+z_0)}{4(1+z_0+z_0^2)} \tag{36}$$

The continuity condition of $w^{p,0}$ and $w^{s,0}22$ at pseudo yield surface is automatically satisfied by substituting the expression for z_0', u_0', r 27,36.

From 30,34 we obtain

$$\begin{aligned} &(\tau^{p,-1})^2=(\tau_{rr}^{p,-1})^2+(\tau_{\theta\theta}^{p,-1})^2+\tau_{rr}^{p,-1}\tau_{\theta\theta}^{p,-1}+(\tau_{rz}^{p,0})^2=\\ &\frac{4B^2}{(\dot\gamma^{p,0})^2}\Big((u_0')^2+(\frac{u_0}{r})^2+u_0'(r)\frac{u_0}{r}\Big)+\frac{B^2}{(\dot\gamma^{p,0})^2}(\frac{\partial u^{p,1}}{\partial z})^2=B^2 \end{aligned} \tag{37}$$

We denote

$$\eta=\sqrt{4\Big((u_0')^2+(\frac{u_0}{r})^2+u_0'\frac{u^0}{r}\Big)} \tag{38}$$

Substituting $\tau_{rz}^{p,0}=-\frac{Bz}{z_0}$ 30,34 into 37 leads to the following expression

$$\frac{B^2\eta^2}{(\dot\gamma^{p,0})^2}=\frac{B^2(z_0^2-z^2)}{z_0^2} \tag{39}$$

After minor calculation we have

$$\dot\gamma^{p,0}=\frac{\eta z_0}{\sqrt{z_0^2-z^2}} \tag{40}$$

Substituting the expression 38 into 31 gives

$$(\frac{\partial u^{p,1}}{\partial z})^2+\eta^2=(\dot\gamma^{p,0})^2 \tag{41}$$

Inserting 40 into 41 we obtain the equation for $\frac{\partial u^{p,1}}{\partial z}$ Solving this equation and integrating we obtain

$$u^{p,1} = \eta\sqrt{z_0^2 - z^2} + u_1^* \tag{42}$$

where $u_1^*(r)$ is an unknown function of integration. Obviously the function $u_1^*(r)$ is the value of the first order velocity u^1 at the pseudo-yield surface $z = z_0(r)$. We now focus on the first-order approximation. Integrating the equation **??** gives

$$p^{p,1} = -\tau_{rr}^{p,-1} - \tau_{\theta\theta}^{p,-1} + \psi(r) \tag{43}$$

where $\psi(r)$ is an unknown function of integration. Enforcing continuity of the pressure at the pseudo-yield surfaces yields $\psi(r) = p_1(r)$. Substituting 43 into the first equation of 33 gives

$$\frac{\partial}{\partial r}(2\tau_{rr}^{p,-1} + \tau_{\theta\theta}^{p,-1}) - p_1' + \frac{\tau_{rr}^{p,-1} - \tau_{\theta\theta}^{p,-1}}{r} + \frac{\partial \tau_{rz}^{p,1}}{\partial z} = 0 \tag{44}$$

Inserting 34,40 into 30 we find that

$$\tau_{rr}^{p,-1} = \frac{2B}{\eta z_0} u_0' \sqrt{z_0^2 - z^2} \quad , \tau_{\theta\theta}^{p,-1} = \frac{2B}{\eta z_0} \frac{u_0}{r} \sqrt{z_0^2 - z^2} \tag{45}$$

Substituting $\tau_{rr}^{p,-1}$ and $\tau_{\theta\theta}^{p,-1}$ into 44 leads to

$$\begin{aligned} 2B\frac{d}{dr}\Big(\frac{2u_0' + \frac{u_0}{r}}{\eta z_0}\Big)\sqrt{z_0^2 - z^2} + 2B\Big(\frac{2u_0' + \frac{u_0}{r}}{\eta z_0}\Big)\frac{z_0' z_0}{\sqrt{z_0^2 - z^2}} - \\ -p_1' + \frac{2B(u_0' - \frac{u_0}{r})}{r\eta z_0}\sqrt{z_0^2 - z^2} + \frac{\partial \tau_{rz}^{p,1}}{\partial z} = 0 \end{aligned} \tag{46}$$

After integrating (in respect that $\tau_{rz}^{p,1}\mid_{z=0} = 0$)

$$\begin{aligned} \tau_{rz}^{p,1} = -B(z\sqrt{z_0^2 - z^2} + z_0^2 \arcsin\frac{z}{z_0})\Big(\frac{d}{dr}\Big(\frac{2u_0' + \frac{u_0}{r}}{\eta z_0}\Big) + \frac{(u_0' - \frac{u_0}{r})}{r\eta z_0}\Big) - \\ -2B\Big(\frac{2u_0' + \frac{u_0}{r}}{\eta z_0}\Big) z_0' z_0 \arcsin\frac{z}{z_0} + p_1' z \end{aligned} \tag{47}$$

Enforcing continuity of the first order shear stress τ_{rz} at the pseudo-yield surface leads to

$$g(r) = -\frac{\pi B}{2}\Big(\frac{d}{dr}\Big(\frac{z_0}{\eta}(2u_0' + \frac{u_0}{r})\Big) + \frac{z_0}{\eta r}(u_0' - \frac{u_0}{r})\Big) \tag{48}$$

The unknown $u_1^*(r)$ can be determined by matching the first order velocities at $z = z_0$

$$\begin{aligned} u_1^* &= \frac{p_1'}{2}(z_0^2 - 1) + g(z_0 - 1), \\ u^s(r,z) &= -\frac{B}{2z_0}(z^2 - 1) + B(z - 1) + \varepsilon(\frac{p_1'}{2}(z^2 - 1) + g(z - 1)), \\ u^p(r,z) &= \frac{B}{2z_0}(1 - z_0^2) + \varepsilon(\eta\sqrt{z_0^2 - z^2} + \frac{p_1'}{2}(z_0^2 - 1) + g(z_0 - 1)) \end{aligned} \tag{49}$$

However, function p_1 is still unknown. To find p_1, we consider the flowrate constraint:

$$\int_0^1 u(r,\, z)\, dz = \int_0^{z_0} u^p\, dz + \int_{z_0}^1 u^s\, dz =$$
$$\frac{B(z_0-1)^2(z_0+2)}{6z_0} + \varepsilon\Big(\frac{z_0^2\pi\eta}{4} + \frac{p_1'}{3}(-1+z_0^3) + \frac{g}{2}(-1+z_0^2)\Big) = \frac{r}{2} \tag{50}$$

$Q(r) = Q_0(r) + \varepsilon Q_1(r)$. We require that $Q = Q_0$, $Q_1 = 0$. Thus we obtain the equation for the border z_0:

$$z_0^3 - z_0\Big(3 + \frac{3r}{B}\Big) + 2 = 0 \tag{51}$$

which is equivalent to 27 and expression for p_1'

$$p_1' = \frac{3z_0^2\eta\pi}{4(1-z_0^3)} - \frac{3g(1+z_0)}{2(1+z_0+z_0^2)} = -\frac{\pi B\eta z_0'}{2} - 2gu_0' \tag{52}$$

As we have mentioned above the numerical modeling [10] shows the presence of the rigid zones near the stagnation points at the center of the plate. It is interesting to investigate the obtained asymptotic solution near the stagnation point. We will examine the second invariant of the stress τ at the plate at $z = 1$.

The leading order τ^0 in the shear region equals the shear stress $\tau^0(r,1) = |\tau_{rz}^0(r,1)| = \frac{B}{z_0(x)}$ and exceed the yield stress because $z_0(r) \le 1$. Thus, in accordance with the zero-order solution, the Bingham fluid in the shear region is yielded.

Then we take into account the first order approximation and consider the stress along the $z = 1$

$$\tau(r,1) = \tau^0(r,1) + \varepsilon\tau^1(r,1) + O(\varepsilon^2) = |\tau_{rz}^0(r,1) + \varepsilon\tau_{rz}^1(r,1) + O(\varepsilon^2)| = \frac{B}{z_0(r)} - \varepsilon(p_1'(r) + g(r)) + O(\varepsilon^2),$$

where $p_1'(r)$ and $g(r)$ are determined by 52, 48 To analyze the behavior of the function $\tau_{rz}^0(r,1)$ near the axis of symmetry, we use the Taylor series. We introduce the new variable $\tilde{\zeta} = \left(\frac{r}{B}\right)^{1/2}$ and expand the function $z_0(r(\tilde{\zeta}))$, using 28 in Taylor series. After that, we return to the variable r and obtain

$$z_0(r) = 1 - \sqrt{\frac{r}{B}} + \frac{r}{3B} + \frac{r}{18B}\sqrt{\frac{r}{B}} - \frac{r^2}{27B^2} - \frac{5r^2}{216B^2}\sqrt{\frac{r}{B}} + \frac{r^3}{243B^3} \tag{54}$$

$$\tau_{rz}^0 = -B(1 + \sqrt{\frac{r}{B}} + \frac{2r}{3B} + \frac{5r}{18B}\sqrt{\frac{r}{B}} + \frac{r^2}{27B^2} - \frac{11r^2}{216B^2}\sqrt{\frac{r}{B}} - \frac{13r^3}{243B^3} + O(r)^{7/2}) \tag{55}$$

$$\tau_{rz}^1 = \frac{\pi}{\sqrt{3}}(\frac{3\sqrt{B}}{4\sqrt{r}} - \frac{67}{192} - \frac{581\sqrt{r}}{2304\sqrt{B}} + O(r)) \tag{56}$$

The first term in the brackets on the right-hand side of 56 is positive and tends to infinity when $r \to 0$ and the other terms are finite for $0 \le r \le 1$. For $r \to 0$ we have $-\tau_{rz}(r,1) = \frac{B}{z_0(r)} - \varepsilon(p_1'(r) + g(r)) \to -\infty$. This means that near $r = 0$ there exists a point $r = r_0$ at which $-\tau_{rz} = B$ and $\tau = B$. For $r \le r_0$ we have $\tau < B$ and the material is unyielded.

Figure 1:

So, asymptotic analysis predicts a region of unyielded material near the plane. Consequently our asymptotic expansion breaks down. We give in Fig. 1(a) the magnitude of the tangential stress $-\tau_{rz}(r,1)$, acting on the disc surface for $B=1$, $\varepsilon=0.1$. The solid line corresponds to $\tau_{rz}(r,1)=\tau^0_{rz}(r,1)+\varepsilon\tau^1_{rz}(r,1)$, dotted line — tangential stress zero-order $\tau^0(r,1)$, dashed line — $z=B$. The solid line is below the dotted line and approaches it for $r\to 1$. The abscissa of the point of intersection of the solid line and a dotted horizontal line is the radius yield surface on the disk ($\tau=B$). Fig. 1(b) shows the stress intensity along the disk surface for $\varepsilon=0.1$ and four Bingham numbers, 0.1, 1,10,100. The solid line corresponds to $\tau(r,1)=\tau^0(r,1)+\varepsilon\tau^1(r,1)$, dotted line — $\tau^0(r,1)$. The uppermost curve in this figure is for the highest Bn value, they decrease gradually as Bn decreases . The stresses are increase monotonically with increasing r and decrease sharply near the axis of symmetry $r=0$. The stresses equal to the corresponding Bingham numbers are marked points on the solid lines. The yield surfaces (i.e. the surfaces on which $\tau=B$) are marked points on the solid lines. As expected, the size of the unyielded spot increases with B.(This figure qualitatively follow fig14a presented in)This figure is in qualitative agreement with fig.14a presented in [9].

4 Slip condition

We consider another boundary condition at the contact with the upper plate ($z=1$):

$$\hat{u}=\begin{array}{ll} -\Big(1-\frac{\hat{\tau}_c}{|\hat{\tau}_{rz}|}\Big)\frac{\hat{\tau}_{rz}}{\hat{c}_f}, & |\hat{\tau}_{rz}|>\hat{\tau}_c,\\ 0, & otherwise\end{array} \tag{57}$$

here $\hat{s}_0$ is the slip yield stress, $\hat{c}_f$ is . The slip yield dimensionless number B_w is defined as the ratio of $\hat{s}_0$ to a characteristic stress $\hat{\mu}\hat{U}/\hat{R}$: $B_w=\hat{s}_0\hat{R}/\hat{\mu}\hat{U}$, the friction dimensionless number C_f is defined by $C_f=\hat{c}_f\hat{R}/\hat{\mu}$. We introduce $\beta=1/C_f$. The three dimensionless numbers B, B_w and β characterize the squeeze problem with the stick-slip law at the wall. We can rewrite in dimensionless form:

$$u=\begin{array}{ll} -\beta\tau_{rz}\Big(1-\frac{B_w}{|\tau_{rz}|}\Big), & |\tau_{rz}|>B_w,\\ 0, & otherwise\end{array} \tag{58}$$

This slip model divides the wall boundary into the slip region and no-slip region. The stick and the slip regions are separated by the transition point r_T. β is a parameter used to adjust the level of the slip velocity in comparison to the wall shear and a B_w is a second parameter used to adjust the length of the slip region. The choice $\beta=0$ leads naturally to a no-slip boundary condition, whereas when B_w becomes zero this model reduces to the standard Navier slip model and slip occurs over the entire wall boundary with a constant slip coefficient.

In this section we will assume, that the condition $|\tau_{rz}|>B_w$ is satisfied, therefore the velocity boundary condition becomes the following:

$$u\mid_{z=1}=-\beta(\tau_{rz}-B_w\frac{\tau_{rz}}{|\tau_{rz}|}),\quad w\mid_{z=1}=-1. \tag{59}$$

4.1 Shear region

Following section 2 we look for a solution in which the main variables of the problem can be expressed as power series of ε, namely

$$u=u^0+\varepsilon u^1+\varepsilon^2u^2\ldots,$$

$$w = w^0 + \varepsilon w^1 + \varepsilon^2 w^2 \dots, \quad (60)$$
$$p = p^0 + \varepsilon p^1 + \varepsilon^2 p^2 \dots,$$
$$\tau_{ij} = \tau_{ij}^0 + \varepsilon \tau_{ij}^1 + \varepsilon^2 \tau_{ij}^2 \dots$$

We start from the zero-order approximation. In the shear region the equations 15 remain valid and we have

$$p^0 = p_0(r), \quad \tau_{rz}^0 = z p_0'(r). \quad (61)$$

Instead of the boundary conditions 11, we have

$$u^0 = \beta(-p_0' - B_w), \quad w = -1. \quad (62)$$

From 18,20,62 we find

$$u^0 = \frac{p_0'}{2}(z^2 - 1) + B(z - 1) + \beta(-p_0' - B_w) \quad (63)$$

We utilize the continuity equation from15 (last one), the second boundary condition 62 and obtain

$$w^0 = -\frac{z^3 - 3z + 2}{6}\Big(p_0'' + \frac{p_0'}{r}\Big) - \frac{B}{2r}(z - 1)^2 + \beta\Big(p_0'' + \frac{p_0' + B_w}{r}\Big)(z - 1) - 1. \quad (64)$$

Figure 2:

Just as in Section 2, we will consider two subdomains: the shear regions (near the plates), where the yield criterion at leading order is overcome ($\tau^0 = |\tau_{rz}^0| = |z p_0'(r)| > B$), and the plastic region (surrounding the centerplane), where the yield criterion is not reached. These regions are separated by pseudo-yield surface $z_0(r) = \frac{B}{|p_0'(r)|}$. The expression for the velocity $u^0(r, z)$ can be written as follows:

$$u^0(r, z) = \begin{array}{ll} \frac{B}{2z_0}(1 - z_0)^2 + \beta(\frac{B}{z_0} - B_w), & z \in [0,\, z_0], \\ -\frac{B}{2z_0}(z^2 - 1) + B(z - 1) + \beta(\frac{B}{z_0} - B_w), & z \in (z_0,\, 1]. \end{array} \quad (65)$$

To determine pseudo-yield surface $z = z_0(r)$, we use the flow rate for the channel

$$\int_0^1 u^0\, dz = \frac{B(z_0 - 1)^2(z_0 + 2)}{6z_0} + \beta(\frac{B}{z_0} - B_w) = \frac{B(z_0^3 - 3z_0 + 2) + 6\beta(B - B_w z_0)}{6z_0} = \frac{r}{2}. \quad (66)$$

The rearrangement of 66 gives the cubic equation where the unknown is the pseudo-yield surface z_0:

$$z_0^3 - 3z_0\Big(1 + \frac{r}{B} + 2\beta\frac{B_w}{B}\Big) + 2 + 6\beta = 0. \quad (67)$$

We denote

$$f_0(z_0) := z_0^3 - 3z_0(1 + \frac{r}{B} + 2\beta\frac{B_w}{B}) + 2 + 6\beta. \quad (68)$$

As $f_0(0) = 2+6\beta > 0$, $f_0(1) = -3\frac{r}{B} - 6\beta(\frac{B_w}{B} - 1) < 0$, then for $r > 0$, $z_1 < 0 < z_2 < 1 < z_3$. Thus for all $r > 0$, $\beta > 0$,$B_w > B$ on the interval $(0,\ 1)$ equation 67 has the only root:

$$z_0(r) = -2\sqrt{1 + \frac{r}{B} + 2\beta\frac{B_w}{B}}\cos\Big(\frac{1}{3}\arccos\Big(\frac{1+3\beta}{\sqrt{(1+\frac{r}{B}+2\beta(\frac{B_w}{B}))^3}} - \frac{2\pi}{3}\Big)\Big). \tag{69}$$

Now, differentiating z_0 and u_0 with respect to r, we obtain

$$z_0' = -\frac{3z_0^2}{2B(1+3\beta-z_0^3)}, \quad u_0' = \frac{3(1+2\beta-z_0^2)}{4(1+3\beta-z_0^3)}. \tag{70}$$

Exploiting the continuity equation we find w^0 in the plastic region. Finally, we obtain

$$w^0(r,z) = \begin{cases} -z(u_0' + \frac{u_0}{r}), & z \in [0,\ z_0], \quad (71) \\ (z^3 - 3z + 2 - 6\beta(z-1))\Big(-\frac{Bz_0'}{6z_0^2} + \frac{B}{6rz_0}\Big) - \frac{B}{2r}(z^2-1) + \beta\frac{B_w}{r}(z-1) - 1, & z \in (z_0, \quad (72) \end{cases}$$

Using formulae 66, 70, it is easy to see that the axial velocity w^0 is continuous on the pseudo-yield surface $z_0(r)$.

The boundary conditions 62 hold for $\frac{B}{z_0} > B_w$, otherwise no-slip boundary conditions hold. The condition $|\tau_{rz}^0(r_*, 1)| = B_w$ defines $r = r_*$ as the "stick-slip" transition point of zero-order . Obviously, $z_0^* = z_0(r_*) = B/B_w$, thus, in accordance with 66,

$$r_* = \frac{1}{3}(\frac{B^3}{B_w^2} - 3B + 2B_w). \tag{73}$$

We denote $\tilde{z_0}(r)$ pseudo-yield surface, corresponding to the equation 67, notation $z_0(r)$ remains for the solution of the equation 27. We observe 36, 70 that $z_0' < 0$, $\tilde{z_0'} < 0$, so the pseudo-yield surfaces $z = z_0(r)$, $z = \tilde{z_0}(r)$ are the decreasing functions of radius. If $z_0(r_*) = B/B_w < z_0(1)$, then the curves $z_0(r)$, $\tilde{z_0}(r)$ intersect at $r_* > 1$ (horizontal line $z = B/B_w$ is below $z_0(r)$ at $0 \le r \le 1$). Thus, the material adheres to the entire surface of the disks. It is evident that $z_0(r_*) = \tilde{z_0}(r_*)$. In order to determine the relative position of the graphs $z_0(r)$ and $\tilde{z_0}(r)$, we subtract the expression 27 from 67 and obtain:

$$(z_0 - \tilde{z_0})(3 + \frac{3r}{B} - (z_0^2 + z_0\tilde{z_0} + \tilde{z_0}^2)) = \frac{6\beta B_w}{B}(\tilde{z_0} - \frac{B}{B_w}). \tag{74}$$

If $r < r_*$, then $\tilde{z_0} - \frac{B}{B_w} > 0$ and $z_0 > \tilde{z_0}$ and vice versa (the second factor on the left side is positive, since $z_0, \tilde{z_0} \in (0,\ 1])$.

Fig. 2 (a) shows the graphs of the pseudo-yield surfaces $z = z_0(r)$, $z = \tilde{z_0}(r)$ for three values of the Bingham number $B = 0.1,\ 1,\ 100$. The black lines correspond to the pseudo-yield surfaces $z = z_0(r)$ with no-slip condition. For the slip condition pseudo-yield surfaces $z = \tilde{z_0}(r)$ are painted by color lines: red lines correspond to $B_w/B = 1.2$, blue line - $B_w/B = 1.5$, turquoise line - $B_w/B = 2$. We can see that the size of the stick region enlarges monotonically as the ratio B_w/B increases (B is fixed); consequently, the smaller is the B_w/B, the higher is the corresponding curve $z = \tilde{z_0}(r)$. We observe that for $B = 10$ slip occurs only for $B_w = 12$. For $B_w = 15$ and $B_w = 20$ no-slip condition is performed on the entire surface of the disks. For $B = 100$ and for all three $B_w = 120, 150, 200$ the no-slip condition is valid on the whole surface of the disks.

On Fig. 2(b) we have plotted $z = z_0(r)$, $z = \tilde{z_0}(r)$ for $B = 1$, $B_w = 2$ and various slip coefficients β. As mentioned above, this parameter adjust the level of the local slip velocity in comparison to the wall shear. More specifically, when the value of β increases the value of slip velocity grows. It can be easily seen that for the small values of β the curve $\tilde{z_0}(r)$ is close to the curve $z_0(r)$. When β becomes bigger the graph of the function $\tilde{z_0}(r)$ is close to a horizontal line in accordance with 70.

Let us return to Fig. 1 (a) and remember that the absolute value of shear stress $|\tau_{rz}|$ is smaller than $|\tau^0_{rz}|$, so $|\tau^0_{rz}(r_*)| = B_w$, but $|\tau_{rz}(r_*)| < B_w$. Accounting for the first approximation for the shear stress leads to decreasing of the absolute magnitude of the shear stress, as can be clearly seen from Fig. 1. Therefore, at the point r_* shear stress is above the critical B_w and actually slippage occurs at point r_T ($r_T > r_*$). Thus the fulfillment of the slip condition for zero approximation does not guarantee the fulfillment of this condition if we take into consideration the first-order approximation. There is a transition region (r_*, r_T) where $|\tau^0_{rz}| > B_w$ and $|\tau^0_{rz} + \varepsilon\tau^1_{rz}| < B_w$. We call this case fake slip and consider later.

In this subsection we deal with the case $|\tau_{rz}| > B_w$. If the condition $|\tau^0_{rz} + \varepsilon\tau^1_{rz}| > B_w$ is fulfilled then the slip occurs and we have the following boundary condition for the first approximation:

$$u^1 \mid_{z=1} = -\beta\tau^1_{rz}. \tag{75}$$

Equation 75 can be written in the general form

$$u^1 \mid_{z=1} = u^1_b, \tag{76}$$

where

$$u^1_b = -\beta\tau^1_{rz}. \tag{77}$$

The equations 16 remain valid but instead of no-slip conditions 11 we consider 76. We solve these equations as in the subsection 2.1 and have the following expressions:

$$p^1 = p_1(r), \quad \tau^1_{rz} = z \cdot p_1'(r) + g(r), \quad u^1 = \frac{p_1'(r)}{2}(z^2 - 1) + g(r)(z - 1) + u^1_b. \tag{78}$$

4.2 Plastic region.

In the plastic region we introduce the asymptotic expansion 29. The expressions 30–31 hold. For the plastic region the solution repeats the solution for the plastic region with no-slip condition. Instead of expressions 34, we obtain the following items:

$$p^0(r) = p_0(r), \quad p_0' = -\frac{B}{z_0}, \quad \tau^0_{rz} = -\frac{Bz}{z_0}, \quad u^0 = u_0(r) = \frac{B}{2z_0}(1 - z_0)^2 + \beta(\frac{B}{z_0} - B_w). \tag{79}$$

The equations 37–47 are valid, but for the functions z_0, z_0', u_0, u_0' we use the expressions 69,70,79 from the previous subsection. Given the continuity of pressure and stress on the pseudo-yield surface we obtain:

$$\begin{aligned}
p^s(r) = &\quad p_0(r) + \varepsilon p_1(r) + \dots, \quad p^p(r) = p_0(r) + \varepsilon(p_1(r) - \frac{2B}{\eta z_0}(u_0' + \frac{u_0}{r})\sqrt{z_0^2 - z^2}) + \dots, \\
\tau^s_{rz}(r) = &\quad -\frac{Bz}{z_0} + \varepsilon(zp_1'(r) + g(r)) + \dots, \\
\tau^p_{rz}(r, z) = &\quad -\frac{Bz}{z_0} + \varepsilon\Big(p_1' z + \frac{2g}{\pi}\arcsin\frac{z}{z_0} - Bz\sqrt{z_0^2 - z^2}\Big(\frac{d}{dr}\Big(\frac{2u_0' + \frac{u_0}{r}}{\eta z_0}\Big) + \frac{u_0' - \frac{u_0}{r}}{r\eta z_0}\Big)\Big) + \dots,
\end{aligned} \tag{80}$$

where

$$g(r) = -\frac{\pi B}{2}\Big(\frac{d}{dr}\Big(\frac{z_0}{\eta}(2u_0' + \frac{u_0}{r})\Big) + \frac{z_0}{\eta r}(u_0' - \frac{u_0}{r})\Big). \tag{81}$$

From continuity of u^1 at $z = z_0$ it follows that

$$u^{1s} = \frac{p_1'}{2}(z^2-1)+g(z-1)+u_b^1, \quad u^{1p} = \eta\sqrt{z_0^2 - z^2}+u_1^* \Rightarrow u_1^* = \frac{p_1'}{2}(z_0^2-1)+g(z_0-1)+u_b^1.$$

Substituting 78 into expression 77 gives

$$u_b^1 = -\beta(p_1'(r) + g(r)). \tag{82}$$

We can express the velocity in terms of $z_0(r)$:

$$\begin{aligned} u^s(r,z) &= -\tfrac{B}{2z_0}(z^2-1) + B(z-1) + \beta(\tfrac{B}{z_0} - B_w) + \varepsilon(\tfrac{p_1'}{2}(z^2-1) + g(z-1) + u_b^1), \\ u^p(r,z) &= \tfrac{B}{2z_0}(1-z_0^2) + \beta(\tfrac{B}{z_0} - B_w) + \varepsilon(\eta\sqrt{z_0^2 - z^2} + \tfrac{p_1'}{2}(z_0^2-1) + g(z_0-1) + u_b^1). \end{aligned}$$

To find p_1 we consider the flow rate through the gap:

$$\int_0^1 u(r,\, z)\, dz = \int_0^{z_0} u^p\, dz + \int_{z_0}^1 u^s\, dz =$$

$$\frac{B(z_0^2 - 3z_0 + 2) + 6\beta(B - B_w z_0)}{6z_0} + \varepsilon\Big(\frac{z_0^2\pi\eta}{4} + \frac{p_1'}{3}(-1+z_0^3) + \frac{g}{2}(-1+z_0^2) + u_b^1\Big) = \frac{r}{2}. \tag{83}$$

Just as in the subsection 2.2 we obtain expression for p_1':

$$p_1' = \frac{3z_0^2\pi\eta}{4(1-z_0^3)} - \frac{3g(1-z_0^2)}{2(1-z_0^3)} + \frac{3u_b^1}{(1-z_0^3)}. \tag{84}$$

Using 82 we have:

$$p_1' = \frac{3z_0^2\eta\pi}{4(1+3\beta-z_0^3)} - \frac{3g(1+2\beta-z_0^2)}{2(1+3\beta-z_0^3)} = -\frac{\pi B\eta z_0'}{2} - 2gu_0'. \tag{85}$$

The coordinate of the transition point r_T can be found from the fact that at the transition point the shear stress is equal to the critical value $-B_w$: $\tau_{rz}^0(r_T) + \varepsilon\tau_{rz}^1(r_T) = -B_w$. We find the root of the next equation $z_0(r)$ numerically:

$$\frac{B}{z_0} + \varepsilon(\frac{\pi B\eta(z_0)z_0'}{2} + g(z_0)(2u_0'(z_0) - 1)) = B_w, \tag{86}$$

and then substitute it into the formula 66 and define r_T.

4.3 Fake slip

Let us return to the fake slip case. There is an interval (r_*, r_T), for which at the boundary $z = 1$ no-slip condition is satisfied, i.e., velocity at the boundary should be zero. Since the zero-order approximation of the velocity at the boundary is positive, this value should be compensated by the boundary condition for the first approximation. Therefore:

$$u_b^1 = -\frac{\beta}{\varepsilon}(\frac{B}{z_0} - B_w) \tag{87}$$

Figure 3:

In both regions, fake slip and slip, the solutions of the first-order are described by the same formulae except the expression for the function u_b^1 in the right-hand side of the boundary condition. The expression 84 after the substitution of 87 takes the following form:

$$p_1' = \frac{z_0^2 \eta \pi 3}{4(1-z_0^3)} - \frac{3g(1+z_0)}{2(1+z_0+z_0^2)} - \frac{3\beta}{\varepsilon(1-z_0^3)}(\frac{B}{z_0} - B_w). \tag{88}$$

Figure 3 shows the graphs of the shear stress for Bingham number $B = 1$ and different wall Bingham numbers $B_1 = 1.5$, 2, 2.5. The purple line corresponds to τ_{rz}^0 - zero approximation with no-slip condition, the black line corresponds to $\tau_{rz} = \tau_{rz}^0 + \varepsilon\tau_{rz}^1$ with no-slip condition. The solutions at the fake slip and the slip regions are depicted in color: red line corresponds to $B_w = 1.5$, blue line - $B_w = 2$, green line - $B_w = 2.5$. The asterisks on r-axis mean transition from the no-slip region to the fake slip region. In other words, they are the abscissas $r = r_*$ of the intersection points of the purple line and the dotted lines $z = B_w$ ($|\tau_{rz}^0(r_*)| = B_w$). The color of the asterisk corresponds to the value of B_w. The circles on r-axis mean transition from the fake slip region to the slip region, namely, they are the abscissas $r = r_T$ of the intersection points of the dotted line $z = B_w$ and the corresponding color line. Thus, each graph is composed from three parts: for $r \leq r_*$ it is solution of the no-slip problem, for $r_* < r < r_T$ it is solution of the fake slip problem, for $r \geq r_T$ it is solution of the slip problem. One can easily see that rising of B_w entails shortening of the fake slip region. The slip condition reduces the absolute value of the shear stress compared with the stick condition. The smaller is ratio B_w/B (for fixed B), the sooner τ_{rz} exceeds the value B_w and the closer to the axis of symmetry is the fake slip region. We note that for case when the fake slip region is located near the axis of symmetry (for example, $B_w = 1.5$), the absolute value of the shear stress in this region (red line) is slightly less in comparison with solution with no-slip condition (black line). If the fake slip region is located in the middle ($B_w = 2$) or closer to the edge of the disk ($B_w = 2.5$), the shear stress in the fake slip region coincide with solution with no-slip condition. In the slip region shear stresses are nearly the linear functions and visually the slopes are almost identical.

In Figure 3(b) we examine the influence of the parameter β on the value of the shear stress with fixed $B = 1$. We choose the color of the lines as well as in the previous figure, while the solid line corresponds to $\beta = 1$, dashed line corresponds $\beta = 0.5$, dotted line –$\beta = 0.1$. The fluid velocity at the disks surfaces has positive value, which increases as β increases. Since the mass flow rate is kept constant, as we vary the slip coefficient β, the increase in the fluid velocity in the disk surfaces region will result in decreased velocity gradients there, and in the center region the fluid will be correspondingly slowed down as β is increased, otherwise the mass conservation requirement will be violated.A reduction in velocity gradient will necessarily lead to a decrease in shear stress, as indicated in 3(b) As expected, for the small values of β ($\beta = 0.1$) the graph is close to the line, corresponding to the no-slip solution.

The fluid velocity at the disks surfaces has positive value, which increases as β increases. Since the mass flow rate is kept constant, as we vary the slip coefficient β, the increase in the fluid velocity in the disk surfaces region will result in decreased velocity gradients there, and in the center region the fluid will be correspondingly slowed down as β is increased, otherwise the mass conservation requirement will be violated. A reduction in velocity gradient will necessarily lead to a decrease in shear stress , as indicated in Fig. 6b, the pressure gradient sharply

decreases with increased p at any radial coordinate.

We see that the location of the transition point r_T is independent of β when $B_w = 2$ and $B_w = 2.5$. For $B_w = 1.5$ magnitude of the shear stress in the fake region increases with decreasing of β, so the critical value B_w is reached at smaller radius, therefore, the transition points r_T are moved to the left. But the difference between them is negligibly small and these points are almost indistinguishable.

Figure 4: Velocity profiles .

Figure 4 shows the radial velocity profiles: red lines correspond to the no-slip solution and black lines correspond to the slip solution, solid lines are used for composite solution, dotted lines - for zero-order solution. Fig. 4 (a) corresponds to $B = 10$, $B_w = 13$, $\varepsilon = 0.1$, radial location is $r = 0.9$. This location belongs to the slip region. We see that the stick solution(corresponding to the no-slip condition) has zero velocity at the disk while the slip solution has positive velocity at the disk. Wherein the velocity magnitude for zero-order approximation is bigger than the composite solution. In consequence of the flow rate conservation, slip solution velocity at the plane of symmetry ($z = 0$) is reduced compared with stick solution. The axial distributions of the radial velocity profiles at selected radial coordinates are shown in Fig. 4(b),(c). Fig. 4 (b) corresponds to $B = 10$, $B_w = 13$, $\varepsilon = 0.05$, radial locations are (from left to right) $r = 0.6, 0.7, 0.8, 0.9, 1$. The first profile is located in the no-slip region, thus solutions coincide and we see the only red line. Next position $r = 0.7$ is located in the fake slip region. The red and black contours are almost identical, excluding small deviation near the pseudo-yield surface. The locations $r = 0.8, 0.9, 1$ are in the slip region. It is easy to see that under the slip condition the velocity profiles are compressed in the radial direction compared with no-slip case. At any radial coordinate r, in the region close to the disks surfaces the velocity of the fluid which corresponds to the slip condition, is more than the velocity corresponding to the no-slip condition. Since the same mass flow rate is imposed, in order to satisfy the mass-conservation constraint, the increase in fluid velocity in the wall regions will be compensated for by the decrease in fluid velocity as the center region.

Fig. 4 (c) corresponds to $B = 1$, $B_w = 2$, $\varepsilon = 0.05$, radial locations are (from left to right) $r = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$. The first profile belongs to the no-slip region, in all other positions the slip condition is performed.

5 NUMERICAL SOLUTION

In order to check our asymptotic solution were performed the simulation for the squeeze flow of viscoplastic Bingham media. This section is dedicated to the solution of the system 3 using an Augmented Lagrangian method.This method to solve the problem of Bingham fluid flow is detailed in a previous papers.3–8 We denote $\Omega = [-L, L] \times [-H, H]$, $\Gamma = \partial\Omega$,

$$\Gamma_u = \{(x, y) \mid -L \le x \le L, y = \pm H\}; \Gamma_s = \Gamma \setminus \Gamma_u; \tag{89}$$

$$\boldsymbol{u}_V = \begin{cases} (0, -V) & \text{if } y = H, \quad (90) \\ (0, V) & \text{if } y = -H. \quad (91) \end{cases}$$

We introduce following convex set

$$\boldsymbol{U}_V = \{\boldsymbol{u} \in (H^1(\Omega))^2 \mid \boldsymbol{u} = \boldsymbol{u}_V \text{ on } \Gamma_u, \nabla \cdot \boldsymbol{u} = 0\} \tag{92}$$

and the functional $J : \boldsymbol{U}_V \to R$ such that

$$J(\boldsymbol{u}) = \frac{\mu}{2}\int_\Omega |\boldsymbol{\gamma}(\boldsymbol{u})|^2 d\boldsymbol{x} + \tau_0 \int_\Omega |\boldsymbol{\gamma}(\boldsymbol{u})| d\boldsymbol{x}. \tag{93}$$

The set $\boldsymbol{U}_V$ consists of function in $(H^1(\Omega))^2$ that are divergence free and for which the essential boundary conditions are satisfied. The solution of problem 3 – 8 expressed as a minimization point of J on $\boldsymbol{U}_V$:

$$J(\boldsymbol{u}) = \min_{\boldsymbol{v}\in\boldsymbol{U}_V} J(\boldsymbol{v}). \tag{94}$$

The difficulty in solving problem 93, 94 is related to non-differentiability of the right-hand size second term of 93

We introduce the following functional space $\boldsymbol{Q} = \{\boldsymbol{q} |\ \boldsymbol{q} \in (L^2(\Omega))^4; \boldsymbol{q}^T = \boldsymbol{q}\}$, the additional variable

$$\boldsymbol{q} = \boldsymbol{\gamma}(\boldsymbol{v}) \in \boldsymbol{Q}. \tag{95}$$

The constraint 95is relaxed thanks to the introduction of Lagrange multiplier that coincides with the stress $\boldsymbol{\tau} \in \boldsymbol{Q}$. The augmented Lagrangian $L_r : \boldsymbol{U} \times \boldsymbol{Q} \times \boldsymbol{Q} \to R$ is given by

$$\mathcal{L}_r(\boldsymbol{u};\boldsymbol{q};\boldsymbol{\tau}) = \frac{\mu}{2}\int_\Omega |\boldsymbol{q}|^2\, d\boldsymbol{x} + \tau_0 \int_\Omega |\boldsymbol{q}|\, d\boldsymbol{x} + \frac{1}{2}\int_\Omega (\boldsymbol{\gamma}(\boldsymbol{u}) - \boldsymbol{q}) : \boldsymbol{\tau}\, d\boldsymbol{x} + \frac{r}{2}\int_\Omega |\boldsymbol{\gamma}(\boldsymbol{u}) - \boldsymbol{q}|^2\, d\boldsymbol{x}, \quad r \geq 0. \tag{96}$$

The minimization problem 93, 94 is replaced by saddle point problem of $\mathcal{L}_r$ which we can solve by the following Uzawa-type algorithm proposed in [**?**]

$$\{\boldsymbol{q}^0, \boldsymbol{\tau}^1\} \text{ are given } \in \boldsymbol{Q} \times \boldsymbol{Q}; \tag{97}$$

for $n = 0, 1, 2 \ldots$, assume $\boldsymbol{q}^n$ and $\boldsymbol{\tau}^n$ are being known, perform the next steps:

Step 1: find $\boldsymbol{u}^{n+1}$ e p^{n+1} such that:

$$\begin{aligned} -r\triangle \boldsymbol{u}^{n+1} + \nabla p^{n+1} &= \nabla\cdot(\boldsymbol{\tau}^n - r\boldsymbol{q}^n) \text{ in } \Omega, \\ \nabla\cdot\boldsymbol{u}^{n+1} &= 0 \text{ in } \Omega, \\ r\boldsymbol{\gamma}(\boldsymbol{u}^{n+1})\cdot\boldsymbol{n} &= (r\boldsymbol{q}^n - (-p^{n+1}\boldsymbol{I} + \boldsymbol{\tau}^n))\cdot\boldsymbol{n} \text{ on } \Gamma_s, \\ \boldsymbol{u}^{n+1} &= \boldsymbol{u}_V \text{ on } \Gamma_u. \end{aligned}$$

Step 2: compute explicitly $\boldsymbol{q}^{n+1}$ as

$$\boldsymbol{q}^{n+1} := \begin{array}{ll} 0, & \text{if } |\boldsymbol{\tau}^n + r\boldsymbol{\gamma}(\boldsymbol{u}^{n+1})| < \tau_0, \\ (1 - \frac{\tau_0}{|\boldsymbol{\tau}^n + r\boldsymbol{\gamma}(\boldsymbol{u}^{n+1})|})\frac{\boldsymbol{\tau}^n + r\boldsymbol{\gamma}(\boldsymbol{u}^{n+1})}{(r+\mu)}, & \text{otherwise.} \end{array} \tag{98}$$

Step 3: explicit update of the Lagrange multiplier $\boldsymbol{\tau}^{n+1}$

$$\boldsymbol{\tau}^{n+1} := \boldsymbol{\tau}^n + r(\boldsymbol{\gamma}(\boldsymbol{u}^n) - \boldsymbol{q}^n). \tag{99}$$

If $|\boldsymbol{\tau}^{n+1} - \boldsymbol{\tau}^n| > \varepsilon$ for $\varepsilon > 0$, go to step 1.

The advantage of this algorithm is that it transforms the non-differentiable problem 93, 94 into a family of a modified Stokes problem with an additional right hand side 98 and local explicit calculations.

The system of equations is discretized in space by finite differences on a uniform staggered grid with meshsize $h = h_1 = h_2$, where h_1 and h_2 are the step discretization in each direction. The discrete values of the pressure are located at the center of each cell and the velocity components u and v are located at the middle of the cell faces. The components γ_{11}, γ_{22}, q_{11}, q_{22}, τ_{11}, τ_{22} of strain-rate and Lagrange multipliers tensors are located at the cell centers whereas the non-diagonal components γ_{12}, q_{12}, τ_{12} ($\gamma_{12} = \gamma_{21}$) are located at the grid nodes. This grid, called Marker And Cell (MAC), allows to write first derivatives with a second-order centered scheme [**?**, **?**].

6 CONCLUSIONS

We have obtained the consistent asymptotic solution which is free from lubrication paradox. We provide expressions for velocity, stress,strain. The numerical solution that we have computed using the augmented Lagrangian method has been compared favorably with the asymptotic solutions, providing validation of the approach.

REFERENCES

[1] J. Engmann, C. Servais, A. Burbidge, Squeeze flow theory and applications to rheometry: A review. *Journal of Non-Newtonian Fluid Mechanics*, **132**, 1–27, 2005.

[2] J.R. Scott, Theory and application of the parallel-plate plastometer. *Trans. Inst. Rubber Ind.*, **7**, 169–186, 1931.

[3] R.L. Peek, Parallel-plate plastometry. *Journal of Rheology*, **3 (3)**, 345–372, 1932.

[4] G.H. Covey, B.R. Stanmore, Use of the parallel-plate plastometer for the characterisation of viscous fluids with a yield stress. *Journal of Non-Newtonian Fluid Mechanics*, **8**, 249–260, 1981.

[5] G.G. Lipscomb, M.M. Denn, Flow of Bingham fluids in complex geometries. *Journal of Non-Newtonian Fluid Mechanics*, **14**, 337–346, 1984.

[6] D.K. Gartling and N. Phan-Thien, A numerical simulation of a plastic fluid in a parallel-plate viscometer. *Journal of Non-Newtonian Fluid Mechanics*, **14**, 347–360, 1984.

[7] E.J. O'Donovan, R.I. Tanner, Numerical study of the Bingham squeeze film problem. *Journal of Non-Newtonian Fluid Mechanics*, **15**, 75–83, 1984.

[8] S.D.R. Wilson, Squeezing flow of a Bingham material. *Journal of Non-Newtonian Fluid Mechanics*, **47**, 211–219, 1993.

[9] D.N. Smyrnaios, J.A. Tsamopoulos, Squeeze flow of Bingham plastics. *Journal of Non-Newtonian Fluid Mechanics*, **100**, 165–190, 2001.

[10] A. Matsoukas, E. Mitsoulis, Geometry effects in squeeze flow of Bingham plastics. *Journal of Non-Newtonian Fluid Mechanics*, **109**, 231–240, 2003.

[11] E. Mitsoulis, A. Matsoukas, Free surface effects in squeeze flow of Bingham plastics. *Journal of Non-Newtonian Fluid Mechanics*, **129**, 182–187, 2005.

[12] N. Roussel, C. Lanos, Z. Toutou, Identification of Bingham fluid flow parameters using a complex squeeze test. *Journal of Non-Newtonian Fluid Mechanics*, **135**, 1–7, 2006.

[13] K.J. Zwick, P.S. Ayyaswamy, I.M.Cohen, Variational analysis of the squeezing flow of a yield stress fluid. *Journal of Non-Newtonian Fluid Mechanics*, **63**, 179–199, 1996.

[14] M.J. Adams, I. Aydin, B.J. Briscoe, S.K. Sinha, A finite element analysis of the squeeze flow of an elasto-viscoplastic paste material. *Journal of Non-Newtonian Fluid Mechanics*, **71**, 41–57, 1997.

[15] G.H. Meeten, Effect of plate roughness in squeeze flow rheometry. *Journal of Non-Newtonian Fluid Mechanics*, **124**, 51–60, 2004.

[16] N.J. Balmforth, R.V. Craster, A consistent thin-layer theory for Bingham fluids. *Journal of Non-Newtonian Fluid Mechanics*, **84**, 65–81, 1999.

[17] I.A. Frigaard, D.P. Ryan, Flow of a visco-plastic fluid in a channel of slowly varying width, *Journal of Non-Newtonian Fluid Mechanics*, **123**, 67–83, 2004.

[18] L. Muravleva, Squeeze plane flow of viscoplastic Bingham material. *J. Non-Newtonian. Fluid Mech.*, **220**, 148–161, 2015.

[19] A. Putz, I.A. Frigaard, D.M. Martinez, On the lubrication paradox and the use of regularisation methods for lubrication flows. J. Non-Newtonian Fluid. Mech. 63 (2009) 62-77.

[20] L.Fusi, A.Farina, F.Rosso, Planar squeeze flow of a bingham fluid. *Journal of Non-Newtonian Fluid Mechanics*, **225**, 1–9, 2015.

[21] H.A. Barnes, A review of the slip (wall depletion) of polymer solutions, emulsions and particle slip in viscometers - its cause, character and cure. *Journal of Non-Newtonian Fluid Mechanics*, **56**, 221–251, 1995.

[22] M.M. Denn, Extrusion instabilities and wall slip, Ann. Rev. Fluid Mech. 33, 265-287 (2001).

[23] J.D. Sherwood, D. Durban, Squeeze flow of a power-law viscoplastic solid. *Journal of Non-Newtonian Fluid Mechanics*, **62**, 35–54, 1996.

[24] J.D. Sherwood, D. Durban, Squeeze flow of a Herschel-Bulkley fluid. *Journal of Non-Newtonian Fluid Mechanics*, **77**, 115–121, 1998.

[25] G. Karapetsas, J. Tsamopoulos, Transient squeeze flow of viscoplastic materials. *Journal of Non-Newtonian Fluid Mechanics*, **133**, 35–56, 2006.

[26] S.P. Yang, K.Q. Zhu, Analytical solutions for squeeze flow of Bingham fluid with Navier slip condition. *Journal of Non-Newtonian Fluid Mechanics*, **138**, 173–180, 2006.

[27] A. Ayadi, Exact analytic solutions of the lubrication equations for squeeze-flow of a biviscous fluid between two parallel disks, *Journal of Non-Newtonian Fluid Mechanics*, **166**, 1253–1261 (2011).

[28] J.R.A. Pearson, C.J.S. Petrie, Proceedings of the Fourth International Congress on Rheology, Part 3, Wiley, New-York, 1965, pp. 265–282.

[29] A. Fortin, D. Cote, P.A. Tanguy, On the imposition of friction boundary conditions for the numerical simulation of Bingham fluid flows, Comp. Methods Appl. Mech. Eng. 88 (1991) 97-109.

[30] P. Saramito, N. Roquet, An adaptive finite element method for viscoplastic fluid flows in pipes, Comput. Methods Appl. Mech. Eng. 190 (2001) 53915412.

[31] R.R. Huilgol, Variational principle and variational inequality for a yieldstress fluid in the presence of slip, *Journal of Non-Newtonian Fluid Mechanics*, **75 (2-3)**, 231–251, 1998.

[32] Y.Damianou, M.Philippou, G. Kaoullas, G.C. Georgiou, Cessation of viscoplastic Poiseuille flow with wall slip, *Journal of Non-Newtonian Fluid Mechanics*, **203**, 24–37, 2014.

[33] Y.Damianou, G.C. Georgiou, Viscoplastic Poiseuille flow in a rectangular duct with wall slip, *Journal of Non-Newtonian Fluid Mechanics*,

[34] R. Glowinski, J.L. Lions, R. Tremolieres, Numerical Analysis of Variational Inequalities, North-Holland, Amsterdam, 1981.

[35] E.J. Dean, R. Glowinski, Operator-splitting methods for the simulation of Bingham viscoplastic flow, Chin. Ann. Math. 23 (2002) 187204.

[36] M. Fortin, R. Glowinski, Augmented Lagrangian Methods, North-Holland, 1983.

[37] N.J. Balmforth, I.A. Frigaard, Yielding to stress: recent developments in viscoplastic fluid mechanics, Annu. Rev. Fluid Mech. 46 (2014) 121146.

[38] V.I. Lebedev, On formulae for roots of cubic equation, *Soviet Journal of Numerical Analysis and Mathematical Modelling*, **4**, 315–324, 1991.

[39] L. Muravleva, Uzawa-like methods for numerical modeling of unsteady viscoplastic Bingham medium flows, Appl. Numer. Math

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PSEUDOMULTIGRID GAUSS-SEIDEL METHOD FOR LARGE SCALE AND HIGH PERFORMANCE COMPUTING

S.I. Martynenko, V.M. Volokhov and P.D. Toktaliev

The Institute of Problems of Chemical Physics of the Russian Academy of Sciences
Academician Semenov avenue 1, Chernogolovka, Moscow region, 142432 Russian Federation
e-mail: Martynenko@icp.ac.ru; vvm@icp.ac.ru; Toktaliev@icp.ac.ru

Keywords: multigrid, robustness, structured grids, parallel multigrid

Abstract. *The paper represents a new approach for development of robust multigrid methods. The approach is based on minimization of the number of problem-dependent components in multigrid algorithm. Robust Multigrid Technique (RMT) is application of the essential multigrid principle in single-grid algorithm. Problem-dependent components of the technique are:*
1) the number of smoothing iterations;
2) unknown ordering (for anisotropic problems);
3) underrelaxation parameter (for nonlinear problems);
4) stopping criterion of the multigrid iterations.
The least number of the problem-dependent components makes it possible to use RMT in black box software. Disadvantage of RMT is extra computational efforts as compared with optimized multigrid.

Since coarse grid correction is computed on the single grid, it is possible to avoid the large communication overhead and the processor idleness on the very coarse grids in parallel multigrid.

The paper represents main components of RMT, results of numerical experiments, convergence and complexity analysis, estimation of speed-up and efficiency of parallel algorithm.

1 INTRODUCTION

Classical multigrid methods are among the fastest numerical algorithms for the solution of boundary value problems. Multigrid methods are characterized by their so-called components: the smoothing procedure, the coarsening strategy, the coarse grid operators, the transfer operators from fine grids to coarse and from coarse to fine and the cycle type. These components have to be specified for each concrete problem. In *optimized multigrid* algorithms, one tries to tailor the components to the problem at hand in order to obtain the highest possible efficiency for the solution process [8]. In contrast with direct and many iterative methods, optimized multigrid offers capability of solving the boundary value problems with complexity and storage proportional to the number of unknowns.

On the other hand, the idea of *robust multigrid* algorithms is to choose the components independently of the given problem, uniformly for as large a class of problems as possible [8].

Various variants of the optimized multigrid methods have been proposed and developed [8]. Less progress has been made in development of the robust multigrid algorithms. In framework of the work a multigrid algorithm is called *robust* if
- it is efficient for a large class of problems
- it has the least number of the problem-dependent components
- it is highly parallel.

One approach consists of minimizing the number of problem-dependent components in multigrid solver. Basic idea of Robust Multigrid Technique (RMT) is to apply the essential multigrid principle (about approximation of the smooth part of the error on coarser grids [8]) in monogrid (or single-grid) algorithm [2, 3].

RMT has the advantage that the intergrid operators are trivial and problem-independent. In addition, smoothing on the single grid gives the most powerful coarse correction strategy to make the task of the smoother less demanding. As a result, weak (and well parallelizable) smoothers can be used in the multigrid technique. Block Seidel smoother can handle many applied problems in unified manner starting from the Poisson equation up to the saddle point problems. Absence of coarse grids leads to almost perfect load balance and low communication overhead in parallel multigrid. Disadvantage of RMT is extra computational efforts as compared with optimized multigrid.

Really RMT is an opportunity to accelerate convergence of Gauss-Seidel iterations at solving (non)linear boundary value problems on structures grids.

2 BASIC COMPONENTS OF ROBUST MULTIGRID TECHNIQUE

Basic principles of RMT for solving boundary value problems (BVP) will be explained by studying a one-dimensional model problem and traditional multigrid notation [5].

To overcome the problem of robustness, RMT consists of two parts: analytical (adaptation of BVPs to the technique by representation of the solution as a sum (Σ-modification) or a product (Π-modification) of two functions (coarse grid correction and approximation to the solution)) and computational (generation of the finest and coarse grids, control volume approximation of the modified BVPs on the grid hierarchy (multigrid structure) and solution of the discretized equations by original multigrid solver).

Main difficulty of the robust multigrid algorithm development is to reach close-to-optimal convergence rate with the least number of the problem-dependent components.

First stage of the computational part of RMT consists of the finest grid generation for the following control volume approximation of the modified boundary value problems. The finest

grid G_1^0 consists of two sets $G^{\rm v}(0;1)$ and $G^{\rm f}(0;1)$ of the grid points. In domain $\Omega = [0,1]$ the sets are defined as

$$G^{\rm v}(0;1) = \{x_i^{\rm v} \mid x_i^{\rm v} = h(i-1), \quad i = 1,2,\ldots,N_x^0+1, \quad h = 1/N_x^0\}\,,$$
$$G^{\rm f}(0;1) = \{x_i^{\rm f} \mid x_i^{\rm f} = 0.5\,(x_i^{\rm v} + x_{i+1}^{\rm v}), \quad i = 1,2,\ldots,N_x^0\}\,.$$

The finest grid G_1^0 with $N_x^0 = h^{-1} = 8$ is shown on Fig. 1, where h is a mesh size.

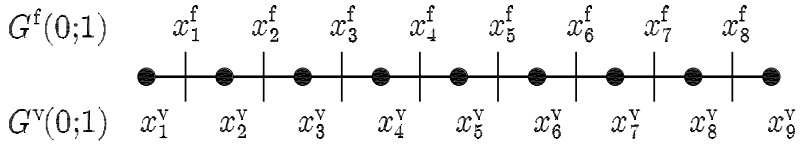


Figure 1: The finest grid G_1^0 with $N_x^0 = h^{-1} = 8$

Coarsening in RMT is based on representation of the finest grid G_1^0 as union of three coarse grids G_1^1, G_2^1 and G_3^1 as shown on Fig. 2. Hereinafter we will use standard multigrid notation for simplicity of RMT description.

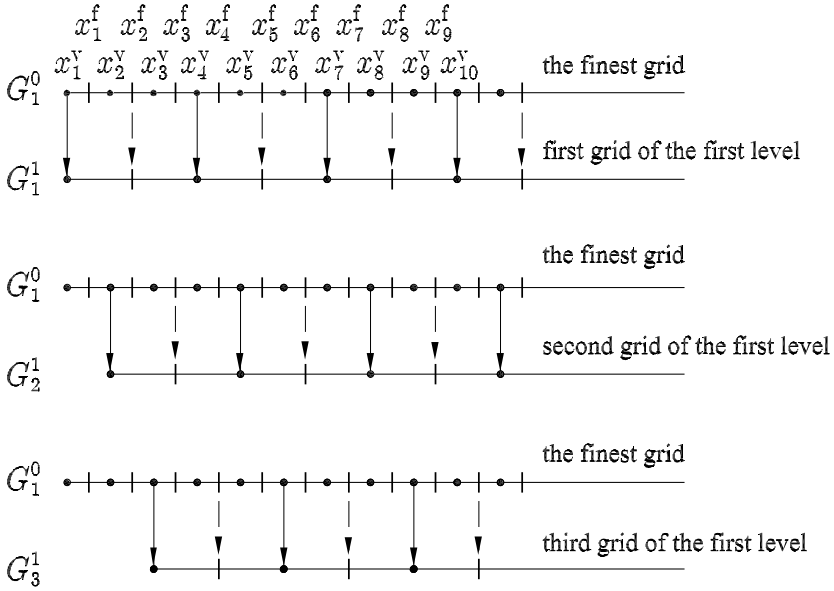


Figure 2: Triple coarsening in Robust Multigrid Technique

It is easy to see the following properties of the coarse grids:

1. the coarse grids G_1^1, G_2^1 and G_3^1 have no common points, i.e.

$$G_n^1 \cap G_m^1 = \emptyset\,, \quad n \neq m\,.$$

2. the finest grid G_1^0 is the union of the coarse grids G_1^1, G_2^1 and G_3^1, i.e.

$$G_1^0 = \bigcup_{k=1}^{3} G_k^1\,.$$

3. all grids are similar to each other, but a mesh size on the coarse grids is three times as large as the mesh size on the finest grid.

4. the functions can be assigned to the grid points x^{v} or to the grid points x^{f}, but in both cases the control volume on the coarse grids is union of three control volumes on the finest grid (Fig. 3).

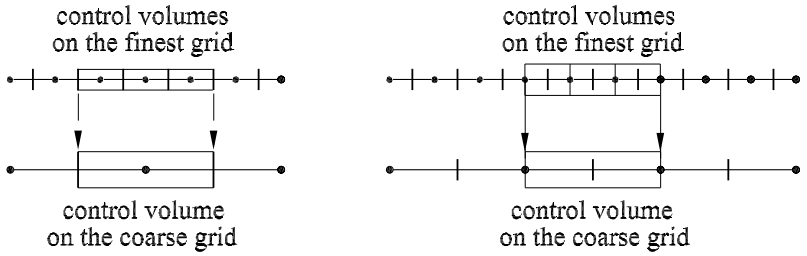


Figure 3: Control volumes on the finest and coarse grids

The finest grid forms the zero level and three coarse grids form the first level. The coarse grid generation is further recurrently repeated: each computational grid G_i^l, $i = 1, \ldots, 3^l$ of a current level l is considered to be the finest grid for the coarse grids G_j^{l+1}, $j = 1, \ldots, 3^{l+1}$ of the next level $l+1$. Nine coarse grids derived from the three grids of the first level form the second level, etc. The coarse grid generation is finished when no further coarsening can be performed. Union of the finest grid and all coarse grids will be called *the multigrid structure* (Fig. 4).

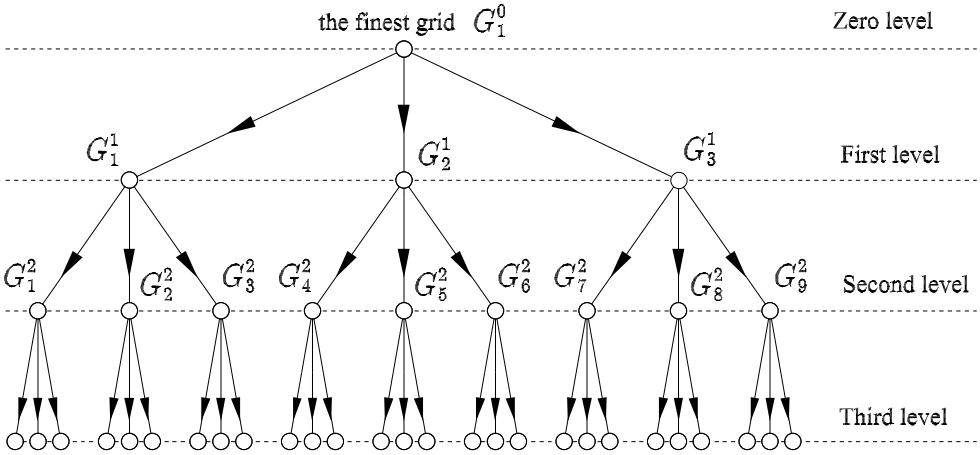


Figure 4: Multigrid structure

Since d-dimensional grid ($d = 2, 3$) can be represented as product of d one-dimensional grids, similar coarsening is performed independently in each spatial direction. Therefore level l consists of 3^{dl} grids in multidimensional case.

The number of levels can be computed in advance. Assuming that majority of the coarsest grids has three grid points, then the number of the finest grid points is $N_x^0 + 1$ or $\approx 3^{L^+ + 1}$, where L^+ is the number of the coarsest level. Therefore

$$N_x^0 + 1 \approx 3^{L^+ + 1} \;\Rightarrow\; L^+ = \left[\frac{\lg(N_x^0 + 1)}{\lg 3} - 1 \right],$$

where square brackets mean integer part.

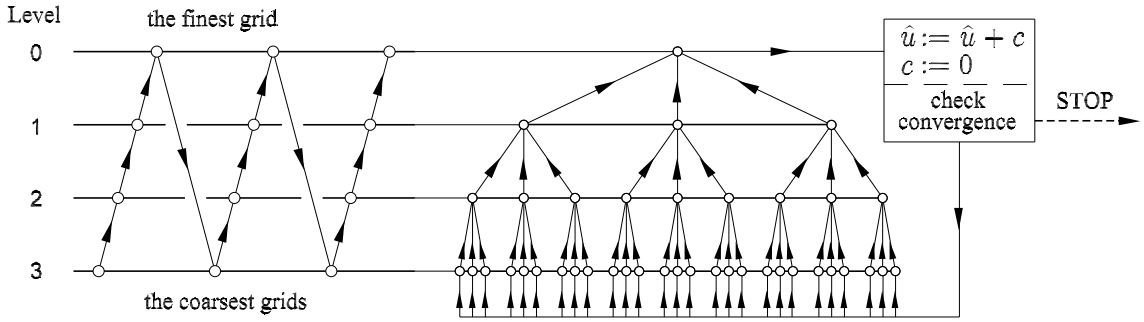


Figure 5: Multigrid cycle

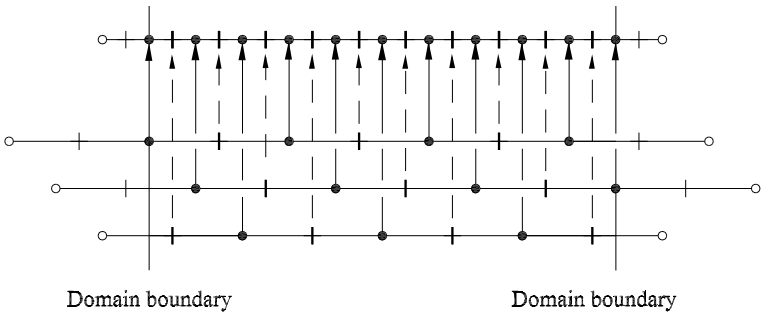


Figure 6: Transfer of correction from coarse grids to fine grid

Multigrid cycle of RMT is shown on Fig. 5. The multigrid schedule of RMT is sawtooth cycle, i.e. a special case of the V-cycle, in which smoothing before coarse grid correction (pre-smoothing) is deleted. Multigrid iterations start on the coarsest level. When the coarsest level solution has been obtained, the transfer to the next finer level is performed. It should be emphasized that the transfer does not add any interpolation errors to the correction $\boldsymbol{c}$ as shown on Fig. 6. Smooth parts of the error are deleted on all grids of the next finer levels in the same manner (computation of the coefficient matrix and right-hand side vector, after that the smoothing iterations). The coarse grid correction to be added to approximation of the solution $\hat{\boldsymbol{u}}$ on the finest grid is $\boldsymbol{c}$ ($\hat{\boldsymbol{u}} := \hat{\boldsymbol{u}} + \boldsymbol{c}$). The multigrid iterations repeatedly improve the approximation to the solution $\hat{\boldsymbol{u}}$ until the current approximation becomes accurate enough.

3 INTERGRID OPERATORS

Details of the control volume approximation on the multigrid structures are given in [5]. Here we focus attention on the intergrid operators of RMT in order to use their properties for the convergence analysis.

For reason of clearness consider one-dimensional Poisson equation

$$\frac{d^2u}{dx^2} = 10e^x, \qquad u(0) = u(1) = 0\,, \tag{1}$$

with exact solution

$$u(x) = 10\big(e^x + (1-e)x - 1\big)\,. \tag{2}$$

The solution $u(x)$ can be represented as

$$u(x) = c(x) + \hat{u}(x)\,, \tag{3}$$

where discrete analogues of the functions $c(x)$ and $\hat{u}(x)$ will be coarse grid correction and approximation to the solution, respectively. Substitution of (3) into (1) gives a Σ-modified form of the initial problem

$$\frac{d^2c}{dx^2} = 10e^x - \frac{d^2\hat{u}}{dx^2}, \qquad c(0) = c(1) = 0 \qquad \hat{u}(0) = \hat{u}(1) = 0\,. \tag{4}$$

The finest and coarse grids for the two-level algorithm are shown on Fig. 7. Let us define the index mapping to simplify operations with the multigrid structure. The notion "a grid of the level l" means one-to-one index mapping of the coarse grid points onto the indices of the finest grid points. In the rest of the paper, the mapping will be denoted by the braces $\{\}$. The index mapping of the grid points from the sets $G^{\rm v}$ and $G^{\rm f}$ is written as $x^{\rm v}_{\{i\}}$ and $x^{\rm f}_{\{i\}}$ respectively, where i and $\{i\}$ are the coarse and finest grid indices. The index mapping gives a close-to-the-finest-grid notation. Really the mapping shows that all derivatives and integrals are approximated on the finest grid, i.e. RMT uses only a single grid for solving BVP.

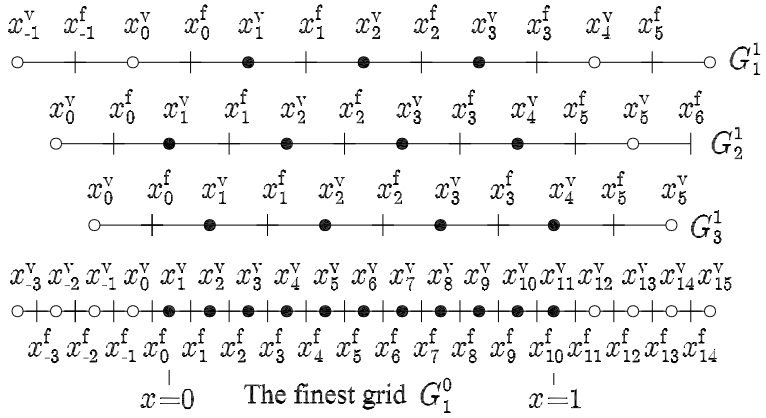


Figure 7: The finest (G_1^0) and three coarse grids (G_1^1, G_2^1 and G_3^1)

Let points $x^{\rm v}$ be vertices and points $x^{\rm f}$ be the control volume faces of the computational grids. Integration of (4) over the control volume $[x^{\rm f}_{\{i-1\}}, x^{\rm f}_{\{i\}}]$ gives

$$\frac{c_{\{i-1\}} - 2c_{\{i\}} + c_{\{i+1\}}}{h^2 3^{2l}} = \frac{1}{h3^l} \int\limits_{x^{\rm f}_{\{i-1\}}}^{x^{\rm f}_{\{i\}}} \left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right) dx\,, \tag{5}$$

where $h3^l$ is mesh size of the grids ($l = 0, 1$).

Fig. 7 shows that the index mapping for the coarse grids G_1^1 is given by

$$x^{\rm v}_{\{0\}} = x^{\rm v}_0, \qquad x^{\rm v}_{\{1\}} = x^{\rm v}_3, \qquad x^{\rm v}_{\{2\}} = x^{\rm v}_6, \qquad x^{\rm v}_{\{3\}} = x^{\rm v}_9, \qquad x^{\rm v}_{\{4\}} = x^{\rm v}_{12},$$
$$x^{\rm f}_{\{0\}} = x^{\rm f}_1, \qquad x^{\rm f}_{\{1\}} = x^{\rm f}_4, \qquad x^{\rm f}_{\{2\}} = x^{\rm f}_7, \qquad x^{\rm f}_{\{3\}} = x^{\rm f}_{10},$$

i.e. the first vertex of the coarse grid G_1^1 coincides with third vertex of the finest grid G_1^0 etc.

Right-hand side of (5) on the finest grid ($l = 0$) is approximated as

$$\frac{1}{h}\int\limits_{x_{i-1}^{\mathrm{f}}}^{x_i^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx \approx 10\,e^{x_i^{\mathrm{v}}} - \frac{\hat{u}_{i-1} - 2\hat{u}_i + \hat{u}_{i+1}}{h^2}\,.$$

Vector of resudials on the finest grid is defined by the components

$$r_1 = 0\,, \quad r_i = 10\,e^{x_i^{\mathrm{v}}} - \frac{\hat{u}_{i-1} - 2\hat{u}_i + \hat{u}_{i+1}}{h^2}\,, \quad r_{N_x^0+1} = 0\,. \tag{6}$$

Eq. (5) in the first vertex $x_{\{1\}}^{\mathrm{v}}$ of the coarse G_1^1 of the first level ($l = 1$) takes the form

$$\frac{c_{\{0\}} - 2c_{\{1\}} + c_{\{2\}}}{h^2 3^2} = \frac{1}{h3}\int\limits_{x_{\{0\}}^{\mathrm{f}}}^{x_{\{1\}}^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx\,.$$

Approximation of the boundary conditions on the coarse grids is discussed in [3, 5]. The equation

$$c_{\{0\}} = \frac{2}{\xi(\xi+1)}\;c\big|_{x=0} + 2\,\frac{\xi-1}{\xi}\,c_{\{1\}} - \frac{\xi-1}{\xi+1}\,c_{\{2\}} + O(h^3 3^{3l})\,,$$

is used for elimination of the coarse grid correction $c_{\{0\}}$ in the virtual vertex $x_{\{0\}}^{\mathrm{v}}$, where

$$\xi = \begin{cases} \dfrac{x_{\{1\}}^{\mathrm{v}}}{h\,3^l}\,, & x = 0 \\[2ex] \dfrac{1 - x_{\{N_k^l+1\}}^{\mathrm{v}}}{h\,3^l}\,, & x = 1 \end{cases}\,.$$

Here the value $c_{\{0\}}$ in the vertex $x_{\{0\}}^{\mathrm{v}}$ located outside the domain $\Omega = [0, 1]$ can be eliminated as follows

$$c_{\{0\}} = -c_{\{1\}} + \frac{1}{5}\,c_{\{2\}}\,.$$

Since the control volume $[x_{\{0\}}^{\mathrm{f}}, x_{\{1\}}^{\mathrm{f}}]$ is union of three control volumes of the finest grid (Fig. 3)

$$[x_{\{0\}}^{\mathrm{f}}, x_{\{1\}}^{\mathrm{f}}] = [x_1^{\mathrm{f}}, x_4^{\mathrm{f}}] = [x_1^{\mathrm{f}}, x_2^{\mathrm{f}}] \cup [x_2^{\mathrm{f}}, x_3^{\mathrm{f}}] \cup [x_3^{\mathrm{f}}, x_4^{\mathrm{f}}]\,,$$

the right-hand side can be rewritten as

$$\frac{1}{h3}\int\limits_{x_{\{0\}}^{\mathrm{f}}}^{x_{\{1\}}^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx = \frac{1}{3}\left(\frac{1}{h}\int\limits_{x_1^{\mathrm{f}}}^{x_2^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx + \right.$$

$$\left. + \frac{1}{h}\int\limits_{x_2^{\mathrm{f}}}^{x_3^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx + \frac{1}{h}\int\limits_{x_3^{\mathrm{f}}}^{x_4^{\mathrm{f}}}\left(10\,e^x - \frac{d^2\hat{u}}{dx^2}\right)dx\right) = \frac{1}{3}\big(r_2 + r_3 + r_4\big)\,.$$

Therefore (5) in the first vertex $x^{\mathrm{v}}_{\{1\}}$ of the coarse grid G^1_1 takes the final form

$$\frac{-3\,c_{\{1\}} + 1.2\,c_{\{2\}}}{9h^2} = \frac{1}{3}\big(r_2 + r_3 + r_4\big)\,.$$

Continuing the same considerations for the vertices $x^{\mathrm{v}}_{\{2\}}$ and $x^{\mathrm{v}}_{\{3\}}$ of the coarse grid G^1_1, we obtain the following system of linear equations

$$\frac{1}{9h^2}\begin{pmatrix} -3 & 1.2 & 0 \\ 1 & -2 & 1 \\ 0 & 1.2 & -3 \end{pmatrix}\begin{pmatrix} c_{\{1\}} \\ c_{\{2\}} \\ c_{\{3\}} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} r_2 + r_3 + r_4 \\ r_5 + r_6 + r_7 \\ r_8 + r_9 + r_{10} \end{pmatrix}. \tag{7}$$

Analogously we have for the coarse grid G^1_2

$$\frac{1}{9h^2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1.5 & -6 \end{pmatrix}\begin{pmatrix} c_{\{1\}} \\ c_{\{2\}} \\ c_{\{3\}} \\ c_{\{4\}} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 \\ r_3 + r_4 + r_5 \\ r_6 + r_7 + r_8 \\ r_9 + r_{10} + r_{11} \end{pmatrix}, \tag{8}$$

and for the coarse grid G^1_3

$$\frac{1}{9h^2}\begin{pmatrix} -6 & 1.5 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} c_{\{1\}} \\ c_{\{2\}} \\ c_{\{3\}} \\ c_{\{4\}} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} r_1 + r_2 + r_3 \\ r_4 + r_5 + r_6 \\ r_7 + r_8 + r_9 \\ 0 \end{pmatrix}. \tag{9}$$

Solutions of the systems (7), (8) and (9) with a zero starting guess ($\hat{u} = 0$) are shown on Fig. 8. In RMT more computational work must be spent on coarse grids in order to allow for the best approximation to the solution on the fine grids. Multiple coarse grid correction strategy used in RMT makes it possible to eliminate the problem-dependent interpolation from the multigrid algorithm. An efficient procedure for the fast integral evaluation on the multigrid structure is given in [5, 6]. Virtual vertices on all grids are needed only for this procedure.

Systems (7), (8) and (9) can be incorporated into common system for convergence analysis of RMT. Coefficient matrix of the common system has a block structure, where the number of blocks is equal to the number of coarse grids. For this example, we have

$$A_1 = \frac{1}{9h^2}\begin{pmatrix} -3 & 1.2 & 0 & & & & & & & & & \\ 1 & -2 & 1 & & & & & & & & & \\ 0 & 1.2 & -3 & & & & & & & & & \\ & & & 1 & 0 & 0 & 0 & & & & & \\ & & & 1 & -2 & 1 & 0 & & & & & \\ & & & 0 & 1 & -2 & 1 & & & & & \\ & & & 0 & 0 & 1.5 & -6 & & & & & \\ & & & & & & & -6 & 1.5 & 0 & 0 \\ & & & & & & & 1 & -2 & 1 & 0 \\ & & & & & & & 0 & 1 & -2 & 1 \\ & & & & & & & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

Total number of the vertices on all coarse grids of level l ($l = 1, 2, \ldots, L^+$) coincides with the number of the vertices on the finest grid ($l = 0$).

Right-hand side vector of the common system is

$$\mathcal{R}_{0\to 1}\boldsymbol{r}_0 = \frac{1}{3}\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \\ r_8 \\ r_9 \\ r_{10} \\ r_{11} \end{pmatrix}, \tag{11}$$

where components of the residual vector $\boldsymbol{r}$ are computed on the finest grid as (6), $\mathcal{R}_{0\to 1}$ is an restriction operator transferring the residual from the finest grid (zero level) to the coarse grids of the first level. It is clear that the restriction operator is the problem-independent component of RMT.

Vector of unknowns of the common system is union of vectors of systems (7), (8) and (9). Taking into account the index mapping, we have

$$\boldsymbol{c}_1 = \begin{pmatrix} c_{\{1\}}(G_1^1) \\ c_{\{2\}}(G_1^1) \\ c_{\{3\}}(G_1^1) \\ c_{\{1\}}(G_2^1) \\ c_{\{2\}}(G_2^1) \\ c_{\{3\}}(G_2^1) \\ c_{\{4\}}(G_2^1) \\ c_{\{1\}}(G_3^1) \\ c_{\{2\}}(G_3^1) \\ c_{\{3\}}(G_3^1) \\ c_{\{4\}}(G_3^1) \end{pmatrix} = \begin{pmatrix} c_3 \\ c_6 \\ c_9 \\ c_1 \\ c_4 \\ c_7 \\ c_{10} \\ c_2 \\ c_5 \\ c_8 \\ c_{11} \end{pmatrix} \tag{12}$$

Here $c_{\{1\}}(G_1^1)$ is value of the correction in the first vertex of the coarse grid G_1^1 etc. as shown on Fig. 9.

Prolongation in the example can be written as

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \\ c_9 \\ c_{10} \\ c_{11} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} c_3 \\ c_6 \\ c_9 \\ c_1 \\ c_4 \\ c_7 \\ c_{10} \\ c_2 \\ c_5 \\ c_8 \\ c_{11} \end{pmatrix} \tag{13}$$

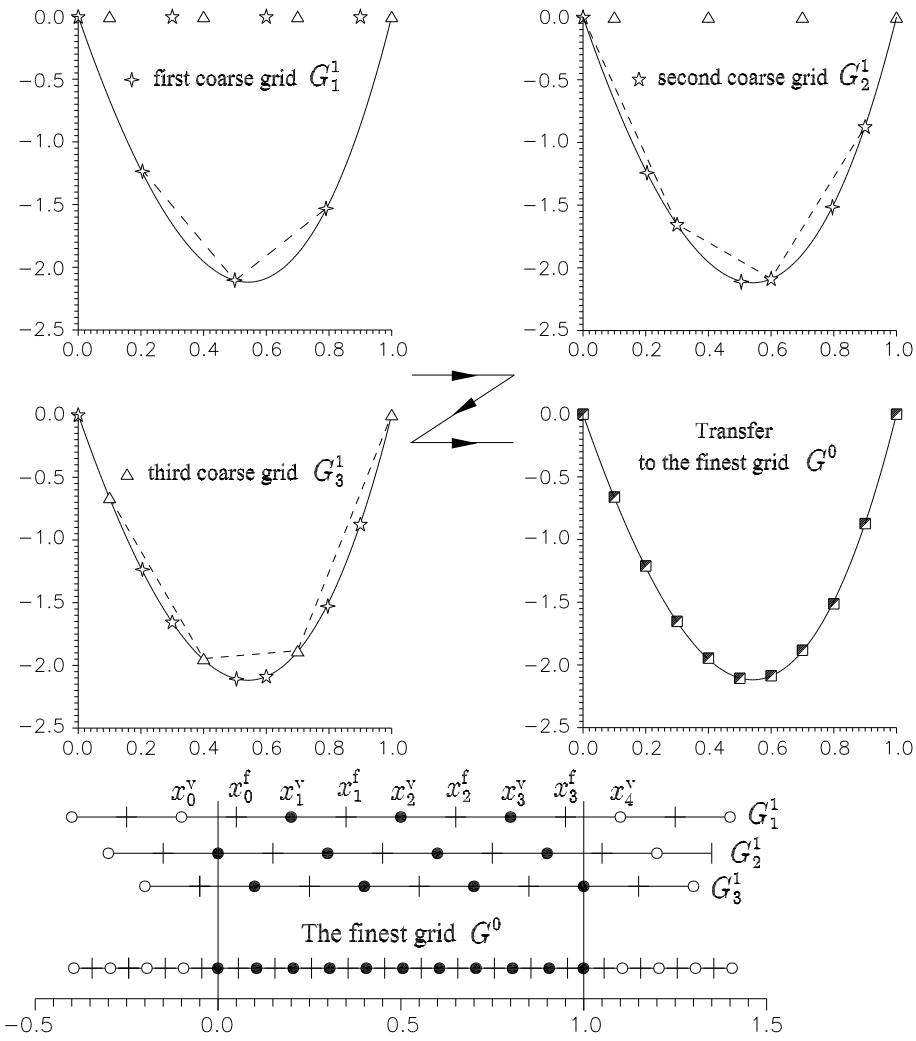


Figure 8: Solutions of the systems (7), (8) and (9) (dotted lines) vs. exact solution (2) (solid line)

or in matrix form

$$\boldsymbol{c}_0 = \mathcal{P}_{1\to 0}\boldsymbol{c}_1, \tag{14}$$

where $\mathcal{P}_{1\to 0}$ is the problem-independent prolongation operator (permutable matrix). Note that $\mathcal{P}_{1\to 0}^{-1} = \mathcal{P}_{1\to 0}^{\mathrm{T}}$.

We summarize some abovementioned results. Approximation of the Σ-modified BVP (4) on three coarse grids of the first level results in the common system of linear equation

$$A_1\boldsymbol{c}_1 = \mathcal{R}_{0\to 1}\boldsymbol{r}_0, \tag{15}$$

where the block-structured coefficient matrix A_1, vector of unknowns $\boldsymbol{c}_1$ and right-hand side vector $\mathcal{R}_{0\to 1}\boldsymbol{r}_0$ are given by (10), (12) and (11), respectively. Assume that the system (15) is solved by direct method, i.e.

$$\boldsymbol{c}_1 = A_1^{-1}\mathcal{R}_{0\to 1}\boldsymbol{r}_0,$$

where the vector $\boldsymbol{c}_1$ is defined by (12). After that the vector $\boldsymbol{c}_1$ is prolongated from the grids of the first level up to the finest grid as (13) (or in the matrix form (14)).

All matrices and all vectors in (15) have no "coarse grid variables" because RMT is the single-grid algorithm. Size of all matrices and all vectors is l-independent. It should be em-

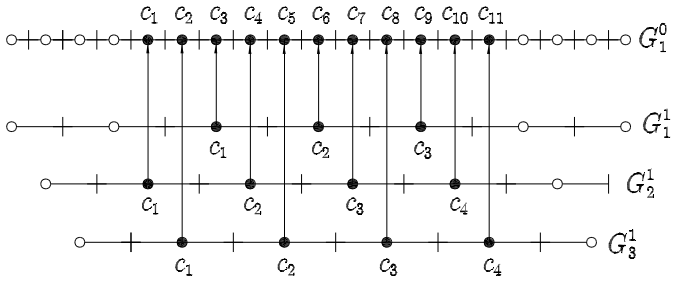


Figure 9: Prolongation of the correction in the one-dimensional example

phasized that multigrid notation is used only for clearness of the description and to follow traditional terminology.

Let us summarize main properties of the problem-independent transfer operators of RMT:

$$\mathcal{R}_{0\to l} = \begin{cases} I, & l = 0 \\ \prod\limits_{k=1}^{l} \mathcal{R}_{l-k\to l-k+1}, & l = 1, 2, \ldots, L^+ \end{cases},$$

$$\mathcal{P}_{l\to 0} = \begin{cases} I, & l = 0 \\ \prod\limits_{k=1}^{l} \mathcal{P}_{k\to k-1}, & l = 1, 2, \ldots, L^+ \end{cases},$$

where I is unit operator.

4 CONVERGENCE ANALYSIS AND COMPLEXITY

Approximation of some linear Σ-modified BVP on the multigrid structure results to series of the system of linear equations of the same size

$$A_l \boldsymbol{c}_l = \mathcal{R}_{0\to l}\big(\boldsymbol{b}_0 - A_0 \hat{\boldsymbol{\varphi}}_0^{(q)}\big), \qquad l = 0, 1, \ldots, L^+,$$

with problem-independent transfer operators $\mathcal{R}_{0\to l}$ and $\mathcal{P}_{l\to 0}$. Here $\hat{\boldsymbol{\varphi}}_0^{(q)}$ is an approximation to the solution $\hat{\boldsymbol{\varphi}}_0 = A_0^{-1}\boldsymbol{b}_0$ after q multigrid iterations and subscript indicates the level index.

Let the smoothers be abbreviated as

$$W_l\big(\boldsymbol{c}_l^{(\nu_l+1)} - \boldsymbol{c}_l^{(\nu_l)}\big) = \mathcal{R}_{0\to l}\big(\boldsymbol{b}_0 - A_0 \hat{\boldsymbol{\varphi}}_0^{(q)}\big) - A_l \boldsymbol{c}_l^{(\nu_l)}$$

or

$$\boldsymbol{c}_l^{(\nu_l+1)} = S_l \boldsymbol{c}_l^{(\nu_l)} + W_l^{-1}\mathcal{R}_{0\to l}\big(\boldsymbol{b}_0 - A_0 \hat{\boldsymbol{\varphi}}_0^{(q)}\big),$$

where $S_l = I - W_l^{-1}A_l$ is the smoothing iteration matrix. We assume that the smoother is convergent iterative method ($\|S_l\| < 1$) and the same number of the smoothing iterations (ν_l) is performed on all grids G_k^l, $k = 1, 2, \ldots, 3^{dl}$ of the level l ($l = 0, 1, 2, \ldots, L^+$, $d = 2, 3$).

Multigrid iterations of RMT is given by

$$\hat{\boldsymbol{\varphi}}_0^{(q+1)} = Q_0 A_0 \hat{\boldsymbol{\varphi}}_0^{(q)} + \big(A_0^{-1} - Q_0\big)\boldsymbol{b}_0,$$

where the matrix Q_0 is defined as

$$Q_l = \begin{cases} S_l^{\nu_l}\Big(d_l\mathcal{R}_{0\to l} + \mathcal{P}_{l+1\to l}Q_{l+1}\Big), & l = 0,1,2,\ldots,L^+-2 \\ S_l^{\nu_l} d_l\mathcal{R}_{0\to l}, & l = L^+-1 \end{cases},$$

$$d_l = A_l^{-1} - \mathcal{P}_{l+1\to l}A_{l+1}^{-1}\mathcal{R}_{l\to l+1}.$$

It is clear that the multigrid iteration matrix is $M = Q_0A_0$, for the details see [5].

Following W. Hackbusch [1], convergence proof of the multigrid methods is based on the definitions:

1. Smoothing property: the matrix S has the smoothing property if there exist a monotonically decreasing function $\eta(\nu_l) : \mathbb{R}_+ \to \mathbb{R}_+$ independent of the mesh size h such that $\eta(\nu_l) \to 0$ at $\nu_l \to \infty$ and

$$\|A_lS_l^{\nu_l}\| \leqslant \eta(\nu_l)\|A_l\|. \tag{16}$$

2. Approximation property: the approximation property holds if there exists a constant $C_A > 0$ independent of the mesh size h such that

$$\|A_l^{-1} - \mathcal{P}_{l+1\to l}A_{l+1}^{-1}\mathcal{R}_{l\to l+1}\| \leqslant C_A\|A_l\|^{-1}. \tag{17}$$

Approximation and smoothing properties should be proved for each problem.

Now we use the multigrid iteration matrix $M = Q_0A_0$ to derive the convergence theorem:

Convergence theorem. Assume the smoothing (16) and approximation (17) properties hold and $\|\mathcal{R}_{0\to l}\| \leqslant C_{\mathcal{R}}$. Then we have h-independent convergence of RMT and

$$\|M\|_2 \leqslant C_A\eta(\nu_0) + C_AC_{\mathcal{R}}\sum_{l=1}^{L^+-1} C^l\eta(\nu_l).$$

The theorem shows that the number of the multigrid iterations is independent of the mesh size h, but computational cost of the iterations is not optimal, i.e. the number of arithmetic operations per the multigrid iteration step is proportional to $N(L^+ + 1)$, where N is the number of grid points and $L^+ + 1$ is the number of grid levels. Since $L^+ + 1 \sim \lg N$, the number of arithmetic operations needed for one multigrid cycle of RMT is $O(N \lg N)$.

We represent results of a standard multigrid solver (V-cycle) and RMT applied to the model problem of the 2D Poisson equation:

$$-\Delta u = f \quad \text{in} \quad \Omega = (0,1)^2,$$

$$u = 0 \quad \text{on} \quad \partial\Omega,$$

with exact solution

$$u(x,y) = Q(x)Q(y), \quad \text{where} \quad Q(\beta) = 10\big(e^\beta + (1-e)\beta - 1\big), \quad \beta = (x,y).$$

In the experiment we use a starting guess $\hat{\boldsymbol{u}}^{(0)} = 0$ and the computational grids 129×129 ($h = 1/128$) and 1025×1025 ($h = 1/1024$). The rate of convergence is measured by looking the residual norm

$$\|\boldsymbol{b}_0 - A_0\hat{\boldsymbol{u}}^{(q)}\|_\infty = \|\boldsymbol{r}^{(q)}\|_\infty, \tag{18}$$

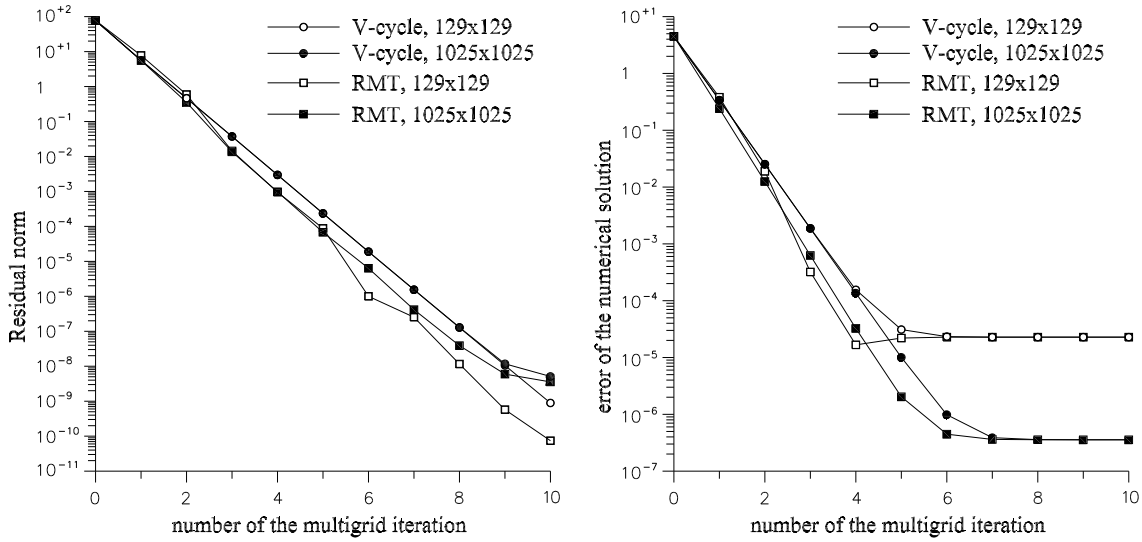


Figure 10: Comparison of the multigrid convergence: V-cycle vs. RMT

and error of the numerical solution defined as maximum of the absolute difference between the exact and numerical solutions, i.e.

$$\max_{ij} |Q(x_i)Q(y_j) - \hat{u}_{ij}^{(q)}|.$$

Fig. 10 represents reduction of the residual norm (18) and the error of the numerical solution as a function of the multigrid iteration q. Point Gauss–Seidel method is used as a smoother. V-cycle is implemented with three pre- and post-smoothing iterations and RMT is implemented with six (post)smoothing iterations. It can be seen that there is no noticeable difference between the convergence behaviors of the multigrid algorithms, but implementation of RMT iterations requires more computational work. Assuming that the computational grid consists of $(2^k + 1)^d$ vertices ($d = 2, 3$) and the computational costs of the transfer operators are negligible compared to the cost of the smoothing iterations, augmentation of the computational efforts in RMT as compared with V-cycle is estimated as

$$\approx \left[\frac{\lg(2^k + 1)}{\lg 3}\right] \frac{2^d - 1}{2^d + 1} \approx \Big[0.63k\Big] \frac{2^d - 1}{2^d + 1}$$

arithmetic operations, where square brackets mean integer part.

5 ROBUST SMOOTHER

Robustness and efficiency of a multigrid method are strongly influenced by the smoother used. Users of multigrid methods employ many different smoothing methods because efficiency of the smoothers are problem-dependent.

Let 2D discrete boundary value problem arising from a five-point discretization be abbreviated as

$$A_{ij}^w u_{i-1j} + A_{ij}^e u_{i+1j} + A_{ij}^s u_{ij-1} + A_{ij}^n u_{ij+1} + A_{ij}^p u_{ij} = b_{ij}.$$

Firstly, we consider a point Gauss-Seidel iteration, which consists of updating the unknown corresponding to the nodes located at the center of the five-point stencil:

$$u_{ij} = \frac{1}{A_{ij}^p}\Big(b_{ij} - A_{ij}^w u_{i-1j} - A_{ij}^e u_{i+1j} - A_{ij}^s u_{ij-1} - A_{ij}^n u_{ij+1}\Big).$$

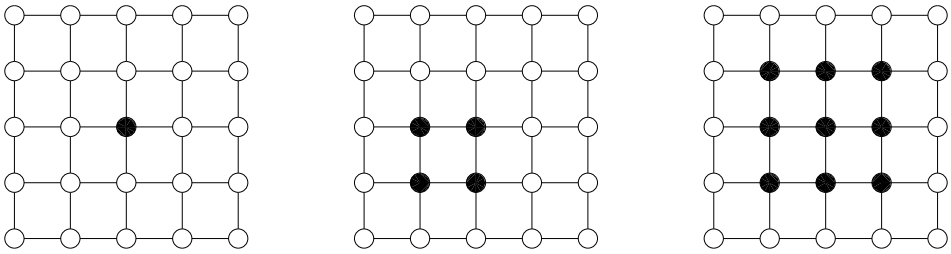


Figure 11: Block ordering of unknowns: blocks 1×1, 2×2 and 3×3

There are many variants of point-type smoothers, that can differ in the choice of the unknown ordering. Simple point-wise smoothers (with any ordering) are not appropriate for the saddle point type problems (i.e. if $A^p_{ij} = 0$).

In general the block ordering of the unknowns can be used for construction of the Gauss–Seidel smoothers. It consists of decomposing the grid into small overlapping subdomains and looping over all of them solving the system arising from the equations corresponding to the points in the subdomain as shown on Fig 11. Point-wise smoothers can be considered as a block iterative methods with blocks 1×1.

Application of blocks 2×2 give the following system

$$\begin{pmatrix} A^p_{ij} & A^e_{ij} & A^n_{ij} & 0 \\ A^w_{i+1j} & A^p_{i+1j} & 0 & A^n_{i+1j} \\ A^s_{ij+1} & 0 & A^p_{ij+1} & A^e_{ij+1} \\ 0 & A^s_{i+1j+1} & A^w_{i+1j+1} & A^p_{i+1j+1} \end{pmatrix} \begin{pmatrix} u_{ij} \\ u_{i+1j} \\ u_{ij+1} \\ u_{i+1j+1} \end{pmatrix} = \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \end{pmatrix},$$

where

$$\begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \tilde{b}_3 \\ \tilde{b}_4 \end{pmatrix} = \begin{pmatrix} b_{ij} - A^w_{ij}\, u_{i-1j} - A^s_{ij}\, u_{ij-1} \\ b_{i+1j} - A^e_{i+1j}\, u_{i+2j} - A^s_{i+1j}\, u_{i+1j-1} \\ b_{ij+1} - A^w_{ij+1}\, u_{i-1j+1} - A^n_{ij+1}\, u_{ij+2} \\ b_{i+1j+1} - A^e_{i+1j+1}\, u_{i+2j+1} - A^n_{i+1j+1}\, u_{i+1j+2} \end{pmatrix}.$$

Solution of the system by a direct method (for example, Gaussian elimination) gives the updated values of u_{ij}, u_{i+1j}, u_{ij+1} and u_{i+1j+1}.

Also there are many variants of box-type smoothers, they can differ in the choice of the subdomains which are solved simultaneously, and in the way in which the local systems are solved. In addition, the different subdomains can be visited in different ways yielding to a wide variety of the box-smoothers.

Note that computational cost of the system solution is g^3 arithmetic operations, where g is the number unknowns in the block. In limit $g \to N$ the ordering results in a direct solver, where N is the number unknowns.

Compared to point-wise smoothers, the crucial advantage of the block smoothers is the ability to deal with zero elements appearing on the diagonal of the coefficient matrix. Very important partial case is a Vanka smoother developed to solve the saddle-point systems arising in the field of computational fluid dynamics, particularly for incompressible flow problems [9]. Other reasons are that it is not too difficult to implement and at the same time it is efficient and robust for a wide class of problem configurations.

6 NUMERICAL EXPERIMENTS

We consider the following partial differential equation (PDE)

$$\frac{\partial}{\partial x}\left(\lambda^x \frac{\partial w}{\partial x}\right) + \frac{\partial}{\partial y}\left(\lambda^y \frac{\partial w}{\partial y}\right) + \frac{\partial}{\partial z}\left(\lambda^z \frac{\partial w}{\partial z}\right) = -f(x,y,z)$$

in the unit cube $\Omega = (0,1)^3$ for the illustration of the RMT convergence. Zero starting guess and uniform grid $101 \times 101 \times 101$ ($h = 1/100$) is used in the numerical experiments. Convergence rate of RMT is illustrated by the average reduction factor of the residual

$$\rho^{(q)} = \left(\frac{\|A_0 \hat{\boldsymbol{u}}^{(q)} - \boldsymbol{b}_0\|_\infty}{\|A_0 \hat{\boldsymbol{u}}^{(0)} - \boldsymbol{b}_0\|_\infty}\right)^{1/q}$$

with q is the number of the multigrid iteration, $\hat{\boldsymbol{u}}^{(0)} = 0$ is a starting guess. The average reduction factor is computed after five multigrid iterations ($q = 5$).

Example 1: Poisson equation ($\lambda^x = \lambda^y = \lambda^z = 1$). Exact solution of the Poisson equation is taken as

$$w_e = e^{x+y+z}, \tag{19}$$

and the error of the numerical solution is defined as

$$e^{(q)} = \max_{ijk} |w_e(x_i, y_j, z_k) - \hat{u}_{ijk}|.$$

Results of the numerical test are given in Table 1. Point Gauss-Seidel (blocks $1 \times 1 \times 1$) and block Gauss-Seidel (blocks $3 \times 3 \times 3$) methods with natural unknown ordering are used as the smoothers in the test. Execution time of the multigrid algorithm with four point Gauss-Seidel smoothing iterations ($\nu = 4$) is taken as a work unit (WU=1) in the numerical experiments. It is clear that RMT works efficiently for the Poisson equation.

Block	ν	$e^{(q)}$	$\rho^{(q)}$	WU
$1 \times 1 \times 1$	4	$7.30 \cdot 10^{-6}$	0.084	1.00
$1 \times 1 \times 1$	5	$7.30 \cdot 10^{-6}$	0.033	1.16
$1 \times 1 \times 1$	6	$7.30 \cdot 10^{-6}$	0.023	1.35
$3 \times 3 \times 3$	3	$7.30 \cdot 10^{-6}$	0.055	1.03
$3 \times 3 \times 3$	4	$7.30 \cdot 10^{-6}$	0.033	1.32
$3 \times 3 \times 3$	5	$7.30 \cdot 10^{-6}$	0.021	1.62

Table 1: Convergence of RMT in **Example 1** ($q = 5$)

Example 2: anisotropic equation ($\lambda^x = \lambda^y = 1, \lambda^z = 0.1$) with the same exact solution (19). Table 2 represents results of the test. Point Gauss-Seidel smoother does not work efficiently because the discrete variables are linked weakly in Z direction ($\lambda^z \ll \lambda^x = \lambda^y$). Application of Gauss-Seidel smoother with the block ordering of the unknowns (blocks $5 \times 5 \times 1$) shows the best results because the smoother takes into account the problem anisotropy.

Example 3: equation with discontinuous coefficients. Assume that unit cube Ω consists of two materials with heat conductivity coefficients λ_i (internal subdomain $\tilde{\Omega}$) and λ_e (external

Block	ν	$e^{(q)}$	$\rho^{(q)}$	WU
$1 \times 1 \times 1$	4	$3.29 \cdot 10^{-3}$	0.752	1.00
$1 \times 1 \times 1$	5	$1.44 \cdot 10^{-3}$	0.637	1.16
$1 \times 1 \times 1$	6	$6.25 \cdot 10^{-4}$	0.538	1.35
$3 \times 3 \times 3$	3	$1.77 \cdot 10^{-4}$	0.522	1.03
$3 \times 3 \times 3$	4	$2.73 \cdot 10^{-5}$	0.333	1.32
$3 \times 3 \times 3$	5	$9.90 \cdot 10^{-6}$	0.216	1.62
$5 \times 5 \times 1$	3	$1.08 \cdot 10^{-5}$	0.268	0.81
$5 \times 5 \times 1$	4	$7.87 \cdot 10^{-6}$	0.142	1.01
$5 \times 5 \times 1$	5	$7.81 \cdot 10^{-6}$	0.081	1.21

Table 2: Convergence of RMT in **Example 2** ($q = 5$)

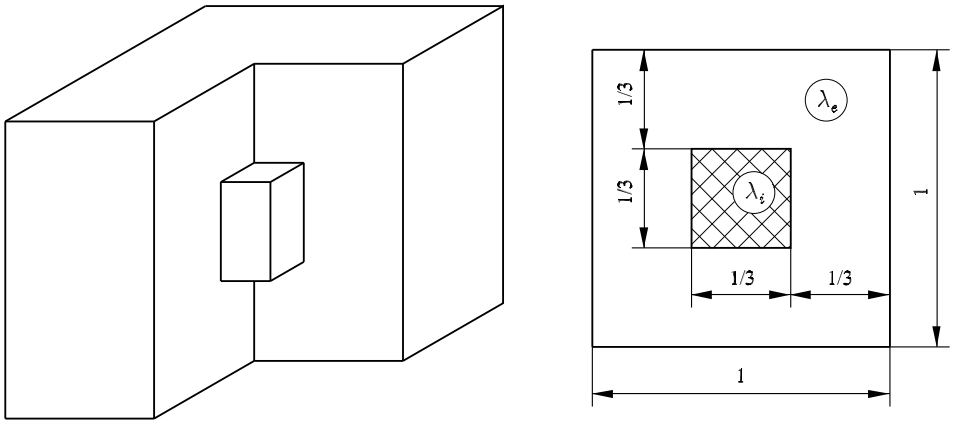


Figure 12: Geometry of the problem with the discontinuous coefficients

subdomain) as shown on Fig. 12, i.e.

$$\lambda^x(x,y,z) = \lambda^y(x,y,z) = \lambda^z(x,y,z) = \begin{cases} \lambda_i, & (x,y,z) \in \tilde{\Omega} \\ \lambda_e, & (x,y,z) \notin \Omega \end{cases}.$$

In case of $\lambda_i = \lambda_e = 1$, the exact solution is given by (19).

Results of the test presented in Table 3 show that the block ordering of the unknowns gives more efficient smoother as compared with the point ordering.

Example 4: nonlinear equation. Let the coefficients λ^x, λ^y and λ^z are defined as

$$\lambda^x = w^{-a}, \quad \lambda^y = w^{-b}, \quad \lambda^z = w^{-c},$$

where a, b and c are some positive constants. Linearization of the discrete nonlinear equation is based on the computation of the coefficients using correction and approximation to the solution obtained on previous level, i.e.

$$\lambda_l^x = (c_{l+1} + \hat{u}_{l+1})^{-a}, \quad \lambda_l^y = (c_{l+1} + \hat{u}_{l+1})^{-b}, \quad \lambda_l^z = (c_{l+1} + \hat{u}_{l+1})^{-c}.$$

Solution of the nonlinear problem with exact solution (19) starts from $\hat{u}^{(0)} = 0$ and $\lambda_x^{(0)} = 1$, $\lambda_y^{(0)} = 1$ and $\lambda_z^{(0)} = 1$. Typical convergence rate of RMT is shown in Table 4. It is easy to see that RMT with the block smoother is an efficient solver for the nonlinear problem.

Block	λ_e	λ_i	ν	$\rho^{(q)}$	WU
$1 \times 1 \times 1$	1	10^{+0}	6	0.023	1.92
$1 \times 1 \times 1$	1	10^{-1}	6	0.228	1.92
$1 \times 1 \times 1$	1	10^{-2}	6	0.198	1.92
$1 \times 1 \times 1$	1	10^{-3}	6	0.227	1.92
$1 \times 1 \times 1$	1	10^{-4}	6	0.273	1.92
$1 \times 1 \times 1$	1	10^{-5}	6	0.410	1.92
$3 \times 3 \times 3$	1	10^{+0}	5	0.021	2.05
$3 \times 3 \times 3$	1	10^{-1}	5	0.193	2.05
$3 \times 3 \times 3$	1	10^{-2}	5	0.170	2.05
$3 \times 3 \times 3$	1	10^{-3}	5	0.150	2.05
$3 \times 3 \times 3$	1	10^{-4}	5	0.149	2.05
$3 \times 3 \times 3$	1	10^{-5}	5	0.149	2.05

Table 3: Convergence of RMT in **Example 3** ($q = 5$)

Block	a	b	c	ν	$e^{(q)}$	$\rho^{(q)}$	WU
$1 \times 1 \times 1$	0.25	0.50	0.75	4	$9.00 \cdot 10^{-6}$	0.317	1.44
$1 \times 1 \times 1$	0.25	0.50	0.75	5	$2.83 \cdot 10^{-6}$	0.169	1.77
$1 \times 1 \times 1$	0.25	0.50	0.75	6	$2.40 \cdot 10^{-6}$	0.115	2.01
$3 \times 3 \times 3$	0.25	0.50	0.75	3	$6.01 \cdot 10^{-6}$	0.256	1.55
$3 \times 3 \times 3$	0.25	0.50	0.75	4	$2.27 \cdot 10^{-6}$	0.110	1.80
$3 \times 3 \times 3$	0.25	0.50	0.75	5	$2.14 \cdot 10^{-6}$	0.075	2.06

Table 4: Convergence of RMT in **Example 4** ($q = 5$)

Example 5: incompressible Navier-Stokes equations. In the application of RMT to the cavity problem shown on Fig. 13, a number of the flow Reynolds numbers (100 and 500) and the computational grids (101×101 and 1001×1001) have been considered. Inlet and outlet velocity components are distributed under the parabolic law:

$$u(0,y) = \begin{cases} 100y(0.2-y), & \text{at } y \leqslant 0.2, \\ 0, & \text{at } y > 0.2, \end{cases} \qquad u(1,y) = u(x,0) = u(x,1) = 0\,,$$

$$v(x,1) = \begin{cases} 100(x-0.8)(1-x), & \text{at } x \geqslant 0.8, \\ 0, & \text{at } x < 0.8, \end{cases} \qquad v(0,y) = v(1,y) = v(x,0) = 0\,,$$

on the other boundaries no-slip conditions are given. Detailed description of the Σ-modification of the Navier-Stokes equations and control volume approximation of the modified Navier-Stokes equations on the multigrid structures are given in [5, 6]. Σ-modification makes it possible to solve discrete Navier-Stokes equations without FAS. Full Vanka method is used as a smoother in the test. Fig. 14 and 15 represent the distribution of the stream function isolines and isobars. Results of test shown in Table 5 demonstrate robustness and efficiency of RMT with full Vanka smoother for the given benchmark problem. Obtained convergence rate depends weakly on the Reynolds number and the mesh size of the computational grids.

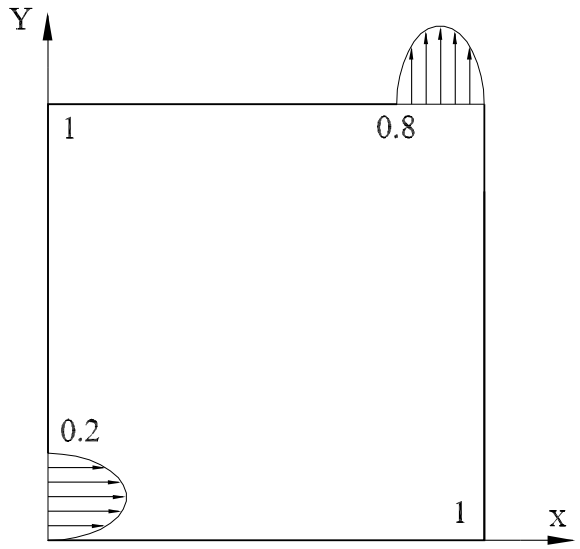


Figure 13: Geometry of the cavity

Grid	Re	ν	$\rho^{(q)}$
101×101	100	4	0.132
1001×1001	100	4	0.151
101×101	500	4	0.167
1001×1001	500	4	0.194

Table 5: Convergence of RMT in **Example 5**

The numerical tests illustrate that the problem-dependent components of RMT are:
1) the number of smoothing iterations;
2) unknown ordering (for anisotropic problems);
3) underrelaxation parameter (for nonlinear problems);
4) stopping criterion of the multigrid iterations.

7 PARALLEL MULTIGRID

Parallelization of classic multigrid methods (CMM) follows in the standard fashion of dividing the domain into subdomains (one for each processor). Each processor is then responsible for updating of the unknowns associated within its subdomain only. However parallel efficiency of CMM can degrade due to coarse grid smoothing. There can be reached situations where the number of processors exceeds the number of coarse grid points. As a result, some processors are idle during these computations. One approaches to overcome poor computation-to-communication ratio on the coarse grids is multiple coarse grid corrections. Nevertheless the choice of appropriate multigrid components and their efficient parallel implementation is highly problem-dependent in CMM [8].

There are two other common measures of parallelism that we now introduce:

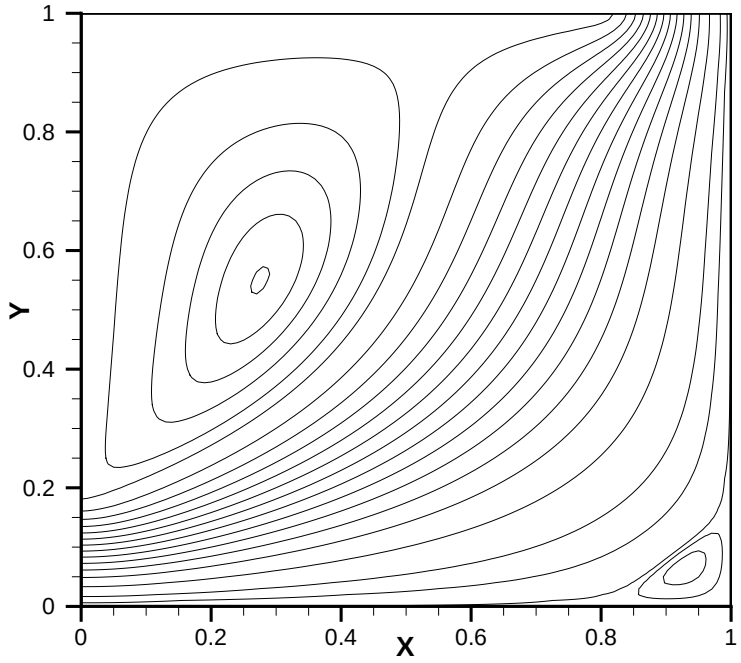


Figure 14: Isolines of the stream functions ($Re = 100$)

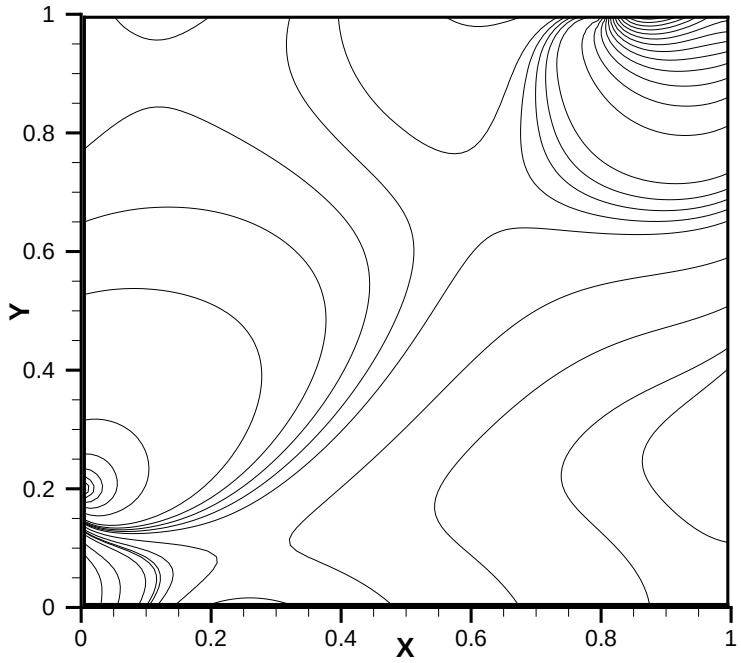


Figure 15: Isobars ($Re = 100$)

Definition 1. The speedup ($\bar{\mathcal{S}}$) and the efficiency ($\bar{\mathcal{E}}$) of a parallel algorithm is

$$\bar{\mathcal{S}} = p\bar{\mathcal{E}} = \frac{T(1)}{T(p)}, \tag{20}$$

where $T(1)$ is the execution time for a single processor and $T(p)$ is the execution time using p processors.

Definition 2. The speedup ($\tilde{\mathcal{S}}$) and the efficiency ($\tilde{\mathcal{E}}$) of a parallel algorithm over the best sequential algorithm is

$$\tilde{\mathcal{S}} = p\tilde{\mathcal{E}} = \frac{\tilde{T}(1)}{T(p)}, \tag{21}$$

where $\tilde{T}(1)$ is the execution time on a single processor of fastest sequential algorithm and $T(p)$ is the execution time of the parallel algorithm on p processors.

General description of the common measures of parallelism is given in [7] and the references therein. In addition to these common measures, we introduce the measures of parallel properties of smoother:

Definition 3. The speedup ($\mathcal{S}$) and the efficiency ($\mathcal{E}$) of a parallel smoother is

$$\mathcal{S} = p\mathcal{E} = \frac{\hat{T}(1)}{\hat{T}(p)}, \tag{22}$$

where $\hat{T}(1)$ is the execution time for a single processor of the smoother and $\hat{T}(p)$ is the execution time using p processors.

Goal of our analysis is to estimate of the speedup $\bar{\mathcal{S}}, \tilde{\mathcal{S}}$ and the efficiency $\bar{\mathcal{E}}, \tilde{\mathcal{E}}$ of a parallel algorithm using the measures of parallelism of the smoothing procedure (22). Let $l = 0$ denote the finest grid, L^+ be number of grid level consisting of 3^{dL^+} coarsest grids ($d = 2, 3$) and l^* ($0 \le l^* \le L^+$) be a depth of the parallelization. On finer grids l ($0 \le l \le l^*$) parallel smoothing causes little overhead and grid partitioning is used for parallelization. To overcome overhead on coarse grids of level l ($l^* < l \le L^+$), the given discrete problem will be decomposed into 3^{dl^*} subproblems without an overlap using properties of the multigrid structure. It is clear that the number of processor (p) should be $p = 3^{dl^*}$ for perfect load balance [4].

Let us consider parallelization of the first depth, i.e. $l^* = 1$ or $p = 3^d$, $d = 2, 3$. Distribution of the coarse grids among the processors are shown on Fig. 16. We assume that the same number of the smoothing iterations are performed on each grid of the multigrid structure. Since the total number of grid points on all grids of the same level is N, the execution time of the smoothing iterations is constant $T_l(1) = const$ ($l = 0, 1, \ldots, L^+$). For $l^* = 1$ we obtain

$$\bar{\mathcal{S}} = 3^d\bar{\mathcal{E}} < 3^d \frac{1 + q^*L^+}{\dfrac{1}{\mathcal{E}_0} + q^*L^+}. \tag{23}$$

Execution time of the sequential V-cycle can be estimated as

$$T(1) \approx \frac{T_0(1)}{1 - 2^{-d}}. \tag{24}$$

Substitution of (24) into (21) gives the estimations of speedup and efficiency of parallel RMT over V-cycle

$$\tilde{\mathcal{S}} = 3^d\tilde{\mathcal{E}} < \frac{3^d}{1 - 2^{-d}} \frac{1}{\dfrac{1}{\mathcal{E}_0} + q^*L^+}. \tag{25}$$

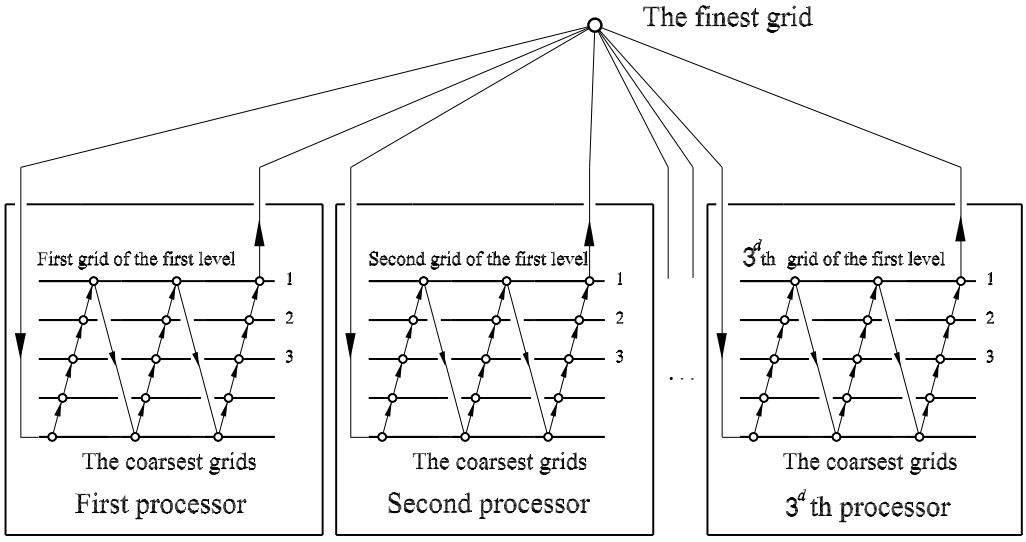


Figure 16: Distribution of subproblems among the processors in case $l^* = 1$ and $q^* = 3$

The estimations (23) and (25) predict that the measures of RMT parallelism depend strongly on efficiency of the parallel smoothing procedure on the finest grid $\mathcal{E}_0$. Since

$$\bar{\mathcal{E}} < \bar{\mathcal{E}}_{\max} = \mathcal{E}_0 \frac{1 + q^* L^+}{1 + q^* L^+ \mathcal{E}_0} \;\Rightarrow\; \frac{\bar{\mathcal{E}}_{\max}}{\mathcal{E}_0} = \frac{1 + q^* L^+}{1 + q^* L^+ \mathcal{E}_0} > 1,$$

the efficiency of parallel RMT will be higher than efficiency of the parallel smoothing procedure on the finest grid.

In the limit case $N \to \infty$ we have:

$$N \to \infty \;\Rightarrow\; L^+ \to \infty \;\Rightarrow\; \begin{cases} \bar{\mathcal{S}} \to 3^d \quad \text{and} \quad \bar{\mathcal{E}} \to 1, \\ \tilde{\mathcal{S}} = O(\lg^{-1} N) \quad \text{and} \quad \bar{\mathcal{E}} = O(\lg^{-1} N), \end{cases}$$

i.e. RMT has full parallelism at $N \to \infty$, but remarkable loss in complexity with classic multigrid is expected.

Eqs. (23) and (25) make it possible to estimate the speedup and efficiency of parallel RMT using only efficiency of parallel smoothing procedure as a function of grid points.

3D Dirichlet boundary value problem for Poisson equation in unit cube is used for numerical experiments. The problem has exact solution e^{x+y+z}, which defines the boundary conditions and source term. Computational grid $245 \times 245 \times 245$ ($N = 14706125 \;\Rightarrow\; L^+ = 4$) is used for the control volume discretization of the problem. OpenMP technology ($l^* = 1$, $d = 3$, $p = 27$) is applied for parallel implementation of RMT. OpenMP is a concurrency platform for multithreaded, shared-memory parallel processing architecture for C, C++ and Fortran. Shared-memory multiprocessor consisting of four processors AMD Opteron 6176 and RAM 128 Gb is used for the computations. Four smoothing iteration steps are performed on each grid with two extra multigrid iterations $q^* = 2$ on coarse levels l ($l^* \le l \le L^+$). The numerical experiments show that obtained efficiencies of parallel smoothing are $\bar{\mathcal{E}}_0 = 0.89$ and $\bar{\mathcal{E}}_* = 0.95$ for the finest and coarse grids, respectively. As a result, obtained efficiency of parallel RMT is 0.92, but estimation (20) predicts $\bar{\mathcal{E}} = \bar{\mathcal{S}} 3^{-d} = 0.98$.

On the other hand, comparison with sequential V-cycle shows that $\tilde{\mathcal{E}} = 0.112$, whereas estimation (21) predicts 0.125. It means that parallel RMT with $p = 27$ only approximately is three time faster than the sequential V-cycle.

Similar estimations for the parallelization of the second depth ($l^* = 2$, $p = 3^{2d}$, $d = 2, 3$) become

$$\bar{\mathcal{S}} = 3^{2d}\bar{\mathcal{E}} < 3^{2d}\frac{2 + q^*(L^+ - 1)}{\dfrac{1}{\mathcal{E}_0} + \dfrac{1}{\mathcal{E}_1} + q^*(L^+ - 1)},$$

$$\tilde{\mathcal{S}} = 3^{2d}\tilde{\mathcal{E}} < \frac{3^{2d}}{1 - 2^{-d}}\frac{1}{\dfrac{1}{\mathcal{E}_0} + \dfrac{1}{\mathcal{E}_1} + q^*(L^+ - 1)}.$$

We summarize the advantages of the multiple coarse grid correction strategy in parallel multigrid:
1) In classical multigrid on the coarse grids algorithms the ratio between communication and computation becomes worse than on fine grids. As a result, a large communication overhead on very coarse grids is expected. This strategy allows us to obtain almost full parallelism on coarse grids by decomposition of the given problem into a fixed number of subproblems without an overlap.
2) The time spent on very coarse grids in W-cycles may become unacceptable. In contrast, the extra multigrid iterations on coarse levels, which are employed in this strategy, lead to further reduction of communication overhead.
3) In classical multigrid algorithms on very coarse grids we may have idle processes. The fixed number of processes in the approach gives almost perfect load balance on the coarse grids.
4) Parallelization of computations on the coarse grids is independent of the smoother.

8 CONCLUSIONS

Application of the essential multigrid principle in single-grid algorithm makes it possible to develop a fast and highly parallel solver of boundary value problems on structured grids. Really RMT can be considered as a problem-independent acceleration technique of Gauss-Seidel iterations. It is expected that RMT will be used in black box software packages where the highest efficiency for a single problem is not so important.

ACKNOWLEDGEMENT

This paper is part of the project “Supercomputer simulation of physical and chemical processes in the high-speed direct-flow propulsion jet engine of the hypersonic aircraft on solid fuels” supported by Russian Science Foundation (project no. 15-11-30012)

REFERENCES

[1] W. Hackbusch, *Multi-grid Methods and Applications*, Berlin, Heidelberg: Springer, 1985.

[2] S.I. Martynenko, Robust multigrid technique for solving partial differential equations on structured grids, *Numerical Methods and Programming*, **1**, 83–100, 2000.

[3] S.I. Martynenko, Robust multigrid technique for black box software, *Computational Methods in Applied Mathematics*, **6**(4), 413–435, 2006.

[4] S.I. Martynenko, Potentialities of the robust multigrid technique, *Computational Methods in Applied Mathematics*, **10**(1), 87–94, 2010.

[5] S.I. Martynenko, *Multigrid Technique: Theory and Applications*, Moscow: FIZMATLIT, 2015. (in Russian)

[6] S.I. Martynenko, Robust multigrid method for solving the Navier-Stokes equations on structured grids, *Modern Applied Science*, **6**(6), 64–72, 2012.

[7] J. Ortega, *Introduction to Parallel and Vector Solution of Linear Systems*, New York: Plenum Press, 1988.

[8] U. Trottenberg, C.W. Oosterlee, A. Schüller, *Multigrid*, London: Academic Press, 2000.

[9] S.P. Vanka, Block-implicit multigrid solution of Navier-Stokes equations in primitive variables, *Journal of Computational Physics*, **65**, 138–158, 1986.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ALGORITHM FOR FAST SIMULATIONS OF SPACE–TIME FINITE ELEMENT METHOD

Marcin Skotniczny[1], **Anna Paszynska**[2], **and Maciej Paszynski**[1]

[1]AGH University of Science and Technology,
al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: {mskotn, maciej.paszynski}@agh.edu.pl

[2]Jagiellonian University
ul. Lojasiewicza 11, Krakw, Poland
anna.paszynska@uj.edu.pl

Keywords: Finite element method, Time–space formulations, h–adaptivity, Multi–frontal solver, Orderings

Abstract. *In this paper we consider element partition trees for multi–frontal solver algorithms utilized in space–time finite element method. The element partition trees are utilized for generating the ordering for the multi–frontal solver algorithm. In particular, we consider three or four dimensional finite element method grids where two or three dimensions represent space and one additional dimension represents time. Additionally, we consider computational grids resulting from h–adaptive algorithms, namely grids refined towards point, edge, face or hyperface singularity. We perform numerical experiments and compare our method with alternative state of the art ordering algorithms available through MUMPS interface.*

1 INTRODUCTION

We propose a heuristic algorithm generating element partition trees for 3D or 4D grids refined towards singularities. This work is an extension of the algorithm proposed for two dimensional problems [1]. When we solve 3D time dependent problem over a large grid, we need to utilize a sequence of grids, that can be refined only in spatial domain, and we cannot take advantage of possible time adaptivity. In this paper we consider an alternative approach that allows for time adaptivity, i.e. allows to use different time steps over different parts of the 3D mesh, at the same time allowing the time steps to change over time. As example, we investigate four dimensional grids refined towards point, edge, face and hyper–face singularities. We generate element partition trees obtained by bisections weighted by element size and estimate the resulting number of floating point operations of the multi–frontal solver algorithm. Our analysis shows that for both three– and four–dimensional point and edge singularity we obtain the linear computational cost $O(N)$, for three– and four–dimensional face singularity we obtain computation cost $O(N^{1.5})$ and for four–dimensional hyper–face singularity we obtain cost of $O(N^2)$ where N is the number of nodes. Values for edge, face and hyper–face singularities correspond to one, two and three dimensional uniform grids respectively. Our method is dedicated to stable time-space formulations using classical finite element method [2] or the DPG method [3].

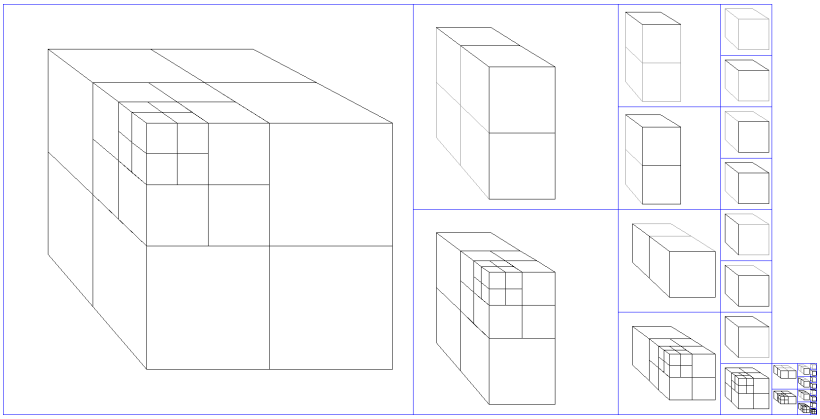


Figure 1: 3D mesh with point singularity and its element partition tree.

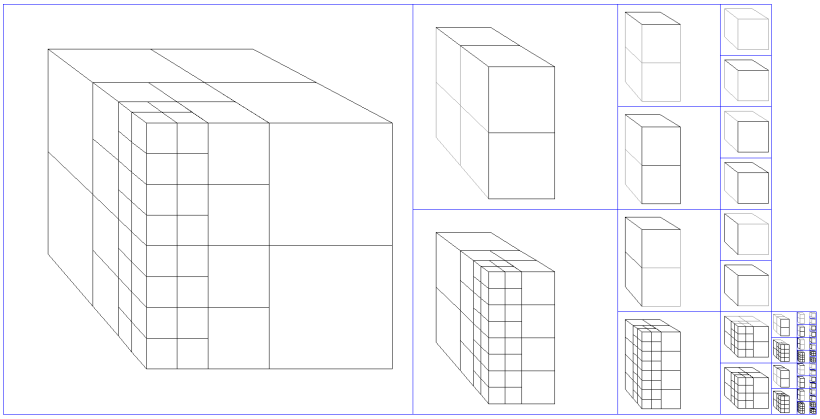


Figure 2: 3D mesh with edge singularity and its element partition tree.

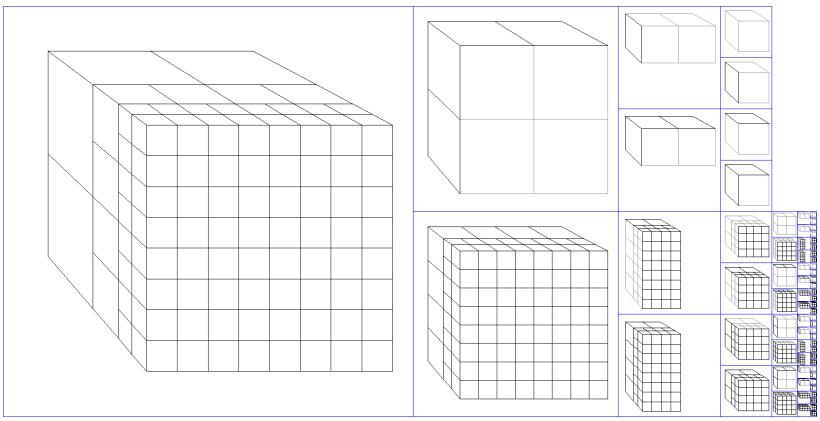


Figure 3: 3D mesh with face singularity and its element partition tree.

2 BISECTIONS WEIGHTED BY ELEMENT SIZE ALGORITHM

In this section we introduce top down algorithm for the construction of element partition trees, called bisections weighted by element size algorithm. The algorithm is based on the multilevel recursive bisection. We make the following asumptions:

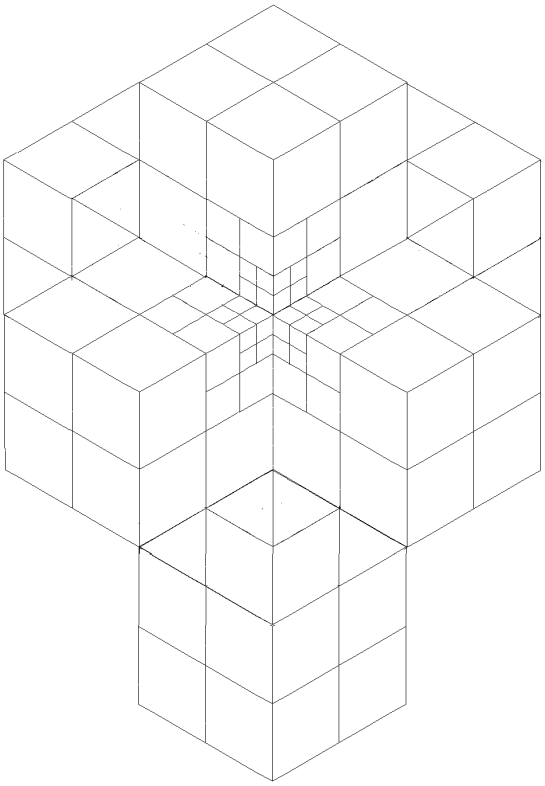
- The input is the graph in which vertices represent mesh elements and edges represent spatial adjacency relation between them.
- We assign weights to vertices equal to the volume of element that each of them represents.
- The volume is proxy for refinement level and is defined as $(\frac{1}{2})^{3R}$ in 3D and $(\frac{1}{2})^{4R}$ in 4D, where R is the refinement level of a mesh element. This way, the total volume over all elements should equal 1.

The input graph is recursively partitioned into 2 equally–weighted parts by using graph partitioning algorithm. The goal of the graph partitioning problem is to compute a balanced partitioning, such that:

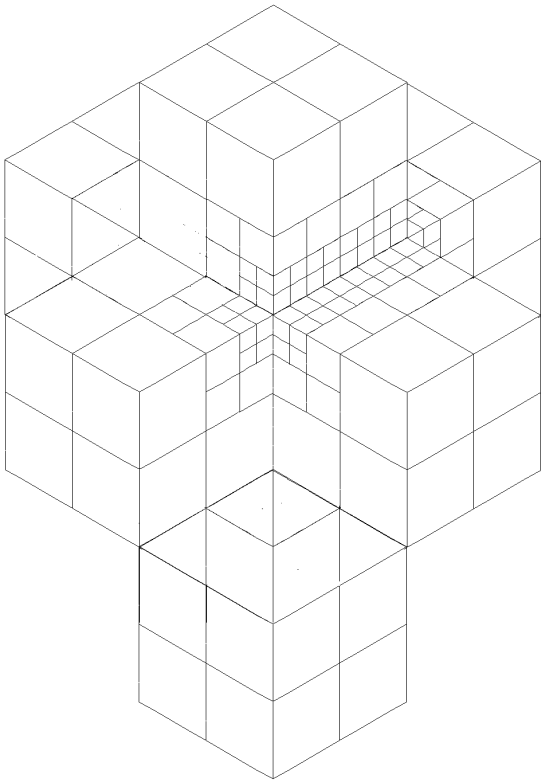
- the number of edges (or the sum of their weights) over the interface is minimal, and
- the number of vertices in each part of the graph is the same (or the sum of weights of vertices in each part of the graph).

To solve the problem, we use METIS_WPartGraphRecursive function to find a balanced partition of a graph where weights on vertices are equal to the volumes of represented mesh elements and weights on the edges equal to 1.

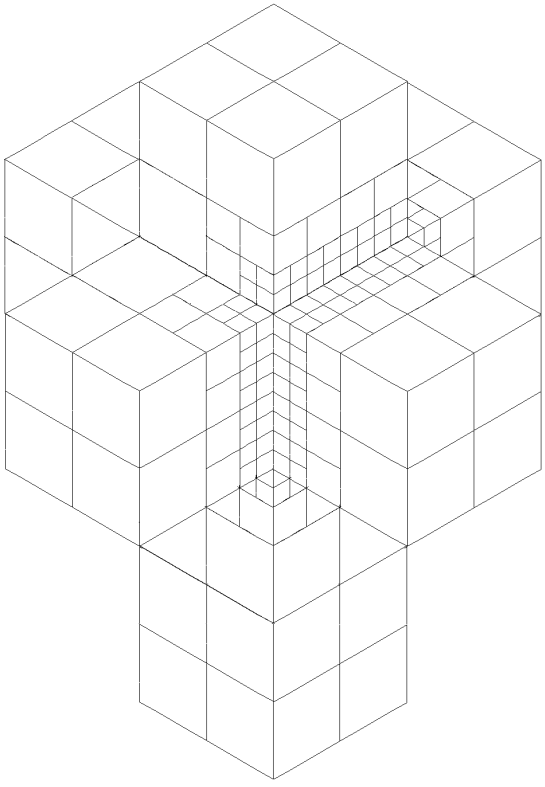
We illustrate the obtained partitions for the recursive partition over 3D meshes with point, edge and face singularities in Figures 1–3. We also show the boundary of corresponding 4D meshes with singularities in Figure 4. The recursive partitions generate the ordering in the following way: we start traversing the partition tree from leaves up to the root in parallel. At leaves we eliminate interiors of elements and on the internal tree nodes we eliminate mesh nodes on faces shared between two partitioned blocks of the mesh. We continue this process recursively up to the root of the partition tree.



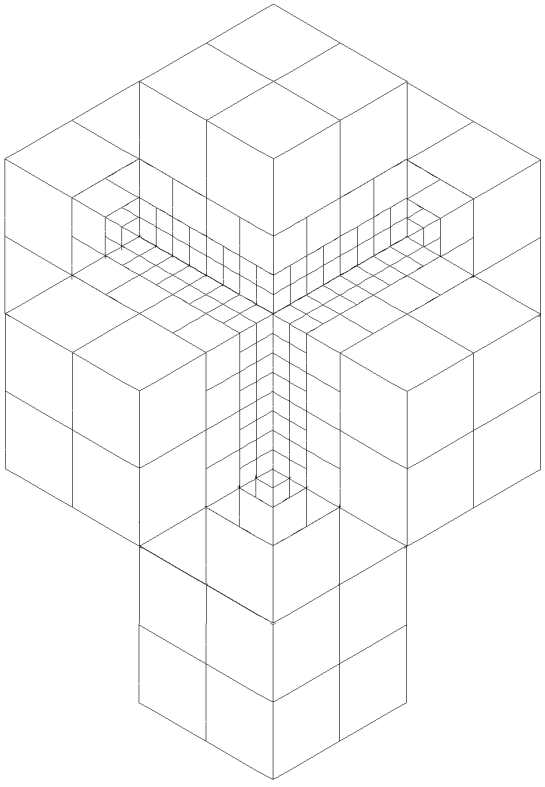
(a) Point singularity.



(b) Edge singularity.

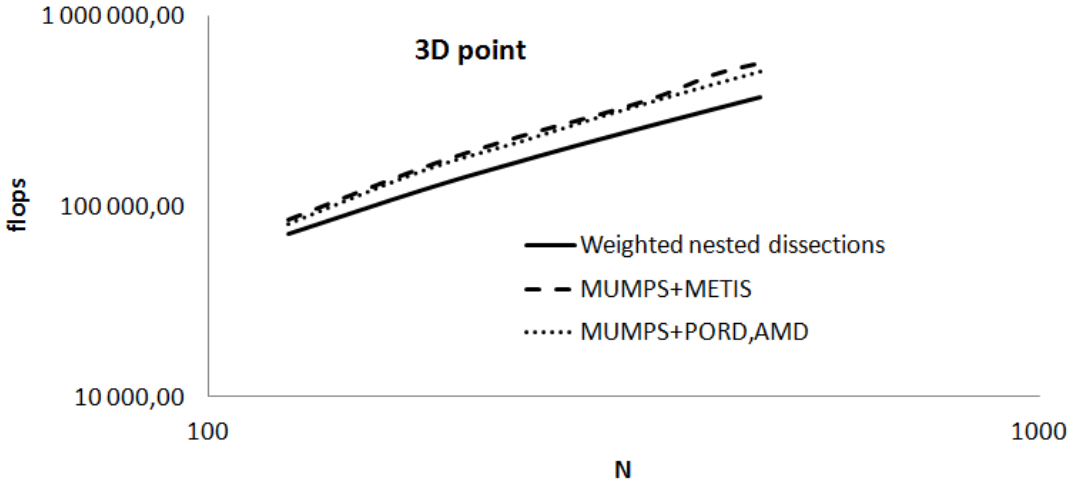


(c) Face singularity.

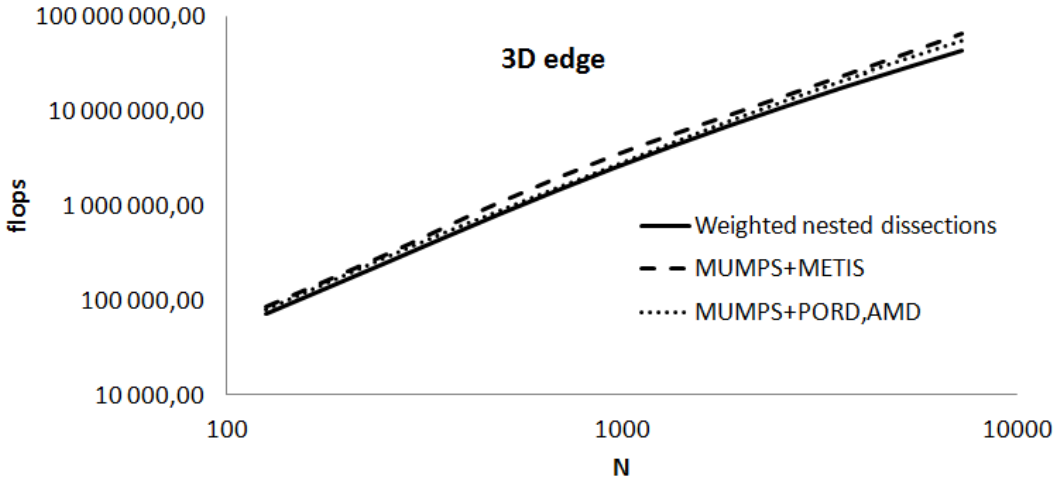


(d) Hyperface singularity.

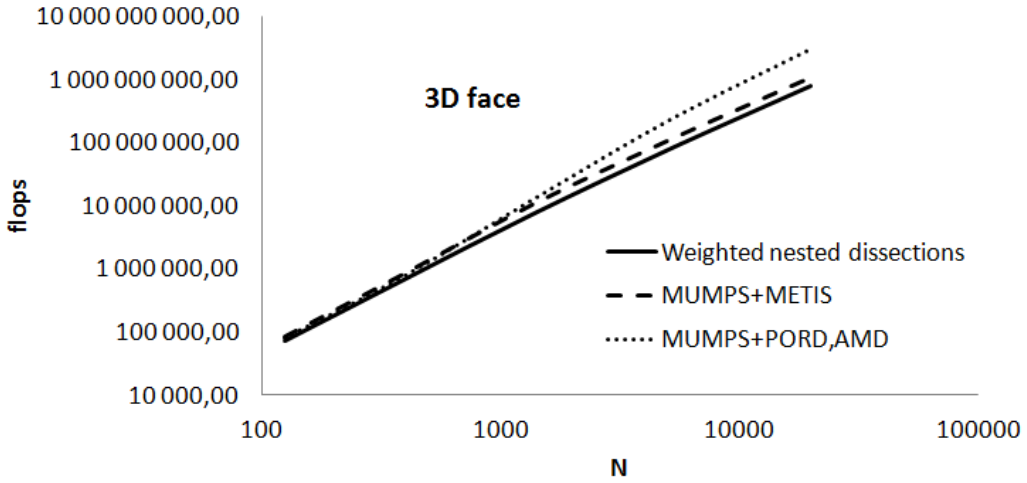
Figure 4: 4D meshes with singularities.



(a) 3D point singularity.



(b) 3D edge singularity.



(c) 3D face singularity.

Figure 5: Scalability of the weighted nested dissections algorithm in 3D.

3 NUMERICAL RESULTS

In this section we present numerical experiments concerning the execution time of our bisections weighted by element size algorithm for 3D and 4D grids with singularities. We show that our method outperforms MUMPS [4, 5] state of the art solver with (approximate) minimum degree or filling orderings [6, 7, 8], METIS [9] or PORD [10] orderings. The comparison is presented in Figures 5 and 6, for 3D and 4D respectively. The FLOPs for the 4D singularities, namely point singularity (kind=0), edge singularity (kind=1), face singularity (kind=2), hyperface singularity (kind=3) are also summarized in Table 1.

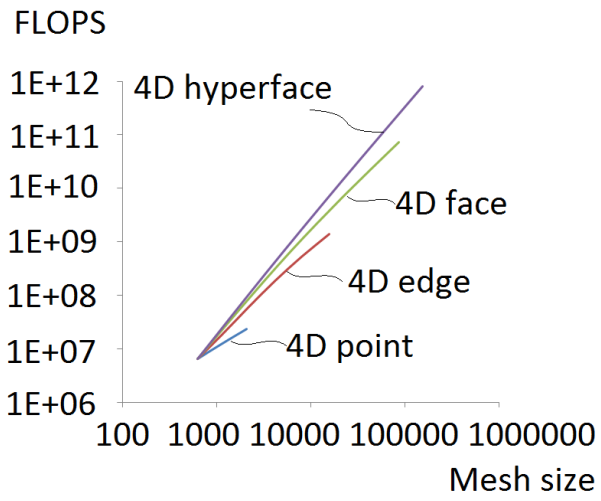


Figure 6: Scalability of the weighted nested dissections algorithm in 4D.

4 CONCLUSIONS

We have presented preliminary results for a promising method for the generation of the orderings for the multi–frontal solver algorithm — the numerical results show that the approach presented can outperform the current state of the art solutions. At the same time we have verified a novel approach to solve space–time formulations considering the time dimension as another space dimension. We intend to investigate this approach and the possibilities it presents for various finite element methods. In particular, it promises benefits over traditional algorithms in the areas of adaptivity and parallelization.

Table 1: Floating point operation costs for four dimensional singularities.

Dimensions	Singularity	Depth	Elements	Nodes	FLOPs
4	0	0	1	81	180441
4	0	1	16	625	6438573
4	0	2	31	865	9260925
4	0	3	46	1105	12083277
4	0	4	61	1345	14905629
4	0	5	76	1585	17727981
4	0	6	91	1825	20550333
4	0	7	106	2065	23372685
4	0	8	121	2305	26195037
4	0	9	136	2545	29017389
4	0	10	151	2785	31839741
4	0	11	166	3025	34662093
4	0	12	181	3265	37484445
4	0	13	196	3505	40306797
4	0	14	211	3745	43129149
4	0	15	226	3985	45951501
4	0	16	241	4225	48773853
4	0	17	256	4465	51596205
4	0	18	271	4705	54418557
4	0	19	286	4945	57240909
4	1	0	1	81	180441
4	1	1	16	625	6438573
4	1	2	46	1161	15728465
4	1	3	106	2177	45010855
4	1	4	226	4153	132140065
4	1	5	466	8049	372950179
4	1	6	946	15785	986944501
4	1	7	1906	31201	2449394343
4	1	8	3826	61977	5755536601
4	1	9	7666	123473	12948968907
4	1	10	15346	246409	28178439933
4	2	0	1	81	180441
4	2	1	16	625	6438573
4	2	2	76	1821	46109087
4	2	3	316	6121	517831573
4	2	4	1276	22389	5999916811
4	2	5	5116	85633	64220568465
4	3	0	1	81	180441
4	3	1	16	625	6438573
4	3	2	136	3291	222691306
4	3	3	1096	21485	12813363367

ACKNOWLEDGEMENTS

The work presented in this paper has been supported by National Science Centre, Poland grant no. DEC-2012/06/M/ST1/00363.

REFERENCES

[1] H. AbouEisha, V. M. Calo, K. Jopek, M. Moshkov, A. Paszynska, M. Paszynski, M.Skotniczny, Optimization of Element Partition Trees for Two-Dimensional h Refined Meshes, submitted to *Computers and Mathematics with Applications* (2016)

[2] T.J.R. Hughes, G.M. Hulbert, Space-Time Finite Element Method for Elastodynamics: Formulations and Error Estimates, *Computer Methods in Applied Mechanics and Engineering* 66 (1988) 339-363

[3] L. Demkowicz, J. Gopalakrishnan, Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations (eds. X. Feng, O. Karakashian, Y. Xing). In: vol. 157. IMA *Volumes in Mathematics and its Applications*, (2014). An Overview of the DPG Method, 149–180

[4] P. R. Amestoy, I. S. Duff, J. Koster and J.-Y. L'Excellent, 2001. A fully asynchronous multifrontal solver using distributed dynamic scheduling, *SIAM Journal of Matrix Analysis and Applications*, 23(1) :15–41.

[5] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent and S. Pralet, 2006. Hybrid scheduling for the parallel solution of linear systems. *Parallel Computing* Vol 32(2):136-156.

[6] P. Heggernes, S.C. Eisenstat, G. Kumfert, A. Pothen, 2001. The Computational Complexity of the Minimum Degree Algorithm, ICASE Report No. 2001-42.

[7] P. R. Amestoy, T. A. Davis, I. S. Du, 1996. An Approximate Minimum Degree Ordering Algorithm, *SIAM Journal of Matrix Analysis & Application*, 17(4):886-905.

[8] G.W. Flake, R.E. Tarjan, K. Tsioutsiouliklis, 2003. Graph clustering and minimum cut trees, *Internet Mathematics* 1:385-408.

[9] G. Karypis, V. Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, *SIAM Journal of Scientiffic Computing*, 20, 1 (1998) 359-392.

[10] J. Schulze, Towards a tighter coupling of bottom-up and top-down sparse matrix ordering methods, *BIT*, 41, 4 (2001) 800.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

HYBRID PARALLELISATION OF AN ALGORITHMICALLY DIFFERENTIATED ADJOINT SOLVER

Pavanakumar Mohanamuraly[1], **Jan C. Hückelheim**[1] **and Jens D. Mueller**[1]

[1]Queen Mary University of London
Mile End Road, E1 4NS, London, United Kingdom
{p.mohanamuraly | j.c.hueckelheim | j.mueller}@qmul.ac.uk

Keywords: Parallel computing, MPI, OpenMP, Algorithmic differentiation, Source transformation, Adjoint CFD

Abstract. *We present a novel approach to parallelise an unstructured node-based finite volume solver using a hybrid MPI and OpenMP paradigm. The basic ingredients of this approach are, (i) zero-halo partitioning of the unstructured mesh and (ii) shared node residual accumulation. These two ingredients preclude the need to explicitly exchange the state information across partitions, allowing the computations to run independently in each partition. As a consequence, we retain the original loop structure for the numerical flux kernels, which can be parallelised using OpenMP directives. Due to the hand-assembly of reverse-differentiated routines in our adjoint solver, the adjoint MPI recipes presented in earlier work can not be applied without modifications. We present a modified adjoint MPI treatment for two MPI operations in the context of our hand-assembled nonlinear iterative solver, namely: (i) shared node accumulation and (ii) in-place all-reduce summation. We demonstrate the parallelisation approach on our in-house solver* mgopt*, which uses the Tapenade AD tool to generate the adjoint code. We show preliminary scalability results for a simple* $2d$ *problem.*

1 PDE constraint optimisation

The present work aims at parallelising the primal and adjoint solvers used in the context of PDE based optimisation, namely aerodynamic shape optimisation problems. The problem can be formulated as follows,

$$\min_{\alpha} J(u, \alpha) \tag{1}$$

$$s.t. \quad R(u, \alpha) = 0 \tag{2}$$

Where, J is referred to as the objective or cost function, which is minimised subject to PDE constraint R. The variable u is called the state and α is the control or design variable. Gradient based optimisation methods require one to calculate the quantity $\frac{dJ}{d\alpha}$. When the number of design variables (dimension of α) is higher than number of output variables (dimension of J), it is computationally efficient to obtain adjoint sensitivity $\frac{dJ}{d\alpha}^T$ given by equation 3. The cost of calculating the adjoint sensitivity is proportional to the dimension of the output J and independent of the dimension of the input α. Algorithmic differentiation (AD) in the reverse or adjoint mode is used to obtain the transposed values. Tapenade[11] AD tool is used in this work to obtain the terms of the adjoint equation 4, which solves for the so-called adjoint variable v.

$$\frac{dJ}{d\alpha}^T = \frac{\partial J}{\partial \alpha}^T + \frac{\partial R}{d\alpha}^T v \tag{3}$$

$$\left(\frac{\partial R}{\partial U}\right)^T v = \left(\frac{\partial J}{\partial U}\right)^T \tag{4}$$

In the present context, the PDE constraint is the compressible steady Navier-Stokes equation, α is the shape design parameter and u is the fluid state. Node-based finite volume discretisation is used in the present work to solve both the primal and the adjoint system. Details of the numerical method and implementation are available in reference [4]. The distributed parallelisation of the node-based finite volume discretisation using Message Parsing Interface (MPI) is presented in the next section.

2 Zero-halo partitioning and MPI adjoints

A popular choice for distributed parallelisation in scientific codes is the use of halo or ghost layers. They are an extra layer(s) of cells placed adjacent to the mesh partition boundary to mimic cells from neighbouring mesh partition. The adjoint of the MPI calls for this approach can be obtained using the framework described in reference [1]. An overview of the halo-layer approach is depicted in figure 1. The main drawback of this approach is that the MPI operations are not self-adjoint, i.e., the primal and adjoint have different MPI call sequence. In a source transformed AD, this forces one to replace MPI calls within the code with its adjoint equivalent. Currently, there are no tools available to automate this task and one has to manually implement the forward and reverse MPI calls. The AMPI[1] library reduces this burden by providing MPI wrappers for both forward and reverse mode. The wrappers typically tape appropriate MPI function arguments in the case of operator overloading or push/pop to stack in the case of source transformation. Lastly, the halo approach gives rise to artificial increase in number of computations per partition by duplicating operations at the halo boundary. This in-turn reduces the strong scaling efficiency. In most practical scientific codes, graph based mesh partitioning tools are used, which try to minimise the graph edge cuts (communication

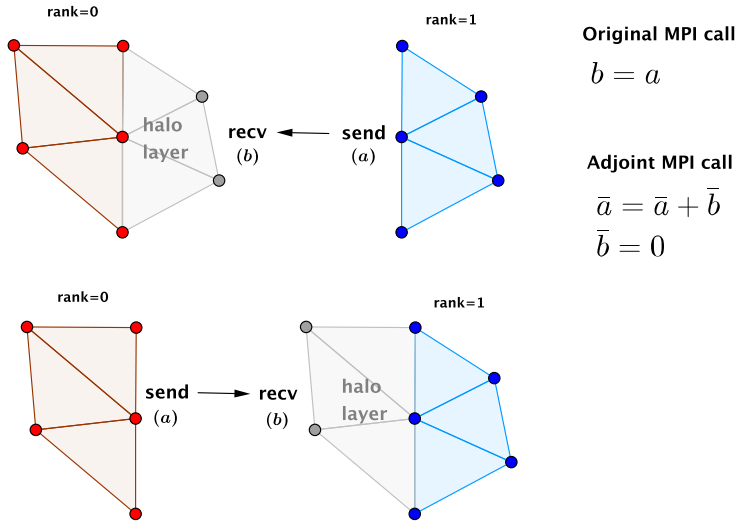


Figure 1: Parallel implementation based on halo layers and their MPI forward-adjoint equivalents

cost) maintaining good load balance. In reference [6], the authors show that for large number of partitions the load balancing severely degrades for graph based methods. The problem is exacerbated in the adjoint [7], which amplifies the load imbalance in the primal causing further reduction in scaling.

In a zero-halo layer approach, the partition local fluxes are computed and accumulated at the partition shared nodes/cells as shown in figure 2. At the partition boundaries, flux calculated at one partition face exactly cancels with its neighbouring partition face pair. As a result, one imposes a no-flux or empty boundary condition at the partition boundary faces. In addition, one should ensure that the flux residuals are accumulated at the shared nodes across partition boundaries to account for the flux residual from the adjacent partition. As a result, flux accumulation at shared-nodes preclude the need for an explicit exchange of the state information. Unlike halo-layer implementation, here it is easier to avoid the duplication of operation and improve scalability. The most attractive feature of this approach is the self-adjoint nature of the MPI call sequence: figure 2(a). This simplifies the implementation in a fixed-point iteration (FPI) type primal/adjoint solver as shown in listing (1) and (2).

Listing 1: Primal FPI

```
...
do iter = 1, n
  call residue(u, R)
  call accumulate(R)
  call update(u, R)
end do
call cost_fun(u, J)
```

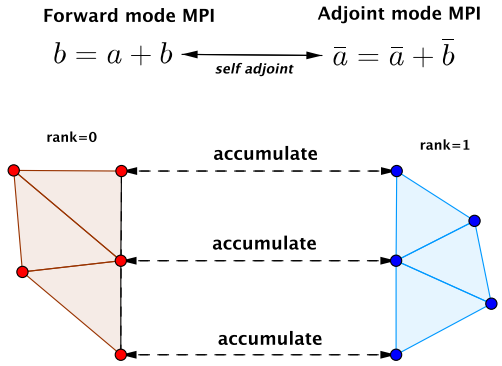
Listing 2: Hand assembled adjoint FPI [4]

```
J̄ = 1
call cost_fun_b(u, (∂J/∂u)^T J̄, J, J̄)
do iter = 1, n
  call residue_b(u, (∂R/∂u)^T v, R, v)

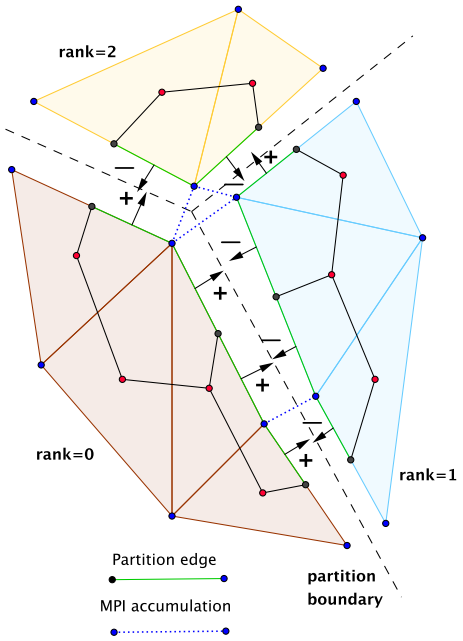
  call accumulate((∂R/∂u)^T v)

  R = (∂R/∂u)^T v - (∂J/∂u)^T
  call update(v, R)
end do
```

Although the adjoint of the MPI operation for shared-node accumulation is symmetric, other



(a) Zero-halo accumulation and the forward-adjoint mode MPI equivalents



(b) Exploded view of a three partition zero-halo flux exchange

Figure 2: Zero-halo implementation, flux accumulation, and self-adjoint MPI operation

MPI operations warrant careful examination. Consider evaluation of a scalar cost function (J), for example, the aerodynamic lift or drag. The pseudo code to calculate the cost function is shown in listing (3). Here the cost function subroutine receives as an input a vector quantity u and returns the scalar J.

Listing 3: Cost function (drag) evaluation (parallel)

```
! Input  : u (vector)
! Output : J (scalar)
subroutine cost_fun(u, J)
  call drag_force( u, J )
  call Allreduce( IN_PLACE, J, ..., SUM )
end subroutine cost_fun
```

The adjoint in the serial case (without Allreduce) is simply obtained by passing the $drag_force$ subroutine to a source transformation engine like Tapenade. On the contrary, it is not straightforward to construct the cost function adjoint in parallel (with IN_PLACE Allreduce). One would be tempted to use the MPI adjoint recipe from reference [2] for the Allreduce operation as shown in listing (4). For the hand assembled adjoint shown in listing (2), verbatim use of the recipe results in an incorrect source term.

Listing 4: Allreduce recipe from [2]

```
! Original call
Allreduce( x, y, ,,, SUM )
! Adjoint call
Allreduce( ȳ, t̄, ,,, SUM )
ȳ = 0
x̄ = x̄ + t̄
```

The hand assembled adjoint source term becomes unambiguous when viewed in the context of solving the (adjoint) linear system shown in equation 4. The subroutine $cost_fun_b$ calculates the right hand side forcing term of equation 4, namely $\left(\frac{\partial J}{\partial U}\right)^T$. In order to do that one has to set the values of $\bar{J} = 1$ and $\bar{u} = 0$. The serial version of the subroutines can now be recast into the mathematical equivalent as shown in listing (5) and (6).

Listing 5: Cost function primal (serial)

```
cost_fun(u, J):

  J = J(u)
```

Listing 6: Cost function adjoint (serial)

```
cost_fun_b(u, ū, J, J̄):

  ū = ū + (∂J/∂U)^T J̄
```

In parallel execution, each rank i calculates J_i local to the partition and an Allreduce sum is performed to obtain the global cost function value $J = \sum_i J_i$. Following a similar argument it can be shown (with reference to figure 3) that the cost-function adjoint calculates a local value of $\left(\frac{\partial J}{\partial U}\right)^T \bar{J}$ in each rank. The only inconsistency appears at the shared nodes, which require accumulation of the source term from neighbouring partition: listing (8). *Note that it is possible to completely avoid the shared-node accumulation inside cost-function by deferring and combining the accumulate with the adjoint residual as shown in listing (10).* Listing (9) is shown for comparison with accumulation in cost function. This idea can be extended to find the MPI adjoints of other types of cost functions like total pressure loss, entropy loss, etc.

Listing 7: Cost function primal (parallel)

```
cost_fun(u, J):

  J_i = J(u)
  J = Σ_i J_i
```

Listing 8: Cost function adjoint (parallel)

```
cost_fun_b(u, ū, J, J̄):

  ū = ū + (∂J/∂U)^T J̄
  call accumulate(ū)
```

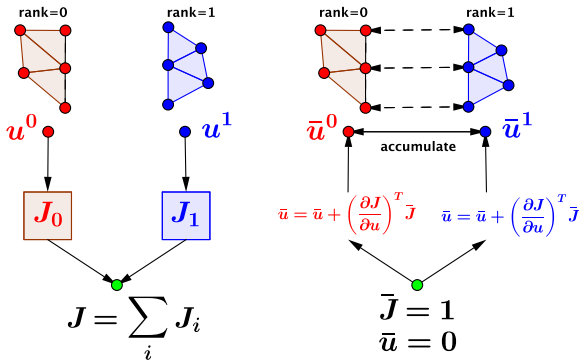


Figure 3: Schematic of cost function and its hand-assembled adjoint implementation

Listing 9: Adjoint FPI (accumulate)

```
J̄ = 1
call cost_fun_b(u, (∂J/∂u)^T J̄, J, J̄)

call accumulate((∂J/∂u)^T)
do iter = 1, n
  call residue_b(u, (∂R/∂u)^T v, R, v)

  call accumulate((∂R/∂u)^T v)

  R = (∂R/∂u)^T v - (∂J/∂u)^T
  call update(v, R)
end do
```

Listing 10: Adjoint FPI (no accumulate)

```
J̄ = 1
call cost_fun_b(u, (∂J/∂u)^T J̄, J, J̄)


do iter = 1, n
  call residue_b(u, (∂R/∂u)^T v, R, v)

  R = (∂R/∂u)^T v - (∂J/∂u)^T

  call accumulate(R)
  call update(v, R)
end do
```

3 Shared memory OpenMP parallelisation

The primal solver is using a shared-memory parallelisation strategy based on edge colouring as presented e.g. in [8, 9]. In this strategy, the edges are coloured in such a way that edges with the same colour do not share any nodes. More formally, for two edges $E_\alpha = (V_i, V_j)$ and $E_\beta = (V_k, V_l)$ we demand that

$$\text{colour}(E_\alpha) = \text{colour}(E_\beta) \Leftrightarrow i, j, k, l \text{ pairwise distinct} \tag{5}$$

We can then perform all flux updates on edges of the same colour simultaneously. This is illustrated in Figure 6. The adjoint CFD solver performs adjoint flux updates in the same fashion, i.e. as a loop over edges. The memory access pattern that describes read and write access of the primal and adjoint CFD solver are shown in Figure 4 and Figure 5.

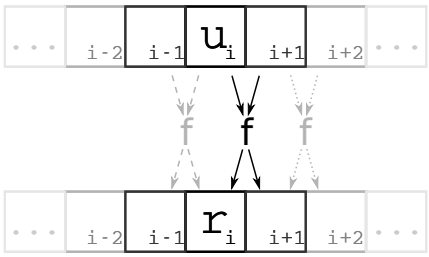


Figure 4: Primal communication pattern

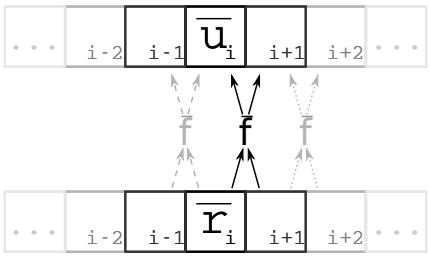


Figure 5: Adjoint communication pattern

Since the memory access of the adjoint and primal solver are identical, the mesh colouring that is used during the primal computation can be reused for the adjoint solver and helps to avoid write conflicts during the adjoint flux update in the same way as in the primal, see Figure 7 for an illustration.

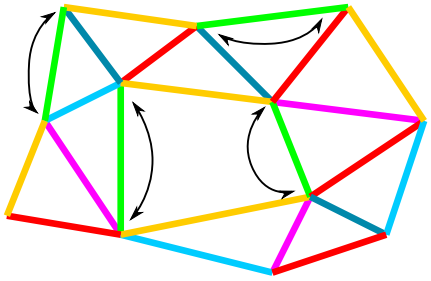


Figure 6: Edge colouring for primal shared-memory parallelisation

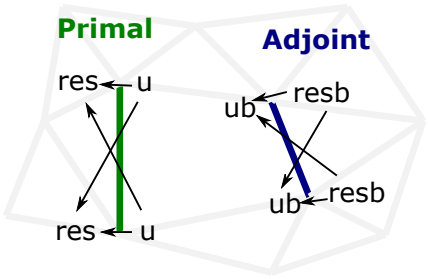


Figure 7: Equivalence of primal and adjoint communication

The AD tool Tapenade does not support OpenMP pragmas at the time of writing. We therefore employ *subroutine outlining* [10], a technique that encapsulates all private variables inside a subroutine and uses all shared variables as subroutine arguments. When this technique is used, the AD tool generates adjoint code in which only an OpenMP parallel loop with the `DEFAULT SHARED` clause has to be inserted in a postprocessing step.

Listing 11: Forward and reverse parallel flux computation

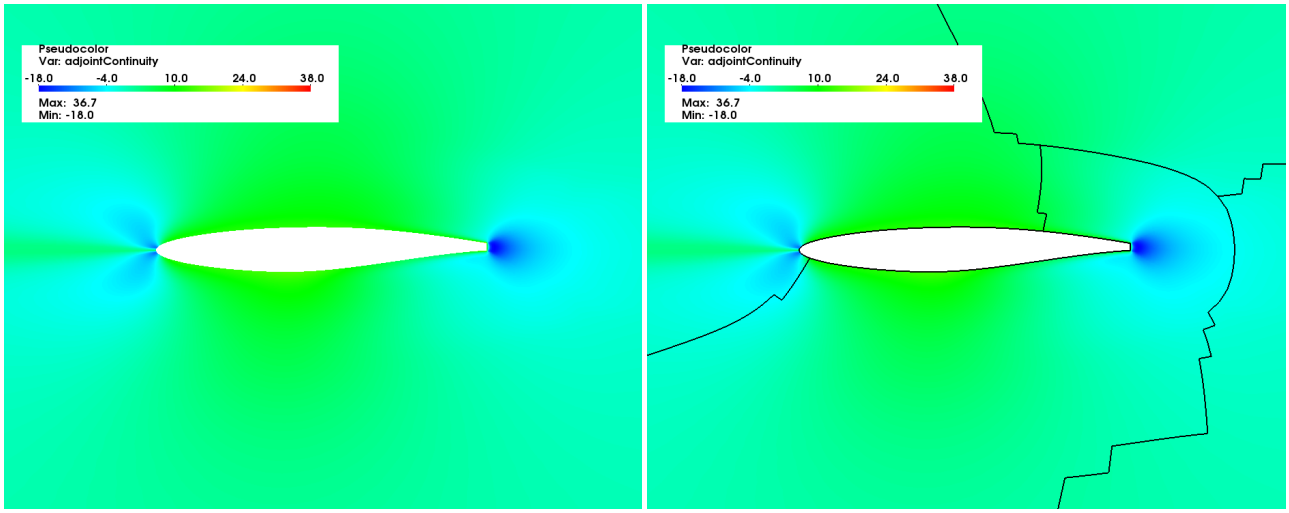
```
do colour=1,nColours !forward
  !$OMP PARALLEL DO DEFAULT SHARED
  do edge=firstEdge(colour),lastEdge(colour)
    call flux_loopbody(edge,res,u)
  end do
  !$OMP END PARALLEL DO
end do
do colour=nColours,1 !reverse
 !$OMP PARALLEL DO DEFAULT SHARED
 do edge=lastEdge(colour),firstEdge(colour)
    call flux_loopbody_b(edge,res,resb,u,ub)
 end do
 !$OMP END PARALLEL DO
end do
```

The flux computation at boundaries could be parallelised in a similar fashion. The primal solver uses a loop over boundary nodes and performs a computation using data from ghost nodes. Each ghost node is exclusively connected with one boundary node, and there is no communication with other nodes during the boundary flux computation. The parallel boundary treatment has not been used for this work.

4 Results

The primal and adjoint solvers in our in-house code $mgopt$ have been parallelised using both MPI and OpenMP paradigm. As an initial attempt, scaling of the inviscid first order primal and adjoint solver on subsonic flow ($M_\infty = 0.3, \alpha_{AOA} = 0^o$) over an $2d\ rae2822$ airfoil is shown. The cost function for the adjoint system is the aerodynamic drag on the airfoil. The mesh has $11,510$ cells and it is run on a four core Intel i7 processor. Pure MPI and OpenMP scaling is shown in figure 9(a) and a hybrid run of two MPI ranks each running two OMP threads within the ranks is compared against four MPI ranks and four OMP thread runs in figure 9(b). The baseline serial timing was obtained by running the solver with MPI code removed and using a single OMP thread. The relative error in the serial and parallel solution matches to machine precision for both primal and adjoint. The adjoint continuity for the serial and parallel case on

four partitions is shown in figure 8. The scaling results are not conclusive due to the small size of the test case. The scalability study for larger $3d$ meshes using the second order solver with the viscous terms is planned for future work.



(a) Serial run (b) Parallel run (on 4 partitions)

Figure 8: Drag adjoint continuity solution contour for subsonic flow over rae2822 airfoil ($M_\infty = 0.3, \alpha_{AOA} = 0^o$)

5 Conclusion

We have demonstrated that naive application of adjoint differentiation for MPI calls can fail for hand-assembled adjoint CFD solvers with zero-halo partitioning. We presented a correct reverse-differentation of MPI `allgather` calls in such scenarios in the context of an unstructured Finite Volume solver for compressible flow. In addition, we exploited the fact that memory access patterns for the primal and adjoint solvers are identical to implement shared-memory parallelism using OpenMP based on edge colouring. The hybrid parallelisation strategy was used in this work to obtain adjoint results based on a drag cost function for inviscid flow around a RAE2822 airfoil with truncated trailing edge. The results from serial and parallel execution runs were validated to be identical to machine precision, and scalability was presented for shared memory, distributed memory and hybrid parallelisation.

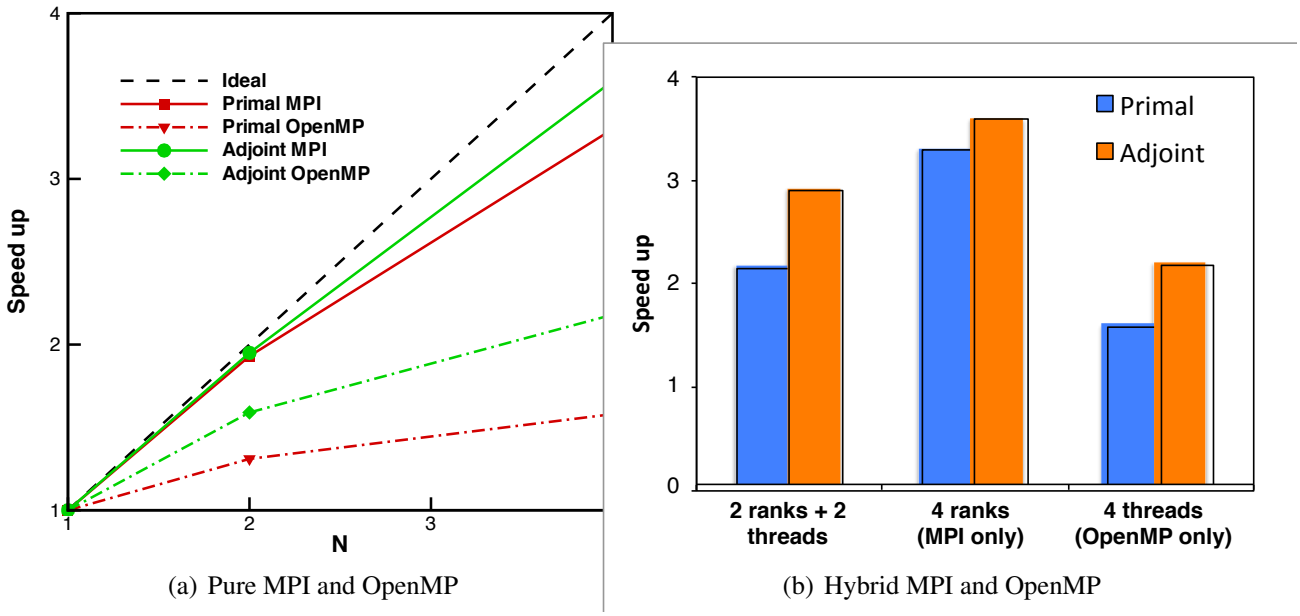
6 Acknowledgement

This research has been supported by the European Commission under the HORIZON 2020 Marie Curie fellowship (grant no. 642959).

REFERENCES

[1] J. Utke, L. Hascoet, P. Heimbach, C. Hill, P. Hovland, and U. Naumann, *Toward Adjoinable MPI*, IEEE International Symposium on Parallel & Distributed Processing, 2009.

[2] J. Utke, L. Hascoet, P. Heimbach, C. Hill, P. Hovland, U. Naumann, M. Schanen, and C. Hill, *Gradient of MPI-parallel codes*, slides from talk, June, 2012.



(a) Pure MPI and OpenMP

(b) Hybrid MPI and OpenMP

Figure 9: Strong scaling for the rae2822 on Intel i7 processor (four core)

[3] Y. Liu, *Hybrid Parallel Computation of OpenFOAM Solver on Multi-Core Cluster Systems*, Master of Science Thesis, KTH Sweden, 2011.

[4] C. Faidon, *Aerodynamic shape optimisation using adjoint sensitivities*, PhD Thesis, Queen Mary Univ. London, 2012.

[5] Hohenwarter, M. and Borcherds, M. and Ancsin, G. and Bencze, B. and Blossier, M. and Delobelle, A. and Denizet, C. and Éliás, J. and Fekete, Á and Gál, L. and Konečný, Z. and Kovács, Z. and Lizelfelner, S. and Parisse, B. and Sturr, G., *GeoGebra 4.4*, Dec 2013 (`http://www.geogebra.org`).

[6] P. Mohanamuraly, and K. Nagarajan, *Revisiting the space-filling curves for storage, reordering and partitioning mesh based data in scientific computing*, IEEE HiPC conference, Dec 2013.

[7] A. Griewank, and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd edition, SIAM, 2008.

[8] Guo X, Gorman G, Sunderland A et al, *Developing hybrid openmp-mpi parallelism for fluidity-next generation geophysical fluid modelling technology*.

[9] Giles MB, Mudalige GR, Sharif Z et al, *Performance analysis and optimization of the op2 framework on many-core architectures*, The Computer Journal, 2011

[10] Liao C, Hernandez O, Chapman B et al, *Openuh: An optimizing, portable openmp compiler*, Concurrency and Computation: Practice and Experience, 2007, 19(18): 2317–2332.

[11] H. Laurent, and P. Valérie, *The Tapenade Automatic Differentiation Tool: Principles, Model, and Specification*, ACM Trans. Math. Softw., vol. 39, no. 3, pp. 20:1-20:43, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EFFICIENT PARALLEL GEOMETRY DISTRIBUTION FOR THE SIMULATION OF COMPLEX FLOWS

Andreas Lintermann[1]

[1]JARA-HPC, RWTH Aachen University
Schinkelstr. 2, 52062 Aachen, Germany
e-mail: Lintermann@jara.rwth-aachen.de

Keywords: Parallel Geometry, Lattice-Boltzmann Method, Respiratory Flow, Computational Fluid Dynamics

Abstract. *Highly resolved intrinsic geometrical shapes used in three-dimensional parallel simulations of fluid flows consume a large portion of the available memory when loaded serially on every process. This demands for a memory efficient implementation of a distributed geometry which is however a non-trivial task when complex spatial domain decomposition methods for the flow domain are involved. To overcome this problem, an algorithm to generate a parallel geometry during the mesh generation is proposed that enables a low-memory subdivision of the geometry based on the decomposition of the flow field. The applied meshing method generates computational grids that can be used for simulations on a quasi-arbitrary number of cores on which the geometry is distributed in an efficient preprocessing step. This allows reducing the number of instances of the geometry in the global memory of the simulation to about one.*
The algorithm is used to generate a parallel geometry for a large shape consisting of $7 \cdot 10^6$ *triangles, i.e., for a geometry representing the whole respiratory tract down to the* 12^{th} *lung generation. For this case, performance and memory consumption measurements are given for simulations on 8,192 up to 131,072 cores and juxtaposed to results obtained from simulations using non-parallel geometries. The findings show that with the new method not only the memory usage could be reduced by the factors of 1,802 and 19,936 for core numbers of 8,192 and 131,072 but also a large speed-up factor of about 51 is obtained in the geometry I/O and preprocessing. Furthermore, the parallel geometry allows using the sweet spot with respect to a combination of distributed and shared memory parallelization leading to an increase of the computational speed of about 1.43.*

1 INTRODUCTION

The three-dimensional parallel simulation of fluid flows in and around objects using non-boundary fitting unstructured meshes requires the geometrical representation of the involved object surfaces to set the boundary conditions and to generate the computational mesh. Such geometries often stem from *Computer Aided Design* (*CAD*) software [1, 2, 3], from *Computer Tomography* (*CT*) images [4, 5, 6], or from laser-scan reconstructions of real-world objects. The number of representing surface elements varies with the resolution of the geometrical model. Hence, for numerical simulations of flows using highly-accurate geometrical representations from, e.g., technical or biomechanical background, the corresponding surface data may exceed the process-local memory or consume a large share of it in parallel simulations. That is, loading the complete geometry for such simulations limits the memory available for solving for the flow variables.

Body-fitted structured meshes have the advantage of not requiring a geometrical representation for preprocessing of the simulation and are well suited to resolve boundary layer flows [7, 8, 9, 10, 11]. However, they necessitate the geometry for the often time consuming manual construction of the computational mesh. In contrast, the generation of hierarchical *Cartesian* meshes can be executed completely automatically in a short amount of time [12, 13, 14]. Such meshes are well suited for complex geometries and are used for a variety of applications [2, 3, 6, 15, 16, 17, 18] in which the computational grids are generated from a geometrical surface representation given in *STereo Lithography* (*STL*) format. Lintermann et al. [6] and Freitas et al. [16] use a *Lattice-Boltzmann Method* (*LBM*) to simulate the biofluidmechanics of the human nasal cavity and the human lung. They use the no-slip *BFL*-rule from Bouzidi et al. [19] as wall boundary condition which requires the availability of the geometry for the distance calculation of boundary cell centers to the wall surface in the discrete *LBM* directions. In [2, 17], a *Finite-Volume Method* (*FVM*) is used for the three-dimensional simulation of the flow in a combustion engine. The same method is applied for the simulation of tip-leakage flows in fan geometries [3] and turbulent helicopter engine jets [18]. In these publications, a *cut-cell* method [2, 20] based on the geometry is used to accurately apply wall boundary conditions. Furthermore, in [2, 15, 17] a *level-set* mesh that allows for an efficient tracking of moving boundaries is constructed from the geometry. That is, the geometry is an essential element for pre- and in-simulation processing.

The geometry in most of these cases is small and consumes only a marginal amount of memory, e.g., one of the nasal cavity geometries in [6] consumes around 210 MB in process memory. However, considering a *High Performance Computer* (*HPC*) like the IBM BlueGene/Q *JUQUEEN* [21] at the *Jülich Supercomputing Center* (*JSC*), which has 1 GB per computational core available, about 20% of the global process memory is occupied by the nasal cavity geometry and cannot be used for the computation. Future computations for the simulation of flows in complex and large geometries may hence experience memory issues that can only be avoided by a parallel implementation of the geometry or by using less cores per node together with a shared memory parallelization.

In this work, a new algorithm to generate a parallel geometry from a *STL* file during the mesh generation process is presented. It is based on the hierarchical *Cartesian* mesh generator introduced in [14] and avoids wasting memory by distributing the geometry corresponding to the applied domain decomposition. This not only reduces the total and local amount of memory occupied by the geometry, but also the time of pre-simulation processing. Furthermore, the computational time is reduced due to an optimal usage of distributed and shared memory paral-

lelization. To the best of the author's knowledge, such an algorithm, especially for large-scale simulations applying complex spatial domain decomposition methods for the flow domain, has not been presented in the literature and may be beneficial for future simulations.

This paper is organized as follows. The numerical methods for the simulation of flows in intricate geometries, for the parallel grid generator, and for the generation of the parallel geometry are discussed in Sec. 2. Sec. 3 discusses the performance of the parallel geometry generation as well as the performance of the simulation method with and without a parallel geometry, and shows the applicability of the new method by running a simulation in a complex geometry. Finally, conclusions are drawn in Sec. 4.

2 NUMERICAL METHODS

The simulation of the flows presented in this paper makes use of the *Lattice-Boltzmann Method* (*LBM*) which is briefly described in Sec. 2.1. Since the generation of a parallel geometry is based on the parallel grid generator presented in [14], an introduction to the grid generator is given in Sec. 2.2. Then, the algorithms for the parallel geometry are discussed in Sec. 2.3.

2.1 Lattice-Boltzmann Method

The *LBM* considers the evolution of *particle probability density distribution functions* (*PPDFs*) f_i in phase space by solving the discretized *Boltzmann* equation, i.e., the *Lattice-Bhatnagar-Gross-Krook* equation [22, 23] given by

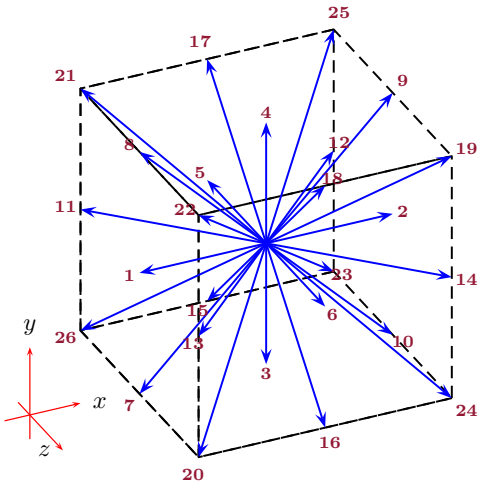
$$f_i\left(\mathbf{x}+\xi_i\delta t, t+\delta t\right) = f_i\left(\mathbf{x},t\right) + \omega\delta t\cdot\left[f_i^{eq}\left[\mathbf{x},t\right) - f_i\left(\mathbf{x},t\right)\right], \tag{1}$$

where $\mathbf{x}$ represents the grid location, ξ_i is the grid velocity vector, t is the time, and δt is the time increment. The equation describes a relaxation towards a thermodynamical equilibrium by the collision operator on the right hand side. That is, the system is relaxed towards the *Maxwell* equilibrium distribution function f_i^{eq} with a collision frequency given by $\omega\delta t$. The left hand side of Eq. 1 describes the streaming operation in which the collision information is propagated to the neighboring grid locations. The index $i \in [0,\ldots,18]$ represents a discrete direction in three dimensions in the *D3Q19* discretization model [24] as shown in Fig. 1.
In Sec. 3, the simulation of the flow in the human respiratory tract at inspiration is used as benchmark case to measure the performance of the code. It uses an adaptive pressure *Dirichlet* condition at the outlets and the *Saint-Vernant and Wantzel* extrapolation boundary condition for the pressure at the inlets [6]. On both sides, at the inlets and outlets, a *Neumann* condition is used for the velocity. For the wall, the no-slip *BFL*-rule [19] is imposed. For more details on the *LBM* and the boundary conditions the reader is referred to [6].

2.2 Parallel grid generation

The parallel grid generator [14] enables a fully automatic creation of large-scale computational *Cartesian* meshes stored in a hierarchical octree in a short amount of time on hundreds of thousand of processes. For process communication it makes use of the *Message Passing Interface* (*MPI*) and I/O is performed by the *parallel NetCDF* library [25]. The meshing algorithm consist of three basic steps:

Step 1 - serial generation of a coarse, uniformly refined mesh
Step 2 - decomposition of this coarse mesh and parallel uniform refinement
Step 3 - local boundary and patch refinement



$$\xi_i = \xi_0 \begin{cases} (\mp 1,0,0),(0,\mp 1,0),(0,0,\mp 1), & i = 1,\ldots,6 \\ (\mp 1,\mp 1,0),(\mp 1,0,\mp 1),(0,\mp 1,\mp 1), & i = 7,\ldots,18 \\ \begin{cases}(0,0,0),\end{cases} & i = 19,\text{(D3Q19)} \\ \begin{cases}(\mp 1,\mp 1,\mp 1),\\(0,0,0),\end{cases} & \begin{matrix} i = 19,\ldots,26,\text{(D3Q19)} \\ i = 27,\text{(D3Q27)} \end{matrix} \end{cases}$$

Figure 1: Three-dimensional *D3Qm* discretization model of the *LBM*, where $m \in \{15, 19, 27\}$.

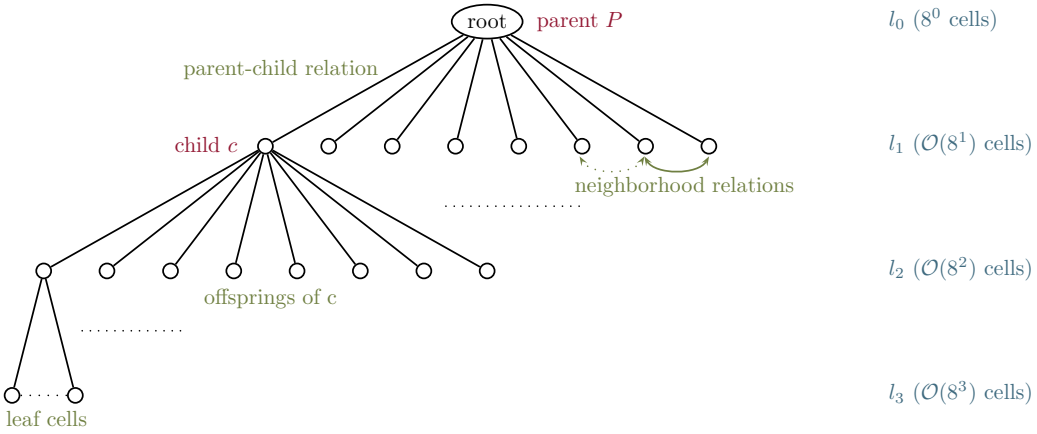


Figure 2: Hierarchical octree of the *Cartesian* mesh including parent-child relations (links between the levels) and the neighborhood relations among cells on equal levels. The upper bound of the number of cells is given by the maximum level, i.e., for l_3 the number of cells is bound by $\mathcal{O}(8^3)$.

In ***Step 1***, the geometry available in *STL* format is first loaded from disk and stored in an *Alternating Digital Tree* (*ADT*) [26]. Then, each *MPI*-process starts to serially refine an initial cube that is placed around the geometry. During each refinement step, the newly generated cells are inserted into an octree (see Fig. 2), where the links between the consecutive levels l_k, l_{k+1} constitute a parent-child relationship and neighborhood relations among cells on the same level are recorded. As depicted in Fig. 3, cells outside the geometry are removed from the octree by tagging cells intersecting with the geometry as boundary cells $c_b \in C_b$ and by using a recursive fill algorithm originating at an inside cell c_{in} to color the cells enclosed by the boundary cell hull. The cells of C_b are determined by an intersection test with candidate triangles retrieved from the *ADT* in $\log$-time. The corresponding inside/outside determination for the originating cell c_{in} is performed using a ray intersection test with the geometry yielding an inside cell for an odd number of intersections and an outside cell for an even number of intersections. The serial refinement is performed until a user-defined level l_α has been reached. At the end of this step, all levels except l_α are removed from the octree.
The remaining cells on level l_α are partitioned in ***Step 2*** using a *Hilbert* decomposition [27] (see Fig. 4). The neighboring processes are determined by a balanced decomposition along the

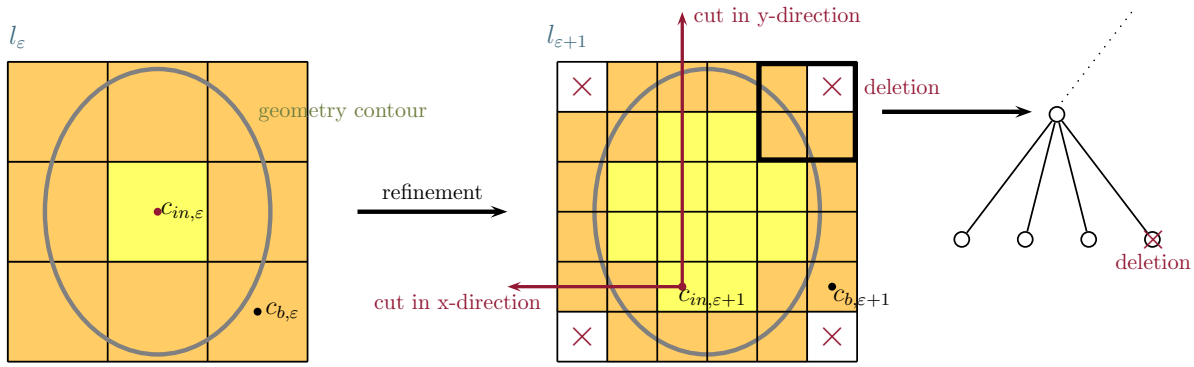


Figure 3: 2D inside/outside detection of cells and removal from the octree. Cells with a geometry intersection on levels l_ε and $l_{\varepsilon+1}$ are marked as $c_{b,\varepsilon}$ and $c_{b,\varepsilon+1}$ (orange). From the inside cells $c_{in,\varepsilon}$ and $c_{in,\varepsilon+1}$, detected by counting ray geometry intersections, inside cells (yellow) are recursively labeled and all non-labeled cells are removed.

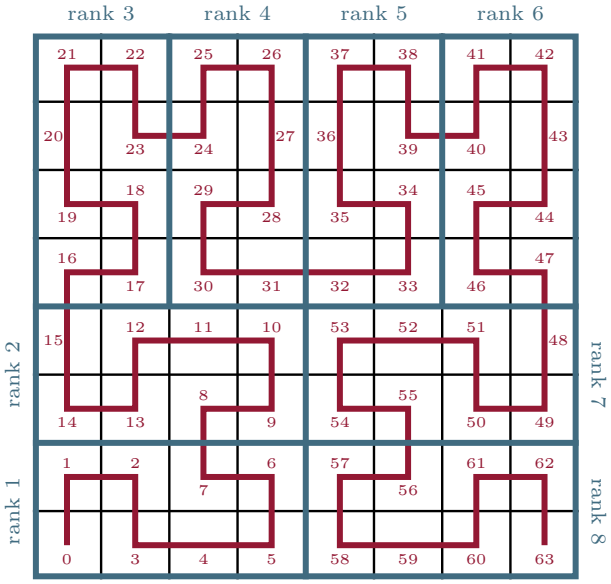
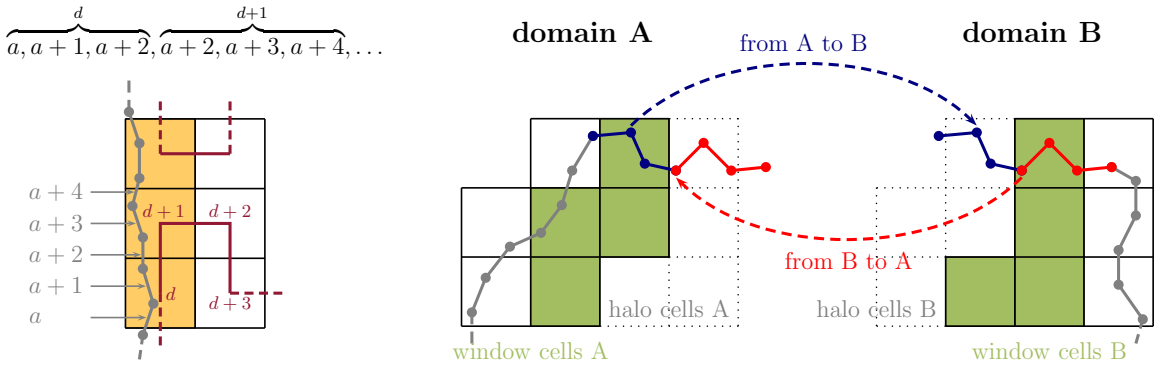


Figure 4: Two-dimensional *Hilbert* curve (red) and the corresponding decomposition (blue) based on the *Hilbert* id.

Hilbert curve and the respective cell neighborhood relations. Then, each process removes the process-foreign cells and continues with parallel refinement of the remaining cells including the removal of outliers until a level l_β has been reached.

In the last ***Step 3***, local refinement based on user-defined patches and/or the boundary distance is performed. For patch refinement, cells inside a previously defined patch geometry are tagged and are then refined. Local boundary refinement is based on a user-defined distance from the original geometry. The algorithm recursively counts the cell distance from the geometry surface across processes and based on the layer count it refines the respective cells. This step is repeated on the finer levels until a final level l_γ has been reached that guarantees a maximum level distance between neighboring cells of one, i.e., $\Delta l = l(c_d) - l(c_e) = 1$ for neighboring cells c_e and c_d. Subsequently, it is necessary to translate local cell ids into global cell ids, which is done by locally counting the number of cells, sharing this information among all processes, and by determining the local offsets from these numbers. Note that the cells below the level l_α are ordered in depth-first manner. That is, the cells on l_α are ordered by the *Hilbert* curve



(a) Triangle detection for the definition of a parallel geometry file.

(b) Triangle exchange between two neighboring domains with one window and halo layer. Each triangle carries an original id such that duplicates are avoided.

Figure 5: Generation of a parallel geometry and window and halo cell exchange in the initialization phase of the simulation.

while for all levels l_ζ with $\zeta > \alpha$ the ordering follows a z-curve. The coarse level cells ordered by the *Hilbert* curve are additionally stored in a list $\mathcal{L}$ which is later used in the simulation to decompose the domain with the help of workloads stored per cell. Furthermore, to avoid large subtrees originating from cells in $\mathcal{L}$, cells with a number of descendants larger than a specified threshold t_d are replaced by their direct children in this list. Finally, the mesh is written to disk in parallel using *parallel NetCDF* [25]. The output of the grid generator enables to run simulations on a quasi-arbitrary number of processes and is only limited by the number of elements in $\mathcal{L}$ and by the available memory per *MPI*-rank.
For further details on the meshing algorithm and its parallel performance, the reader is referred to [14].

2.3 Parallel geometry processing

The exploitation of parallelism for geometries involves the generation of a parallel geometry file from serial information. Furthermore, suitable initializations and boundary conditions for parallel geometries are necessary. In the following, the corresponding algorithms are explained.

Generation of a parallel geometry The order of the cells in the list $\mathcal{L}$ constructed in ***Step 3*** (Sec. 2.2) and the respective workloads constitute the spatial decomposition of the flow domain for a parallel computation. It ensures process-locality for the cells and can hence also be used to generate a parallel geometry that allows for a process-locality of the corresponding triangles. For this purpose, after the generation of $\mathcal{L}$, its elements $c_\mathcal{L}$ are traversed and the original triangle ids $o(T_i)$ of the geometry inside each such cell is collected in a variable two dimensional array, where candidate triangles are retrieved from the *ADT* (see Fig. 5(a)). The first dimension of this array is given by the number of cells $|\mathcal{L}|$ of $\mathcal{L}$ and the second dimension holds the ids

$$o(c_\mathcal{L}) = \{o(T_i) : \exists \mathbf{c}(T_i)\,(\mathbf{c}(T_i) \text{ inside } c_\mathcal{L})\}, \tag{2}$$

where $\mathbf{c}(T_i)$ is a vertex in T_i. Note that elements of $o(c_\mathcal{L})$ may be associated with more than one cell as the respective triangle vertices may be located in different cells. To this end, each process holds a list of triangles incidental to the local fraction of $\mathcal{L}$. Finally, the number of triangles per cell $\Xi(c_\mathcal{L})$, the original triangle ids $o(c_\mathcal{L})$, the vertex coordinates $\mathbf{c}$, and the triangle normals $\mathbf{n}$ are written to disk using *parallel NetCDF* [25]. In sum, the information given in Tab. 1 is

no.	name	meaning	no. elements	size
1	`vertices[3][3]`	triangle coordinates	$\|o(c_\mathcal{L})\|$	9×`double`
2	`normal[3]`	normal	$\|o(c_\mathcal{L})\|$	3×`double`
3	`segmentId`	geometry segment id	$\|o(c_\mathcal{L})\|$	`int`
4	`originalTriId`	original triangle id $o(c_\mathcal{L})$	$\|o(c_\mathcal{L})\|$	`int`
5	`noTriPerMinCell`	number of triangles per cell $\Xi(c_\mathcal{L})$	$\|\mathcal{L}\|$	`int`
6	`noRealTri`	size of $o(c_\mathcal{L})$	1	`int`
7	`boundingBox[6]`	min. and max. of T_i	-	6×`double`
8	`BC`	the boundary condition id	-	`int`

Table 1: Triangle variables and their sizes. The information $1,\ldots,6$ is written to disk, while 7 and 8 are assigned in the solver preprocessing.

stored per triangle in the file. As will be shown in Sec. 3.3 the overhead for the creation of the parallel geometry is negligibly small compared to the mesh generation and the computation of the flow. The parallel geometry file is larger than the original *STL* file since triangles may be listed several times due to the aforementioned multiple cell affiliations. The file size ratio of the *parallel NetCDF* and the *STL* file

$$\sigma = \sigma_{NetCDF}/\sigma_{STL} \tag{3}$$

is on the one hand determined by the location of the triangle vertices and on the other hand by the size ratio of the triangles and the grid distance δ_x. Note that in current simulation setups as they are used, e.g., in the simulation of the flow in the human nasal cavity [6], the file size of both, the *STL* and the *NetCDF* file, is of the order of $\mathcal{O}(1\ GB)$. The spatial resolution δ_x and the geometry resolution usually defines ratios σ of the order of $\mathcal{O}(10)$ for such simulations, i.e., the geometrical resolution is often only slightly higher than the mesh resolution. The results presented in Sec. 3.3 will show that a ratio of $\sigma = 1.92$ is achieved for a highly resolved respiratory tract geometry and a large-scale computational mesh.

Simulation initialization using a parallel geometry Due to the parallel nature of the geometry, some special treatment is required for the initialization of the simulation. This treatment is covered in three major steps:

Step a - parallel reading of the geometry file
Step b - triangle exchange for the halo cells
Step c - boundary condition treatment

In ***Step a***, all *MPI*-ranks open the parallel geometry and the corresponding sub-grid file. From the grid file, the workload array is read and used to find the optimal partition of the cells in $\mathcal{L}$ (cf. [14]). This partition is then used to subdivide the triangles stored in the parallel geometry file. Subsequently, each process reads the triangles based on the number of triangles per cell $\Xi(c_\mathcal{L}) = |o(c_\mathcal{L})|$ in $\mathcal{L}$.
During the initialization of the mesh, the window cells, which are the neighbors to cells on a neighboring *MPI*-rank, are identified and a copy of these cells is created on the neighbor domains as halo cells. The triangles in each such window cell in $\mathcal{L}$ need to be transferred to the respective process neighbors to ensure triangle consistency among all processes. Therefore, in ***Step b*** the original triangle ids $o(c_{\mathcal{L},r})$ for the neighboring rank r are first transferred for all cells $\mathcal{L}_r \subseteq \mathcal{L}$ to r. Then, on process r it is checked which elements of $o(c_{\mathcal{L},r})$ are already present on

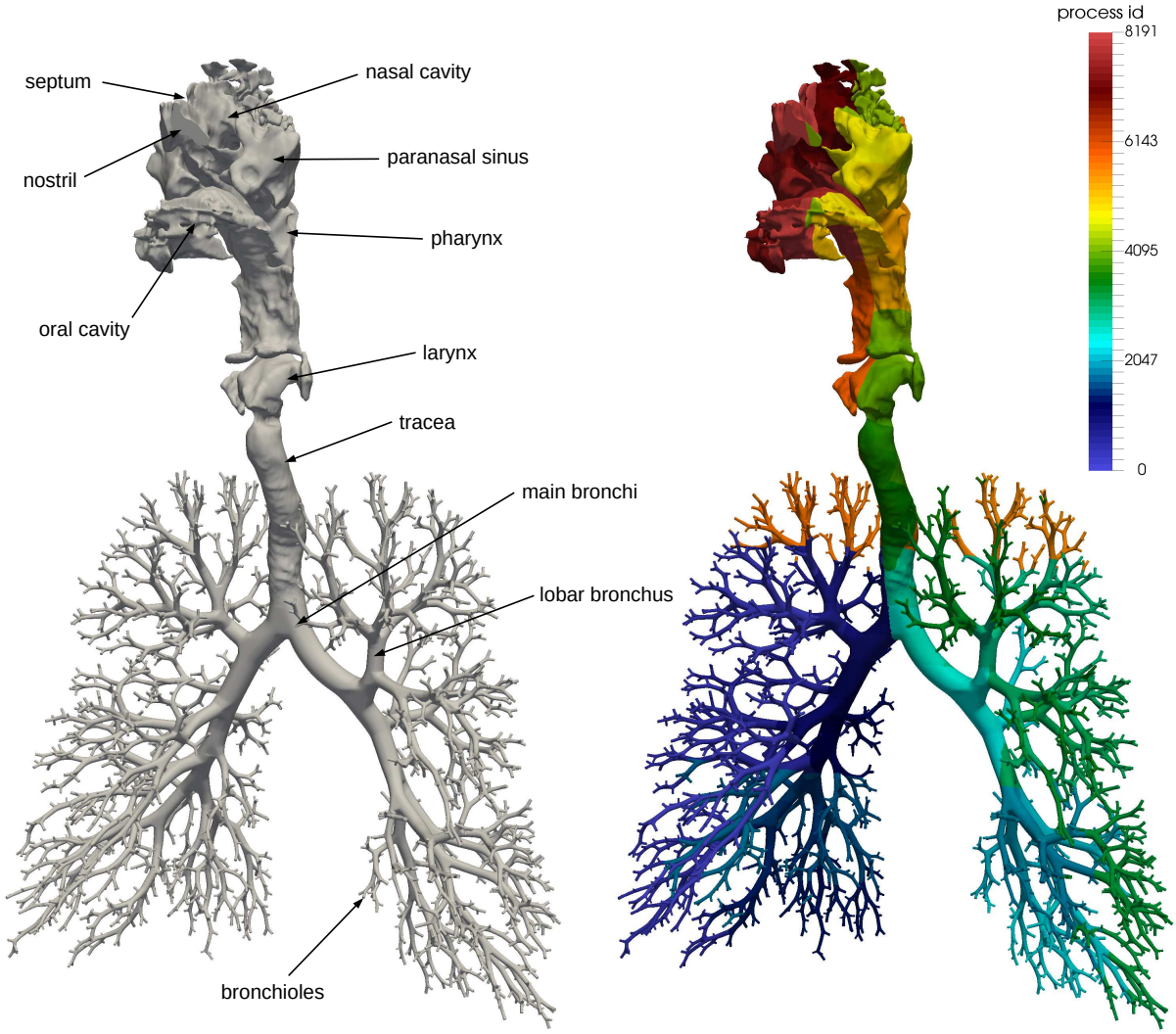


Figure 6: On the left, the geometry of the respiratory tract used for the memory and performance measurements is shown. The right side shows the parallel geometry. Each colored patch corresponds to a single process-id. In total, 8192 cores have been used for the parallel geometry in this case, i.e., there exist at maximum 8192 colored patches.

this domain and which are missing. Subsequently, the originating process is notified about the missing triangles and sends them to r. To this end, all processes have all triangles available for later geometry processing. Fig. 5(b) shows the situation for two neighboring domains.
Some boundary conditions of the flow simulation need special treatment when applying a parallel geometry. For example, the hydraulic diameter $D_h = 4A/C$, based on the area A and the circumference C of a surface segment Ω_s, e.g., of an inlet geometry, is often used as reference length in the REYNOLDS number $Re = (u \cdot D_h)/\nu$, where u is a reference velocity and ν is the kinematic fluid viscosity. Furthermore, the minimal distances of boundary cell centers to the boundary of Ω_s are required to prescribe boundary conditions like parabolic *Dirichlet* inflow profiles for the velocity. The calculation of these values hence necessitates the process-locality of the whole geometry segment, however, Ω_s may be shared between different processes. To overcome these problems, the required values need to be transferred between the different processes in ***Step c***. That is, for a set of shared processes $S = \{p_a, \ldots, p_b\}, 0 \leq a < b \leq n$, where n is the total number of *MPI*-ranks and $p_a, \ldots, p_b$ define a *MPI* sub-communicator, the area is

calculated by

$$A = \sum_{i=p_a}^{p_b} \sum_{k=0}^{l} A_{i,k}, \tag{4}$$

where $A_{i,k}$ is the area of the k^{th} triangle on process i in Ω_s. To obtain the circumference the length of all boundaries of Ω_s is summed

$$C = \sum_{i=p_a}^{p_b} C_i, \tag{5}$$

where C_i is calculated by the sum of the lengths of the edges of the triangles in Ω_s that are incident to only one triangle (boundary edges). However, to calculate the distances it is necessary to have the whole boundary of Ω_s available on $p_a, \ldots, p_b$. Therefore, the boundary edges are gathered on all processes of the corresponding *MPI* sub-communicator.

Note that due to the reduced number of triangles in the process-local *ADT*s, searches, as they are performed for the wall distance calculation of the *LBM BFL*-rule, are at average reduced from

$$\log_2\left(\mathcal{O}(|T_s|)\right) \quad \text{to} \quad \log_2\left(\frac{\mathcal{O}(|T_p|)}{n}\right), \tag{6}$$

where $|T_s|$ is the number of global triangles in the serial case, $|T_p|$ is the number of triangles in the parallel case, and n is the number of *MPI*-ranks. The number of triangles in T_p is usually of the order of $o(|T_s|)$.

3 RESULTS

In this section, the total and the per-process memory consumption for a simulation Ψ of the flow in the complete human respiratory system down to the 12^{th} lung generation using the two configurations of a parallel and a non-parallel geometry are investigated in Sec. 3.1. In addition to the memory consumption, the performance of the simulation preprocessing for the computations is juxtaposed for both, the non-parallel and the parallel geometry version, in Sec. 3.2. Finally, in Sec. 3.3 a brief analysis of the flow field obtained from the *LBM* simulation using the parallel geometry is presented.
The respiratory system geometry consists of $7 \cdot 10^6$ triangles and consumes 645 MB in *NetCDF* format and 334 MB in 32-bit binary *STL* format on disk. The upper respiratory tract down to the end of the trachea is reconstructed from *CT* image data using the methods described in [28]. The image has a spatial transverse, coronal, and saggital resolution of 0.49 mm, 0.49 mm, and 0.7 mm and was recorded by a Siemens *CT* scanner. The lower respiratory tract is generically generated by using the software *Lung4Cer* [29] and has been glued to the upper respiratory tract surface. The geometry of Ψ is stored in 1,747 files (two inlets, 1,744 outlets, and the wall). The corresponding parallel geometry file has a size of $1,239$ MB, i.e., the respective *STL* in *NetCDF* format to parallel geometry *NetCDF* file size ratio is $\sigma = 1.92$. The information on the parallel geometry is stored in a single file. The geometry of the respiratory tract is shown on the left side in Fig. 6.
All measurements were performed on the IBM BlueGene/Q *JUQUEEN* [21] system at the *JSC*. The *JUQUEEN* consists of 28,672 nodes containing IBM PowerPC A2 CPUs at 1.6 GHz, 16 cores, and 16 GB of RAM per node. The overall peak performance is 5.9 PFlops. It is capable of running a maximum number of 4 *OpenMP* threads per core.

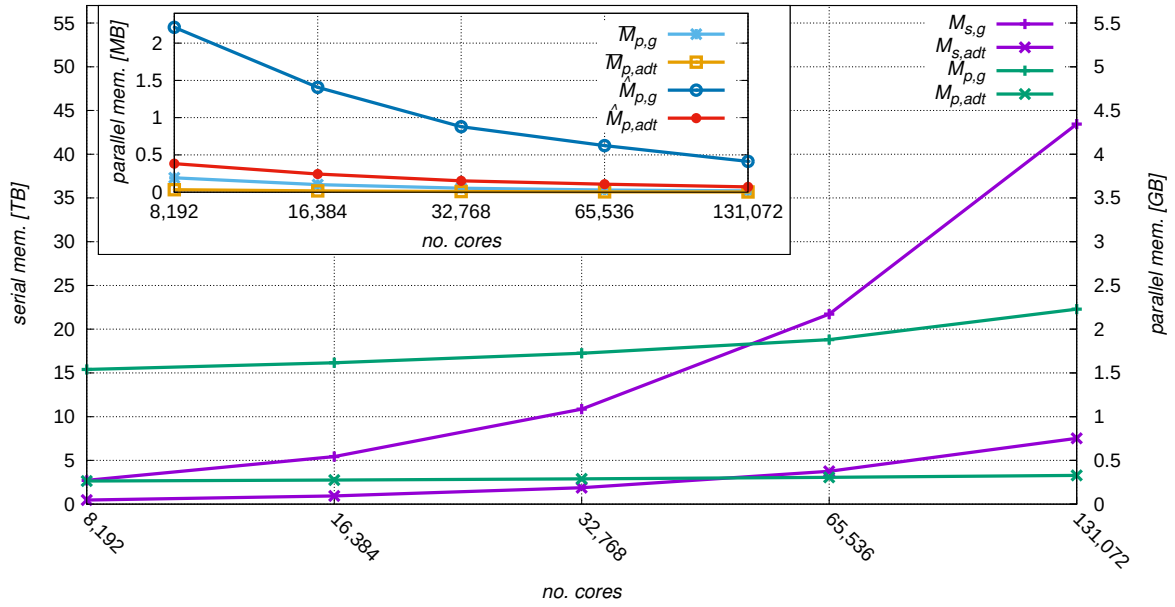


Figure 7: Total memory consumption of the simulations Ψ_s and Ψ_p on $8,192,\ldots,131,072$ cores. The graphs for the serial geometry and the serial *ADT* (purple) use the left ordinate scale counting in TB ($\mathcal{M}_{s,g}$ and $\mathcal{M}_{s,adt}$), while the graphs for parallel geometry and parallel *ADT* ($\mathcal{M}_{p,g}$ and $\mathcal{M}_{p,adt}$, green) use the right ordinate counting in GB. The inset shows the per-process averaged ($\overline{\mathcal{M}}_{p,g}$ and $\overline{\mathcal{M}}_{p,adt}$) and maximum ($\hat{\mathcal{M}}_{p,g}$ and $\hat{\mathcal{M}}_{p,adt}$) memory allocations for the parallel geometry and the *ADT*.

3.1 Memory consumption

For the simulation of the respiratory tract, memory measurements using the non-parallel geometry (simulation Ψ_s) and the parallel geometry (simulation Ψ_p) are performed. For both simulation cases, jobs using 8,192, 16,384, 32,768, 65,536, and 131,072 cores are submitted. Since the geometry consumes approx. $1,390$ MB in memory, *MPI*-processes are started on only every fourth core of the *JUQUEEN* nodes using 16 *OpenMP* threads for Ψ_s. This allows having enough memory available for the simulation. Note that in addition to the variables stored in the *parallel NetCDF* file, further variables are allocated per triangle, i.e., the bounding box, and the boundary condition id (see Tab. 1). For Ψ_p 16 *MPI*-ranks per node and 4 *OpenMP* threads are used for the simulations. The right image in Fig. 6 shows the corresponding triangle distribution on 8,192 cores. Fig. 7 and Tab. 2 show the results of the memory measurements for Ψ, where the left ordinate in Fig. 7 represents the memory consumption $\mathcal{M}_s$ in TB for the computations using the serial geometry and the right ordinate the consumption $\mathcal{M}_p$ of the parallel geometry approach in GB. While the amount of memory consumed by the serial geometry stays constant across all domains, the total allocated amounts of memory are $\mathcal{M}_{s,g}(8,192) \approx 2780.7$ GB and $\mathcal{M}_{s,adt}(8,192) \approx 481.3$ GB for the geometry itself and the *ADT* (left ordinate in Fig. 7), in sum $\mathcal{M}_{s,\Sigma}(8,192) = 3261$ GB ≈ 3.19 TB. Doubling the number of cores results in a doubling of the allocated memory leading to a total amount of memory on 131,072 cores of $\mathcal{M}_{s,\Sigma}(131,072) = 52191.3$ GB ≈ 51 TB for Ψ_s. In contrast, the total amount of consumed memory of the parallel geometry increases slightly from $\mathcal{M}_{p,g}(8,192) = 1.54$ GB for 8,192 cores to $\mathcal{M}_{p,g}(131,072) = 2.28$ GB for 131,072 cores (right ordinate in Fig. 7). The same trend is visible for the *ADT*, with an increase from $\mathcal{M}_{p,adt}(8,192) = 0.27$ GB to $\mathcal{M}_{p,adt}(131,072) \approx 0.34$ GB. Hence, the sum $\mathcal{M}_{p,\Sigma}$ for all core number also increases from $\mathcal{M}_{p,\Sigma}(8,192) = 1.81$ GB to $\mathcal{M}_{p,\Sigma}(131.072) \approx 2.62$ GB. The overall slight upward tendency is caused by the increase of the local triangle count

simulation	$\mathcal{M}_g$ [GB]	$\mathcal{M}_{adt}$ [GB]	$\mathcal{M}_\Sigma$ [GB]	$\overline{\mathcal{M}}_g$ [MB]	$\overline{\mathcal{M}}_{adt}$ [MB]	$\hat{\mathcal{M}}_g$ [MB]	$\hat{\mathcal{M}}_{adt}$ [MB]
$\Psi_s(8,192)$	2,780.684	481.271	3,261.955	-	-	-	-
$\Psi_s(16,384)$	5,561.368	962.542	6,523.910	-	-	-	-
$\Psi_s(32,768)$	11,122.736	1,925.084	13,047.820	-	-	-	-
$\Psi_s(65,536)$	22,245.472	3,850.168	26,095.640	-	-	-	-
$\Psi_s(131,072)$	44,490.944	7,700.336	52,191.280	-	-	-	-
$\Psi_p(8,192)$	1.540	0.270	1.810	0.193	0.033	2.212	0.383
$\Psi_p(16,384)$	1.655	0.282	1.937	0.101	0.017	1.406	0.243
$\Psi_p(32,768)$	1.766	0.296	2.062	0.054	0.009	0.879	0.152
$\Psi_p(65,536)$	1.925	0.314	2.239	0.029	0.005	0.625	0.108
$\Psi_p(131,072)$	2.282	0.336	2.618	0.017	0.003	0.413	0.071

Table 2: Absolute geometry memory consumptions of the simulations Ψ_s and Ψ_p performed on $8,192,\ldots,131,072$ cores of the *JUQUEEN* system. The table shows the total global memory allocation for the geometry and the *ADT* $\mathcal{M}_g$, $\mathcal{M}_{adt}$ and the accumulated memory allocation $\mathcal{M}_\Sigma$ in GB. Furthermore, the process-averaged $\overline{\mathcal{M}}_g$ and $\overline{\mathcal{M}}_{adt}$ and the maxima $\hat{\mathcal{M}}_g$ and $\hat{\mathcal{M}}_{adt}$ are given in MB.

for higher core numbers as a consequence of the increased triangle exchange in the halo cells. The inset of Fig. 7 and Tab. 2 additionally give the process-averaged $\overline{\mathcal{M}}_{\{g,adt\}}$ and maximum $\hat{\mathcal{M}}_{\{g,adt\}}$ memory allocation of the geometry and the *ADT*. It is clearly visible that due to the scattering of the triangles across all domains, the local memory usage is strongly decreased. To this end, the global memory consumption of the simulation Ψ is reduced by a factor of

$$R_{\mathcal{M}}\left[\Psi(8,192)\right] = \frac{\mathcal{M}_{s,\Sigma}(8,192)}{\mathcal{M}_{p,\Sigma}(8,192)} \approx 1,802 \tag{7}$$

using the same number of cores of 8,192 for the serial and the parallel geometry simulation. This factor is with

$$R_{\mathcal{M}}\left[\Psi(131,072)\right] = \frac{\mathcal{M}_{s,\Sigma}(131,072)}{\mathcal{M}_{p,\Sigma}(131,072)} = 19,936 \tag{8}$$

even higher for 131,072 cores.

3.2 Preprocessing performance

The same simulations as analyzed in Sec. 3.1 are used to juxtapose the preprocessing performance of the serial and parallel geometry version, i.e., for Ψ_s and Ψ_p on 8,192 up to 131,072 cores on the *JUQUEEN* system. The analysis considers the performance of the code sections that are associated with the geometry. That is, the performance of the geometry I/O, the calculation of the bounding box for each triangle, the setup of the *ADT*, the calculation of the normal vectors for inlets and outlets, and the *LBM* wall-distance calculation for wall-boundary cells required for the *BFL*-rule [19], is measured. Note that due to the equal processing time for the serial geometry, only the results obtained on 8,192 cores for Ψ_s are juxtaposed to the parallel geometry results. Fig. 8 and Tab. 3 show the results of the performance measurements for Ψ_s and Ψ_p. It is clearly visible that in $\Psi_s(8,192)$ the reading of the geometry consumes with $\mathcal{T}_{s,rdg}(8192) \approx 763.7$ s most of the time. This is caused by the serial reading from disk of

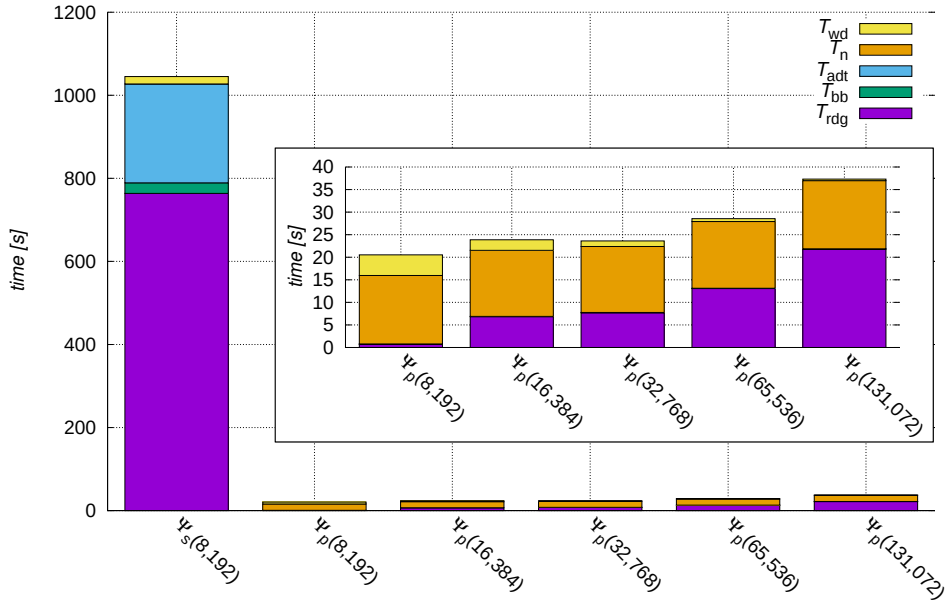


Figure 8: Performance of different code sections involving geometry treatment during the processing of simulation Ψ on different core numbers $8, 192, \ldots, 131,072$ using a serial and parallel geometry. The code section times are recorded for the wall-distance calculation $\mathcal{T}_{wd}$, normal calculation $\mathcal{T}_n$, setting up the *ADT* $\mathcal{T}_{adt}$, calculating the bounding boxes of all triangles $\mathcal{T}_{bb}$, and reading and distributing the geometry $\mathcal{T}_{rdg}$. The inset shows a zoom of the parallel code performance.

	$\mathcal{T}_{wd}$ [s]	$\mathcal{T}_n$ [s]	$\mathcal{T}_{adt}$ [s]	$\mathcal{T}_{bb}$ [s]	$\mathcal{T}_{rdg}$ [s]	$\mathcal{T}_\Sigma$ [s]
$\Psi_s(8,192)$	17.882	0.011	238.175	25.447	763.673	1,045.188
$\Psi_p(8,192)$	4.584	15.168	0.020	0.077	0.685	20.534
$\Psi_p(16,384)$	2.350	14.628	0.007	0.075	6.820	23.880
$\Psi_p(32,768)$	1.217	14.633	0.002	0.073	7.684	23.609
$\Psi_p(65,536)$	0.637	14.784	0.001	0.077	13.057	28.556
$\Psi_p(131,072)$	0.337	15.110	0.0003	0.077	21.794	37.318

Table 3: Absolute runtime in seconds for preprocessing algorithms of the simulations using a serial geometry Ψ_s performed on 8192 cores and an the parallel geometry Ψ_p executed on $8, 192, \ldots, 131,072$ cores of the *JUQUEEN* system. The total runtime $\mathcal{T}_\Sigma$ is subdivided into the time required for the wall-distance calculation $\mathcal{T}_{wd}$, normal calculation $\mathcal{T}_n$, setting up the *ADT* $\mathcal{T}_{adt}$, calculating the bounding boxes of all triangles $\mathcal{T}_{bb}$, and reading and distributing the geometry $\mathcal{T}_{rdg}$.

all the 1747 files containing the $7 \cdot 10^6$ triangles on all processes. The second most expensive algorithm is with $\mathcal{T}_{s,adt}(8,192) \approx 238.2$ s the setup of the *ADT*. Compared to these values, the durations of the wall-distance calculation $\mathcal{T}_{s,wd}(8,192)$, normal calculation $\mathcal{T}_{s,n}(8,192)$, and bounding box calculation $\mathcal{T}_{s,bb}(8,192)$ are rather small. The total required time is given by $\mathcal{T}_{s,\Sigma}(8,192) \approx 1045.2$ s. Using the parallel geometry reduces the preprocessing time of Ψ drastically, i.e., using the same amount of cores the total runtime is reduced from $\mathcal{T}_{s,\Sigma}(8,192)$ to $\mathcal{T}_{p,\Sigma}(8,192) \approx 20.6$ s by a factor of

$$R_\mathcal{T}\left[\Psi(8,192)\right] = \frac{\mathcal{T}_{s,\Sigma}(8,192)}{\mathcal{T}_{p,\Sigma}(8,192)} \approx 51. \tag{9}$$

The inset in Fig. 8 shows a zoom of the parallel geometry runtimes. Due to the increase of the number of computational subdomains using more cores the reading process of the *parallel NetCDF* library slows down (see also Tab. 3). Furthermore, an increased number of triangles needs to be exchanged between *MPI*-ranks due to the growth of the global amount of halo cells for higher core numbers. As it is evident from Tab. 8, the normal calculation stays almost constant for all core counts. This effect can be explained by the fact that the total number of geometry segments that require a normal computation (1746 inlets/outlets) stays constant. Most of the computation is performed locally and hence the influence of the augmentation of the number of sharing processes is small. A comparison of $\mathcal{T}_{s,n}(8192) = 0.011$ s with those of the parallel version shows that the serial normal calculation is much faster since it does not require any communication and an inside/outside detection can directly be performed locally. Finally, looking at the other times, it is obvious that due to the decreasing number of triangles per process under an increasing number of cores the setup of the *ADT* and the bounding box calculations are negligible. Furthermore, the time for the wall-distance calculation continuously decreases from $\mathcal{T}_{p,wd}(8, 192) \approx 4.58$ s to $\mathcal{T}_{p,wd}(131, 072) \approx 0.34$ s.

3.3 Flow in the respiratory tract

To test the new algorithm, a simulation using 32,768 cores and 4 OpenMP threads per *MPI*-rank is performed on the *JUQUEEN* system, i.e., 2,048 nodes are allocated. The computational mesh has been created by the parallel mesh generator [14] on 4,096 cores in 339.2 s. The coarse-level mesh used for the domain decomposition is on level $l_\alpha = 9$ and the final level is $l_\gamma = 12$ resulting in a total number of cells of $266.5 \cdot 10^6$ and a grid spacing of $\delta_x = 0.11$ mm on l_γ. Note that in Lintermann et al. [14] a mesh $\mathcal{K}$ consisting of $9.82 \cdot 10^9$ could be generated within ≈ 406.8 s. However, the coarse-level mesh of $\mathcal{K}$ is at level $l_{\alpha,\mathcal{K}} = 6$ and the used geometry was a simple cube. That is, serial refinement in the current simulation setup and the inside/outside determination are caused by the large level difference of $l_\Delta = l_\alpha - l_{\alpha,\mathcal{K}} = 3$ and the complex geometry more time consuming. The share of the parallel geometry creation in the mesh generation is with 9.1 s or with 2.7% of the meshing time negligible.
The simulation uses the boundary conditions described in Sec. 2.1 for inspiratory flow and a REYNOLDS number of $Re = (u \cdot D_h)/\nu = 1,515$. Therefore, a bulk velocity magnitude u based on a volume flux of 360 ml/s, the hydraulic diameter D_h of the trachea (see Fig. 6), and the kinematic viscosity of air are applied. To obtain a quasi-steady flow field, approx. $3 \cdot 10^6$ *LBM* iterations have been performed which required around 4 days wall-clock time for the computation on 32,768 cores, i.e., around $3.2 \cdot 10^6$ core-hours. The computations are restarted from checkpoints every 12 hours. All variables introduced in this section are given in non-dimensional *LBM* units.
Fig. 9 shows a cut-plane of the velocity magnitude at inspiration reaching from the left nasal cavity to the first bifurcation in the lung. It is clearly visible that a jet emerges at the epiglottis and continues towards the larynx and below. Further upstream, the flow is laminar while downstream of the larynx the flow becomes transitional and laminarizes again when reaching the first bifurcation of the lung. The left zoom of Fig. 9 shows contours of the Δ-*criterion* [30] given by

$$\Delta_c = \left(\frac{Q}{3}\right)^3 + \left[\frac{\det\left(\nabla \otimes \vec{v}\right)}{2}\right]^2 > 0, \tag{10}$$

which is based on the Q-*criterion* [31]

$$Q = \frac{1}{2}\left(|\Omega|^2 - |S|^2\right) > 0, \tag{11}$$

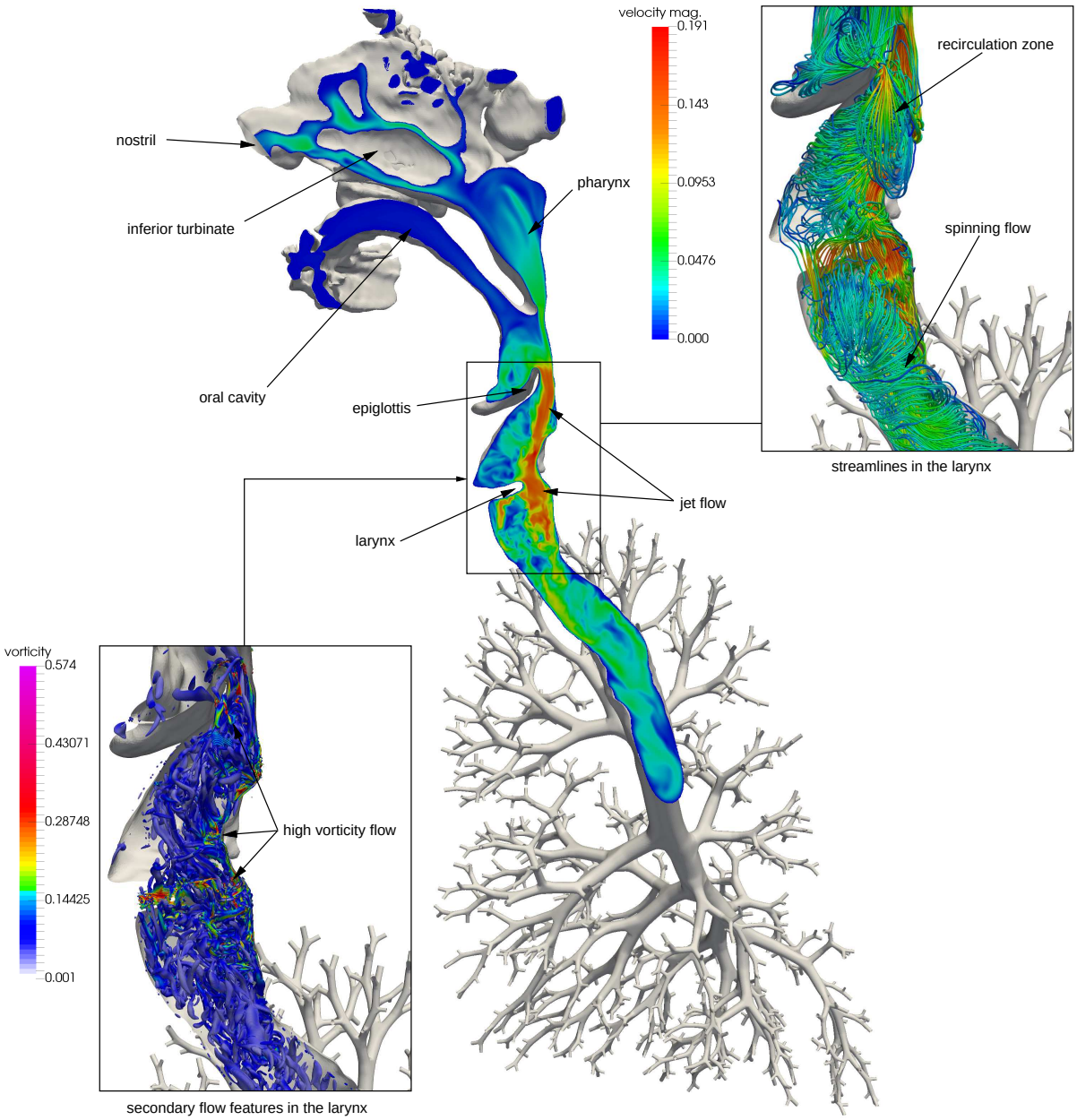


Figure 9: Flow in the human respiratory system at inspiration. In the center, a cut-plane colored by the velocity magnitude is shown. The zooms in the left and right corner show contours of the Δ-*criterion* colored by the vorticity magnitude and streamlines colored by the velocity magnitude.

where $S = \frac{1}{2}\left[\nabla \otimes \vec{v} + (\nabla \otimes \vec{v})^T\right]$ is the strain tensor, $\Omega = \frac{1}{2}\left[\nabla \otimes \vec{v} - (\nabla \otimes \vec{v})^T\right]$ is the vorticity tensor and $|\cdot|$ denotes the *Frobenius norm*. The contours are colored by the vorticity magnitude and evidence the existence of secondary flow structures emerging from the epiglottis. Considering the streamlines shown in the zoom on the right of Fig. 9 a recirculation zone is visible between the epiglottis and the larynx. Downstream of the larynx a spinning wall-bound flow appears that helps to laminarize the flow towards the first bifurcation of the lung.
Note that the simulation of the intrinsic flow in such a complex and highly-resolved geometry is not possible using 16 *MPI*-ranks per *JUQUEEN* node and 4 *OpenMP* threads per task when loading the geometry in serial. In such a case, 4 *MPI*-ranks per node have to be used at maximum and 16 *OpenMP* threads need to be spawned per task. However, in addition to the

waste of memory, *OpenMP* scalability tests on one *JUQUEEN* node have shown that the sweet spot is actually at 16 *MPI*-ranks and 4 *OpenMP* threads and the corresponding computations of the *LBM* are about 1.43 times faster than using 4 *MPI*-ranks and 16 *OpenMP* threads. For the benchmark simulation that means that instead of running 8 batch jobs à 12 hours on the *JUQUEEN* a computation with the serial geometry requires about 11.4 batch jobs à 12 hours to perform the same amount of *LBM* iterations as in the parallel geometry case. Therefore, applying the parallel geometry distribution method not only reduces the allocated memory and increases the preprocessing speed, but also allows for faster simulations.

4 CONCLUSION

In this paper a new and efficient algorithm to generate parallel geometries during the mesh generation has been presented. The algorithm takes advantage of the ordering of coarse-level cells that are output by the parallel grid generator and that determine a domain decomposition of the flow field using a *Hilbert* curve. This leads to a subdivision of the geometry that goes along with the mesh decomposition, i.e., the mesh generator allows for simulations on a quasi-arbitrary amount of cores using distributed geometry elements that are process-local. Instead of loading the geometry on each process in serial and hence consuming a large memory share that could otherwise be used for the simulation, the new method reduces the share to a small fraction. That is, for the simulation of the flow in a respiratory tract on 8,192 cores the total amount of memory allocated by the geometry could be reduced by a factor of 1802.2 from 3.19 TB to 1.8 GB and for 131,072 cores even by a factor of 19,936 from $\approx$ 51 TB to 2.6 GB. Furthermore, the algorithm allowed reducing the preprocessing time by a factor of about 51 from 1045.2 s to 20.5 s for the simulation on 8,192 cores. The test simulations with higher core numbers of up to 131,072 showed that the global memory and the preprocessing time increases only slightly by using the new method. Finally, some results of the flow field computation have been presented to underline the validity of the algorithm and show that the parallel geometry allows using the sweet spot with respect to a combination of distributed and shared memory parallelization. This leads to a 1.43 times faster computation compared to the serial geometry case using the same number of nodes.
Further investigations will cover the performance increase for other solvers using the same meshing technique, i.e., an analysis of the performance of a *Finite-Volume* solver using *level-set* methods [2, 17] for the representation of moving boundaries and a *discontinuous Galerkin* method for the computation of aeroacoustics [32] will be performed.

5 Acknowledgments

The author gratefully acknowledges the computing time granted by the *JARA-HPC* Vergabegremium and provided on the *JARA-HPC* Partition, part of the supercomputer *JUQUEEN* [21] at Forschungszentrum Jülich. Furthermore, the author would like to thank the Institute of Aerodynamics, RWTH Aachen University, for the provision of their code *ZFS*.

REFERENCES

[1] T. Nakashima, M. Tsubokura, M. Vázquez, H. C. Owen, Y. Doi, Coupled Analysis of Unsteady Aerodynamics and Vehicle Motion of a Heavy-Duty Truck in Wind Gusts, Computers & Fluids 80 (2012) 1–9. doi:10.1016/j.compfluid.2012.09.028.

[2] M. Meinke, L. Schneiders, C. Günther, W. Schröder, A cut-cell method for sharp moving boundaries in Cartesian grids, Computers & Fluids 85 (2013) 135–142. doi:10.1016/j.compfluid.2012.11.010.

[3] A. Pogorelov, M. Meinke, W. Schröder, Cut-cell method based large-eddy simulation of tip-leakage flow, Physics of Fluids 27 (7) (2015) 075106. doi:10.1063/1.4926515.

[4] G. J. M. Garcia, N. Bailie, D. a. Martins, J. S. Kimbell, Atrophic rhinitis: a CFD study of air conditioning in the nasal cavity, Journal of applied physiology (Bethesda, Md. : 1985) 103 (3) (2007) 1082–92. doi:10.1152/japplphysiol.01118.2006.

[5] S. K. Kim, Y. Na, J.-I. Kim, S.-K. Chung, Patient specific CFD models of nasal airflow: overview of methods and challenges, Journal of Biomechanics 46 (2) (2013) 299–306. doi:10.1016/j.jbiomech.2012.11.022.

[6] A. Lintermann, M. Meinke, W. Schröder, Fluid mechanics based classification of the respiratory efficiency of several nasal cavities, Computers in Biology and Medicine 43 (11) (2013) 1833–1852. doi:10.1016/j.compbiomed.2013.09.003.

[7] I. Hörschler, M. Meinke, W. Schröder, Numerical simulation of the flow field in a model of the nasal cavity, Computers & Fluids 32 (1) (2003) 39–45. doi:10.1016/S0045-7930(01)00097-4.

[8] M. Konopka, M. Meinke, W. Schröder, Large-eddy simulation of shock-cooling-film interaction at helium and hydrogen injection, Physics of Fluids 25 (10) (2013) 106101. doi:10.1063/1.4823745.

[9] B. Roidl, M. Meinke, W. Schröder, A reformulated synthetic turbulence generation method for a zonal RANS-LES method and its application to zero-pressure gradient boundary layers, International Journal of Heat and Fluid Flow 44 (2013) 28–40. doi:10.1016/j.ijheatfluidflow.2013.03.017.

[10] B. Roidl, M. Meinke, W. Schröder, Boundary layers affected by different pressure gradients investigated computationally by a zonal RANS-LES method, International Journal of Heat and Fluid Flow 45 (2014) 1–13. doi:10.1016/j.ijheatfluidflow.2013.11.004.

[11] V. Statnikov, T. Sayadi, M. Meinke, P. Schmid, W. Schröder, Analysis of pressure perturbation sources on a generic space launcher after-body in supersonic flow using zonal turbulence modeling and dynamic mode decomposition, Physics of Fluids 27 (1) (2015) 016103. doi:10.1063/1.4906219.

[12] K. Nakahashi, L. Kim, Building-Cube Method for Large-Scale High Resolution Flow Computations, in: 42nd AIAA Aerospace Sciences Meeting and Exhibit, no. January, Reno, Nevada, 2004, AIAA 2004–434. doi:10.2514/6.2004-434.

[13] C. Burstedde, L. C. Wilcox, O. Ghattas, p4est : Scalable Algorithms for Parallel Adaptive Mesh Refinement on Forests of Octrees, SIAM Journal on Scientific Computing 33 (3) (2011) 1103–1133. doi:10.1137/100791634.

[14] A. Lintermann, S. Schlimpert, J. Grimmen, C. Günther, M. Meinke, W. Schröder, Massively parallel grid generation on HPC systems, Computer Methods in Applied Mechanics and Engineering 277 (2014) 131–153. doi:10.1016/j.cma.2014.04.009.

[15] C. Günther, M. Meinke, W. Schröder, A flexible level-set approach for tracking multiple interacting interfaces in embedded boundary methods, Computers & Fluids 102 (2014) 182–202. doi:10.1016/j.compfluid.2014.06.023.

[16] R. K. Freitas, W. Schröder, Numerical investigation of the three-dimensional flow in a human lung model, Journal of Biomechanics 41 (11) (2008) 2446–2457. doi:10.1016/j.jbiomech.2008.05.016.

[17] L. Schneiders, C. Günther, J. Grimmen, M. Meinke, W. Schröder, Sharp resolution of complex moving geometries using a multi-cut-cell viscous flow solver, in: 22nd AIAA Computational Fluid Dynamics Conference, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2015, AIAA 2015–3427. doi:10.2514/6.2015-3427.

[18] M. O. Cetin, S. R. Koh, M. H. Meinke, W. Schröder, Aeroacoustic Analysis of a Helicopter Engine Jet Including a Realistic Nozzle Geometry, in: 21st AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2015, AIAA 2015–2533. doi:10.2514/6.2015-2533.

[19] M. Bouzidi, M. Firdaouss, P. Lallemand, Momentum transfer of a Boltzmann-lattice fluid with boundaries, Physics of Fluids 13 (11) (2001) 3452–3459. doi:10.1063/1.1399290.

[20] D. Hartmann, M. Meinke, W. Schröder, A strictly conservative Cartesian cut-cell method for compressible viscous flows on adaptive grids, Computer Methods in Applied Mechanics and Engineering 200 (9-12) (2011) 1038–1052. doi:10.1016/j.cma.2010.05.015.

[21] M. Stephan, J. Docter, JUQUEEN: IBM Blue Gene/Q Supercomputer System at the Jülich Supercomputing Centre, Journal of large-scale research facilities JLSRF 1 (2015) A1. doi:10.17815/jlsrf-1-18.

[22] P. L. Bhatnagar, E. P. Gross, M. Krook, A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems, Physical Review 94 (3) (1954) 511–525. doi:10.1103/PhysRev.94.511.

[23] D. Hänel, Molekulare Gasdynamik, Einführung in die kinetische Theorie der Gase und Lattice-Boltzmann-Methoden, Springer-Verlag, 2004.

[24] Y. H. Qian, D. D'Humières, P. Lallemand, Lattice BGK Models for Navier-Stokes Equation, Europhysics Letters (EPL) 17 (6) (1992) 479–484. doi:10.1209/0295-5075/17/6/001.

[25] J. Li, M. Zingale, W.-k. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, Parallel netCDF: A High-Performance Scientific I/O Interface, in: Proceedings of the 2003 ACM/IEEE conference on Supercomputing - SC '03, ACM Press, New York, New York, USA, 2003, p. 39. doi:10.1145/1048935.1050189.

[26] J. Bonet, J. Peraire, An alternating digital tree (ADT) algorithm for 3D geometric searching and intersection problems, International Journal for Numerical Methods in Engineering 31 (1) (1991) 1–17. doi:10.1002/nme.1620310102.

[27] H. Sagan, Space-Filling Curves, 1st Edition, Springer, 2007.

[28] G. Eitel, R. K. Freitas, A. Lintermann, M. Meinke, W. Schröder, Numerical Simulation of Nasal Cavity Flow Based on a Lattice-Boltzmann Method, in: New Results in Numerical and Experimental Fluid Mechanics VII, Vol. 112 of Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Springer Berlin / Heidelberg, 2010, pp. 513–520.

[29] H. Kitaoka, A 4D Model Generator of the Human Lung, Forma 26 (2011) 19–24.

[30] M. S. Chong, A. E. Perry, B. J. Cantwell, A general classification of three-dimensional flow fields, Physics of Fluids 2 (1990) 765–777.

[31] J. C. R. Hunt, A. A. Wray, P. Moin, Eddies, Stream, and Convergence Zones in Turbulent Flows, Center For Turbulence Research Report CTR.

[32] M. A. Schlottke, H.-J. Cheng, A. Lintermann, M. H. Meinke, W. Schröder, A direct-hybrid method for computational aeroacoustics, in: 21st AIAA/CEAS Aeroacoustics Conference, Dallas, Texas, USA, 2015, AIAA 2015–3133. doi:10.2514/6.2015-3133.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

LARGE-SCALE FREE-SURFACE FLOW SIMULATION USING LATTICE BOLTZMANN METHOD ON MULTI-GPU CLUSTERS

Naoyuki Onodera[1], **Kunihide Ohashi**[2]

[1] National Maritime Research Institute,
6-38-1, Shinkawa, Mitaka-shi, Tokyo, JAPAN
e-mail: onodera@nmri.go.jp

[2] National Maritime Research Institute,
6-38-1, Shinkawa, Mitaka-shi, Tokyo, JAPAN

Keywords: High-performance computing, GPU, Lattice Boltzmann method, Free-surface flow, Large-eddy simulation.

Abstract. *Turbulent free-surface flows around ship strongly affect maneuverability and safety. In order to understand the details of the turbulent flow and surface deformation, it is necessary to carry out high-order accurate and large-scale CFD simulations. We have developed a CFD code based on LBM (Lattice Boltzmann Method) with a single-phase free-surface model. Since violent flows are turbulent with high Reynolds number, a LES (Large-Eddy Simulation) model has to be introduced to solve the LBM equation. The coherent-structure Smagorinsky model is a state-of-the-art sub-grid scale model. Since this model is able to determine the model constant locally, it is suitable for a large-scale calculation containing complicated solid bodies. Our code is written in CUDA and MPI. The GPU kernel function is tuned to achieve high performance on the TSUBAME 2.5 supercomputer at Tokyo Institute of Technology. We obtained good scalability in weak scaling test. Each GPU handles a domain of 192 × 192 × 192, and 27 components are defined at a grid by the D3Q27 model. The fairly high performance of 809 MLUPS(Mega lattice update per second) is achieved by using 1000 GPUs in single precision. By executing this high-performance computation, turbulent violent flow simulation with real ship data is performed, and details of turbulent flows and free-surface deformations will be simulated with much higher accuracy than ever before.*

1 INTRODUCTION

Turbulent free-surface flows around ship strongly affect maneuverability and safety. In order to understand the details of the turbulent flow and surface deformation, it is necessary to carry out high-order accurate and large-scale CFD (Computational Fluid Dynamics) simulations. Lattice Boltzmann method (LBM) is a class of CFD methods to solve a discrete-velocity Boltzmann equation. Since LBM performs local memory access with a simple algorithm, it is suitable for a large-scale parallel computation. As an example of the large-scale calculation using LBM, the researches are nominated for Gordon bell prize of SC, and very high performance were achieved [1-3]. Since violent flows are turbulent with high Reynolds number, a sub-grid scale model has to be introduced to solve the LBM equation. Large-eddy simulation (LES) is a successful approach for unsteady turbulent flows [4]. The dynamic Smagorinsky model is a prominent subgrid-scale model based on a concept of eddy viscosity [5,6]. However, the model parameter of the DSM requires an averaging in global domain. That makes it difficult to perform a large-scale simulation. The coherent-structure model (CSM) is a remedy for these problems [7,8]. Since the model parameter can be locally determined without averaging, it is suitable for parallel computation. In this paper, we apply the CSM to a large-scale turbulent flow simulation.

For the free surface simulation, we applied interface capture method and free-surface model to the LES-LBM equation [9-11]. In this paper, we compared two types of interface capture methods based on level set and VOF. Our code is written in CUDA, and the GPU kernel function is tuned to achieve high performance on the TSUBAME 2.5 supercomputer at Tokyo Institute of Technology. By executing multi-GPU computation, a large-scale simulation with realistic ship data will be performed.

2 LATTICE BOLTZMANN METHOD

LBM solves the discrete Boltzmann equation to simulate the flow of a Newton fluid. Flow field is expressed by a limited number of particles with streaming and collision process. Physical space and time is discretized by a uniform grid. Since fluid particles move onto the neighbor lattice point after 1-time step in streaming process, discrete error of interpolation does not arise. A macroscopic diffusive and a pressure gradient is expressed by the collisional process, and we adopt BGK model. Here, in the time t and the position x, the time evolution of the discretized velocity function is denoted as follows

$$f_i(x + c_i\Delta t, t + \Delta t) = f_i(x,t) - \frac{1}{\tau}(f_i(x,t) - f_i^{eq}(x,t)). \tag{1}$$

Δt is time interval, τ is the relaxation time, and $f_i^{eq}(x, t)$ is the local equilibrium distribution. The local equilibrium distribution is defined as follows

$$f_i^{eq} = w_i\rho\left(1 + \frac{3c_i \cdot u}{c^2} + \frac{9(c_i \cdot u)^2}{2c^4} - \frac{3u^2}{2c^2}\right), \tag{2}$$

ρ is density, and u is velocity. The corresponding weighting factors w_i and components of velocity vector c_i of D3Q27 model are defined as follows

$$w_i = \begin{cases} \frac{8}{27} & i = 0 \\ \frac{2}{27} & i = 1 \sim 6 \\ \frac{1}{54} & i = 7 \sim 18 \\ \frac{1}{216} & i = 19 \sim 26 \end{cases} \tag{3}$$

$$c_i = \begin{cases} (0,0,0) & i = 0 \\ (\pm c,0,0),(0,\pm c,0),(0,0,\pm c) & i = 1 \sim 6 \\ (\pm c,\pm c,0),(\pm c,0,\pm c),(0,\pm c,\pm c) & i = 7 \sim 18 \\ (\pm c,\pm c,\pm c) & i = 19 \sim 26 \end{cases} \tag{4}$$

Relaxation time in collisional process is denoted by the following formula using dynamic viscosity,

$$\tau = \frac{1}{2} + \frac{3\nu}{c^2 \Delta t}. \tag{5}$$

3 LARGE-EDDY SIMULATION

LES resolves dynamics of large-scale structures on a grid scale (GS), and the effects of smaller-scale turbulent structures are taken into account by using a subgrid-scale (SGS) model. SGS models, based on a concept of eddy viscosity, evaluate the effects of turbulence as

$$\nu_{SGS} = C\bar{\triangle}^2 \left|\bar{S}\right|, \tag{6}$$

$$\bar{S}_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right). \tag{7}$$

where C is the model coefficient, Δ is the filter width, and S_{ij} is the velocity strain tensor.

3.1 Dynamic Smagorinsky model

In the Smagorinsky model (SM), the model coefficient C is constant in an entire computational domain, and the SM does not satisfy a correct asymptotic behavior to a wall. The dynamic Smagorinsky model (DSM) improves those defects by calculating the model parameter dynamically [5,6]. The model parameter of the DSM is determined by using two types of grid filters, and the DSM is the most notable breakthrough in LES. However, the model parameter of the DSM requires averaging in global domain for numerical stability. That makes it difficult to perform a large-scale simulation with complex geometries.

3.2 Coherent-structure Smagorinsky model

The coherent structure model (CSM) is a remedy for these problems [7,8]. Since a turbulent structure determines a model parameter locally, the CSM shows good performance in complex geometries. The model coefficient C_{CSM} is calculated by the coherent-structure function F_{CS}, composed of the second invariant of the velocity gradient tensor Q and the magnitude of a velocity gradient tensor E.

$$C_{CSM} = C' \left|F_{CS}\right|^{3/2}, \tag{8}$$

$$F_{CS} = \frac{Q}{E} \quad (-1 \le F_{CS} \le 1), \tag{9}$$

$$Q = -\frac{1}{2}\frac{\partial \bar{u}_j}{\partial x_i}\frac{\partial \bar{u}_i}{\partial x_j} \tag{10}$$

$$E = \frac{1}{2}\left(\frac{\partial \bar{u}_j}{\partial x_i}\right)^2. \tag{11}$$

The coefficient C' is a fixed model parameter, and it is optimized for this simulation. Since the model parameter can be locally determined without averaging, it is suitable for efficient parallel computation.

Introduction of CSM to LBM enables a large-scale simulation with complicated boundary at a high Reynolds number [11]. The total relaxation time with the eddy viscosity is written as

$$\tau_* = \frac{1}{2} + \frac{3\nu_*}{c^2 \Delta t}, \tag{12}$$

$$\nu_* = \nu_0 + \nu_{SGS}. \tag{13}$$

4 FREE-SURFACE MODEL FOR LATTICE BOLTZMANN METHOD

The free surface model [9-11] is implemented with a single-phase model. In a single-phase model, three kinds of cells are defined: liquid cells, gas cells, and interface cells. Assuming the effects of gas are small and negligible, gas cells are not calculated. Liquid cells and interface cells belong to fluid region, and their velocity distribution functions are calculated. The free surface boundary conditions are implemented on the interface cells. For example, the velocity component c_i coming from the empty cell $x+c_i\Delta t$ is reconstructed as

$$f_{-i}(x, t + \Delta t) = f_i^{eq}(\rho_A, u_B) + f_{-i}^{eq}(\rho_f, u_B) - f_i(x, t). \tag{14}$$

$\rho_A=1$ is atmospheric pressure and u_B is velocity of the interface cell. Since velocities in the gas phase are unknown, they are extrapolated from neighbor interface cells.

5 PERFORMANCE ON THE TSUBAME SUPERCOMPUTER

The TSUBAME 2.5 supercomputer in Tokyo Institute of Technology is equipped with more than 4,000 GPUs (NVIDIA K20X), and the theoretical peak performance is 17 PFlops in single precision. The GPU code runs on the TSUBAME by using CUDA programing and MPI library. The parallel GPU-computation is implemented by flat MPI. Since direct data transfers between GPU-to-GPU is not implemented in CUDA 7.0, MPI communications are conducted by using 3-step communications as GPU-CPU (*cudaMemcpyDeviceToHost*), CPU-CPU (*MPI communication*), and CPU-GPU (*cudaMemcpyHostToDevice*) (Fig.1).

We show weak scaling results in Fig. 2. Each GPU handles a domain of 192 × 192 × 192 meshes, and 27 components are defined at a grid by the D3Q27 model. The horizontal axis indicates number of GPUs, and the vertical axis indicates the performance (MLUPS: Mega-lattice update per second) in single precision.

As a result, we achieved good scalability from 8 GPUs to 1000 GPUs; the performance on 8 GPUs is 70 MLUPS (0.809 sec/step), the performance on 125 GPUs is 109 MLUPS (0.815 sec/step), the performance on 512 GPUs is 428 MLUPS (0.846 sec/step), and the performance on 1000 GPUs is 809 MLUPS (0.875 sec/step). The algorithm/code, performed on 1000 GPUs, is 92% more efficient than when performed on 8 GPUs. Since the model parameter of our code can be locally determined, we obtained fairly good performance in the weak scalabilities.

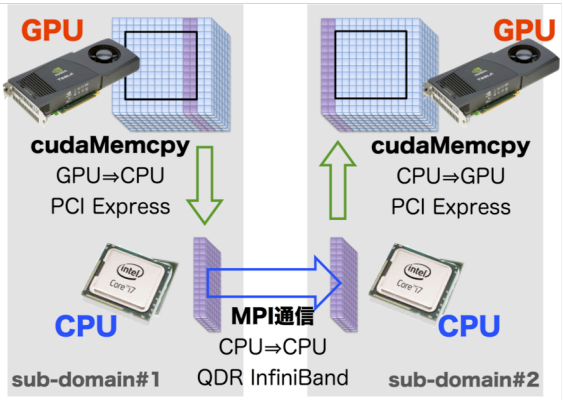


Fig. 1: MPI data transfer between GPU-GPU.

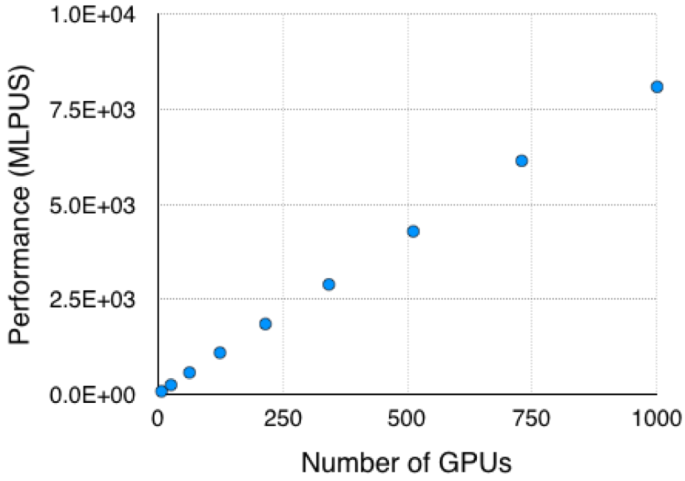


Fig. 2: Weak scalability on TSUBAME 2.5. Each GPU handles a domain of 192 × 192 × 192, and 27 components are defined at a grid by the D3Q27 model.

6 NUMERICAL EXAMPLES

6.1 Two-dimensional dam-break problem

Two-dimensional dam-break problems are considered. Two types of interface capture methods are implemented and the results are compared to experimental results of Martin and Moyce [13]. One interface capture scheme is a level set method and the other scheme is a VOF method. The level set equation is discretized by a 3rd-order WENO-scheme. The VOF equation is discretized by a THINC-WLIC scheme. Figure 3 shows the initial condition with non-dimensional parameters of $a=0.05715\ m(2.25\ inch)$ and $n=1$. Physical parameters used here are liquid density $\rho=1000kg/m^3$, viscosity $\nu=1\times10^{-5}$, and gravity acceleration $g=9.8\ m/s$. The effects of surface tension and wall wettability are not considered.

Figure 4 shows the evolutionary process of the breaking dam. Figures 5 and 6 show the time evolution of top of water column and water front. Results of three different grid resolutions ($N_x=320,512,1024$) are compared to those of experiment. Dimensionless parameters used here are top of water column $H=\eta/(an^2)$, water front $Z=z/a$, and time $\tau=t\,(g/a)^{1/2}$ and $T=nt\,(g/a)^{1/2}$.

The numerical results agree with the experimental results, except at the surge front position with the coarse grid ($N_x=320$) of the level set method. This discrepancy results from the loss of mass at the waterfront. The free surface simulated by the level set method is much smoother than that of VOF.

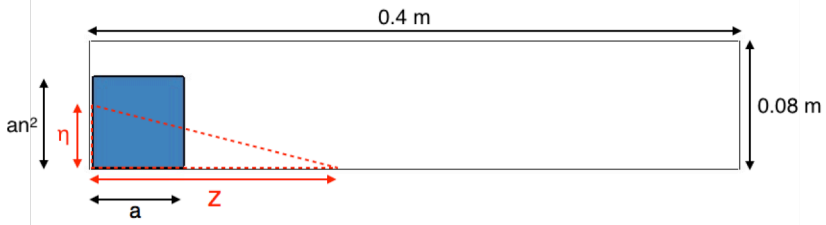


Fig. 3 : Computational condition of breaking dam

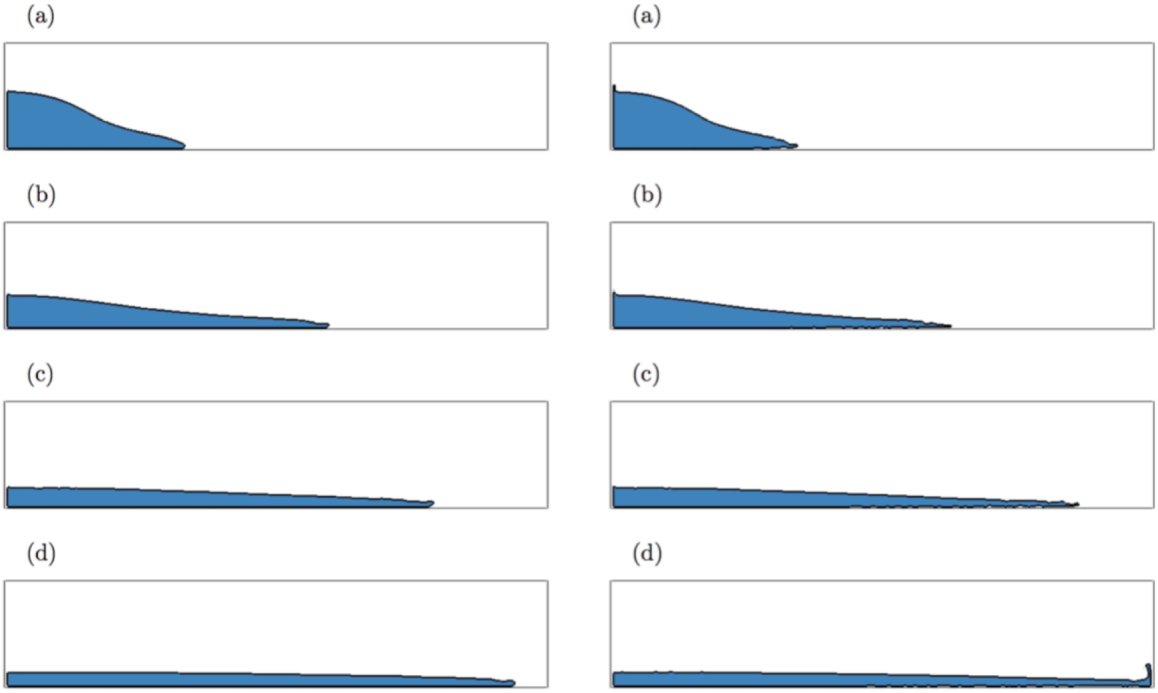


Fig. 4: The process of the 2d-breaking dam at time (a) 0.1, (b) 0.2, (c) 0.3, and (d) 0.4 sec. Left: Level set method; right: THINC-WLIC scheme (512 × 102 cells).

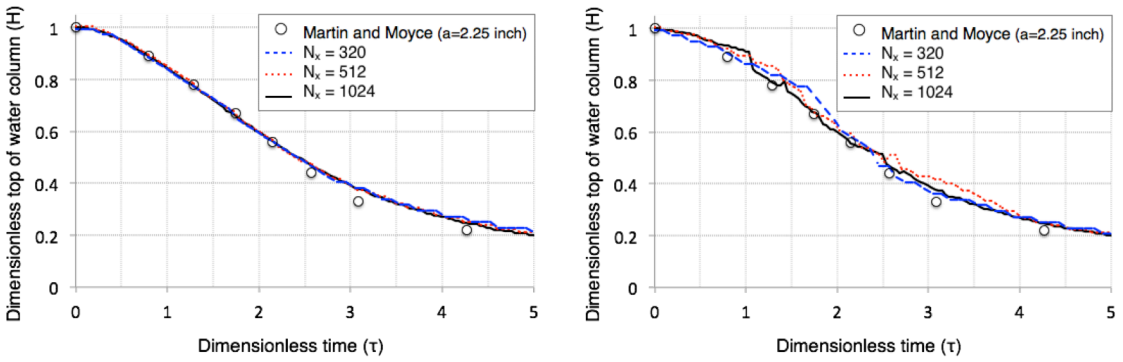


Fig. 5: Comparison of numerical and experimental results for the position of the water column top. Left: level set method; right: THINC-WLIC scheme.

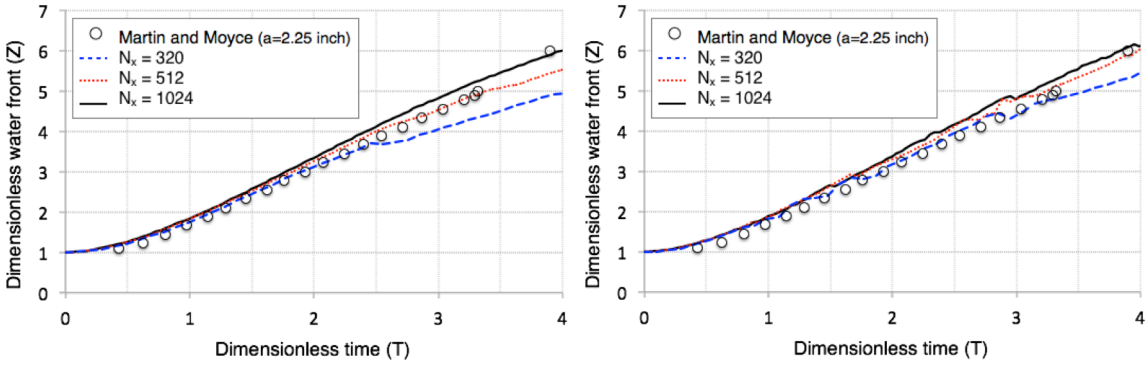


Fig. 6: Comparison of numerical and experimental results for the position of the surge front. Left: level set method; right: THINC-WLIC scheme.

6.2 Three-dimensional dam-break problem

Three-dimensional dam-break problems are considered. A domain size is $0.72\ m \times 0.36\ m \times 0.12\ m$. The width of water column is set to $15\ cm$, and the height of thin water surface is set to $1.8\ cm$. The number of grid points is $(N_1, N_2, N_3) = (1024, 512, 176)$, and the grid resolution is $0.703\ mm$. Figure 7 shows water surface profiles at time step: (a) 0.2, (b) 0.3, (c) 0.4, and (d) 0.5 sec. Although free surface profiles of the level set method and the VOF method are roughly the same, there are large differences at the bubbles in the fluid. Consequently, the level set method does not guarantee the conservation of mass locally. The volume of water by using the level set method increases by $27\ \%$, and volume of the water by using the VOF method decreased by $0.4\ \%$ at $time = 1\ sec$.

6.3 Three-dimensional dam-break problem with a real ship data

A large-scale free-surface problem is considered. We have generated the solid object based on real ship data (Fig. 8 : bulk carrier) and carried out the domain decomposition for multiple-GPU computing. The VOF method (THINC-WLIC) is applied for interface capturing method. We used 80 GPUs for the computation with a $2560 \times 296 \times 736$ mesh. The solid object does not move in this calculation. Figures 9 and 10 show profiles of free-surface at time step (a)6, (b)9, (c)12, (d)15, (e)18, and (f) 21 sec. As shown above, our code can calculate turbulent violent flows with high Reynolds number, and it is concluded that the present scheme is a promising candidate to simulate free-surface flows.

7 CONCLUSION

This paper has presented a large-scale free-surface simulation based on real ship data. The LES-based lattice Boltzmann method is applied to turbulent free-surface flow with a high Reynolds number. Our code is written in CUDA, and the GPU kernel function is tuned to achieve high performance on the TSUBAME 2.5 supercomputer at Tokyo Institute of Technology. We obtained good scalabilities from 8 GPUs to 1000 GPUs. The fairly high performance of 809 MLUPS (Mega lattice update per second) using 1000 GPUs is achieved in single precision. By executing this high-performance computation, turbulent violent flow simulation with a real ship data is considered, and details of the free surface will be simulated with much higher accuracy than ever before.

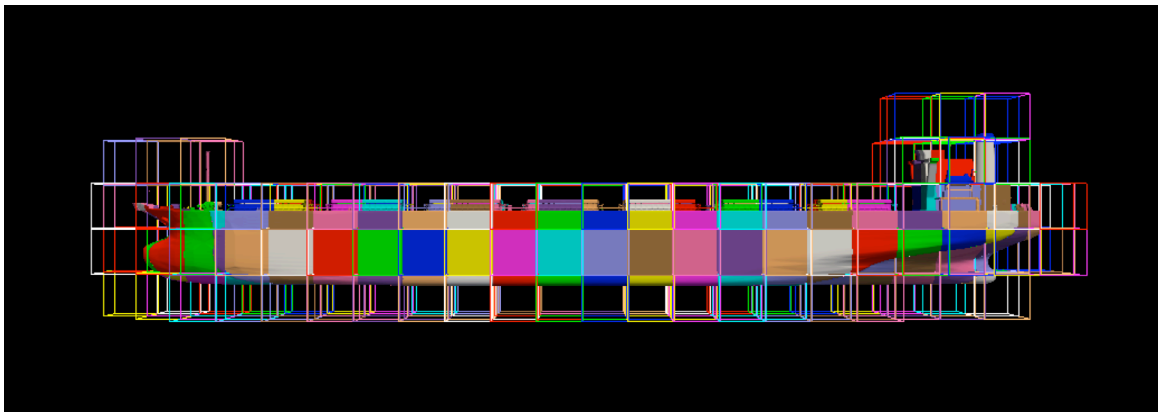


Fig. 8: Level set data of a bulk carrier.

8 ACKNOWLEDGEMENT

This research was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI, a Grant-in-Aid for Young Scientists (B) No. 15K20996.

REFERENCES

[1] Rahimian, I. Lashuk, S. Veerapaneni, A. Chandramowlishwaran, D. Malhotra, L. Moon, R. Sampath, A. Shringarpure, J. Vetter, R. Vuduc, et al. Petascale direct numerical simulation of blood flow on 200k cores and heterogeneous architectures. In Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, pages 1–11. IEEE Computer Society, 2010.

[2] X. Wang and T. Aoki. Multi-gpu performance of incompressible flow computation by lattice boltzmann method on gpu cluster. Parallel Computing, 2011.

[3] M. Bernaschi, M. Fatica, S. Melchionna, S. Succi, and E. Kaxiras. A flexible high-performance lattice boltzmann gpu code for the simulations of fluid flows in complex geometries. Concurrency and Computation: Practice and Experience, 22(1):1–14, 2009.

[4] H. Yu, S.S. Girimaji, and L.S. Luo. Dns and les of decaying isotropic turbulence with and without frame rotation using lattice boltzmann method. Journal of Computational Physics, 209(2):599–616, 2005.

[5] M. Germano, U. Piomelli, P. Moin, and W.H. Cabot. A dynamic subgrid-scale eddy viscosity model. Physics of Fluids A: Fluid Dynamics, 3:1760, 1991.

[6] DK Lilly. A proposed modification of the germano subgrid-scale closure method. Physics of Fluids A: Fluid Dynamics, 4:633, 1992.

[7] H.Kobayashi, F. Ham, and X. Wu. Application of a local sgs model based on coherent structures to complex geometries. International Journal of Heat and Fluid Flow, 29(3), 2008.

[8] H. Kobayashi. Large eddy simulation of magneto hydrodynamic turbulent channel flows with local subgrid-scale model based on coherent structures. Physics of Fluids, 18:045107, 2006.

[9] U Rude and N Thurey. Free surface lattice- boltzmann fluid simulations with and without level sets. In Vision, Modeling, and Visualization 2004: Proceedings, November 16-18, 2004, Standford, USA, page 199. IOS Press, 2004.

[10] Nils Thurey, C Korner, and U Rude. Inter- active free surface fluids with the lattice boltz- mann method. Technical Report05-4. University of Erlangen-Nuremberg, Germany, 2005.

[11] H. Yu, S.S. Girimaji, and L.S. Luo. Dns and les of decaying isotropic turbulence with and without frame rotation using lattice boltzmann method. Journal of Computational Physics, 209(2):599–616, 2005.

[12] T. Shimokawabe, T. Aoki, T. Takaki, T. Endo, A. Yamanaka, N. Maruyama, A. Nukada, and S. Matsuoka. Peta-scale phase-field simulation for dendritic solidification on the tsubame 2.0 supercomputer. In Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, page 3. ACM, 2011.

[13] J.C. Martin and W.J. Moyce, “An experimental study of the collapse of liquid columns on a rigid horizontal plane” Phil. Trans. Royal Soc. of London, Vol. 244, pp.312-324, 1952

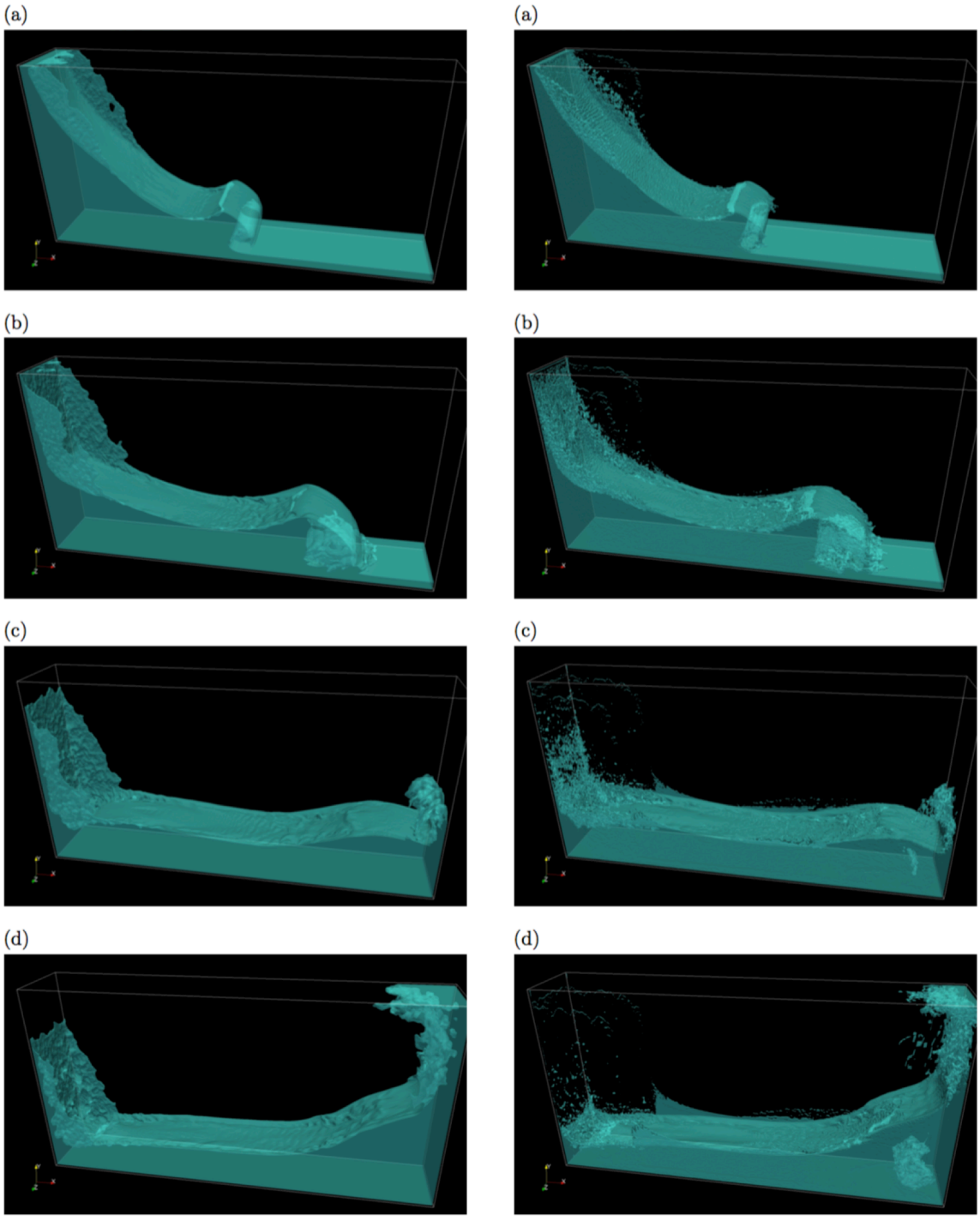


Fig. 7: The water surface profiles of the 3d-breaking dam problem at time step (a) 0.2, (b) 0.3, (c) 0.4, and (d) 0.5 sec. Left: Level set method; right: THINC-WLIC scheme ($1024 \times 512 \times 176$ cells).

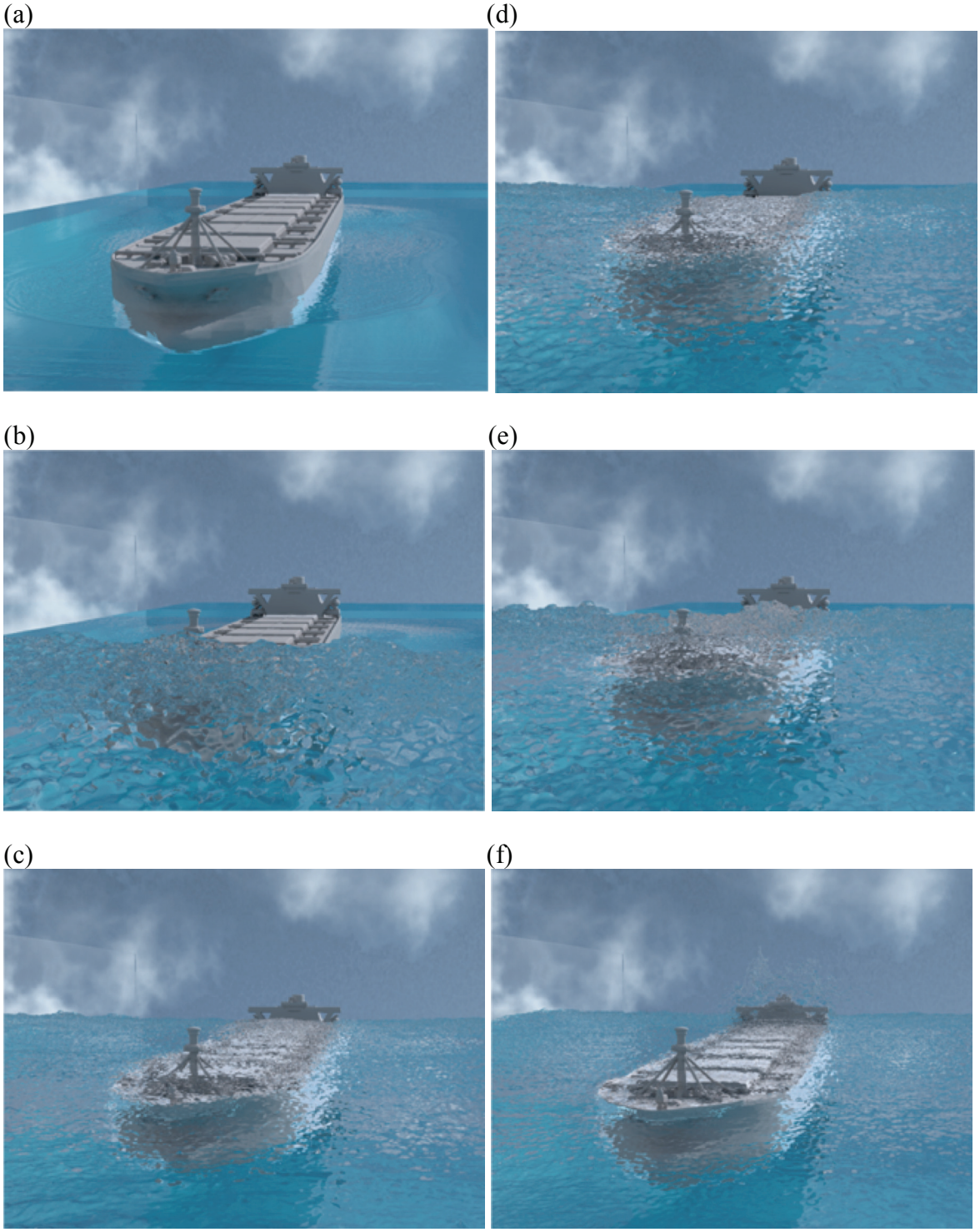


Fig. 9: The evolutionary process of the 3d-breaking dam with a bulk carrier data at time step (a) 6, (b) 9, (c) 12, (d) 15, (e) 18, and (f) 21 sec. (THINC-WLIC scheme : 2560 × 296 × 736 cells).

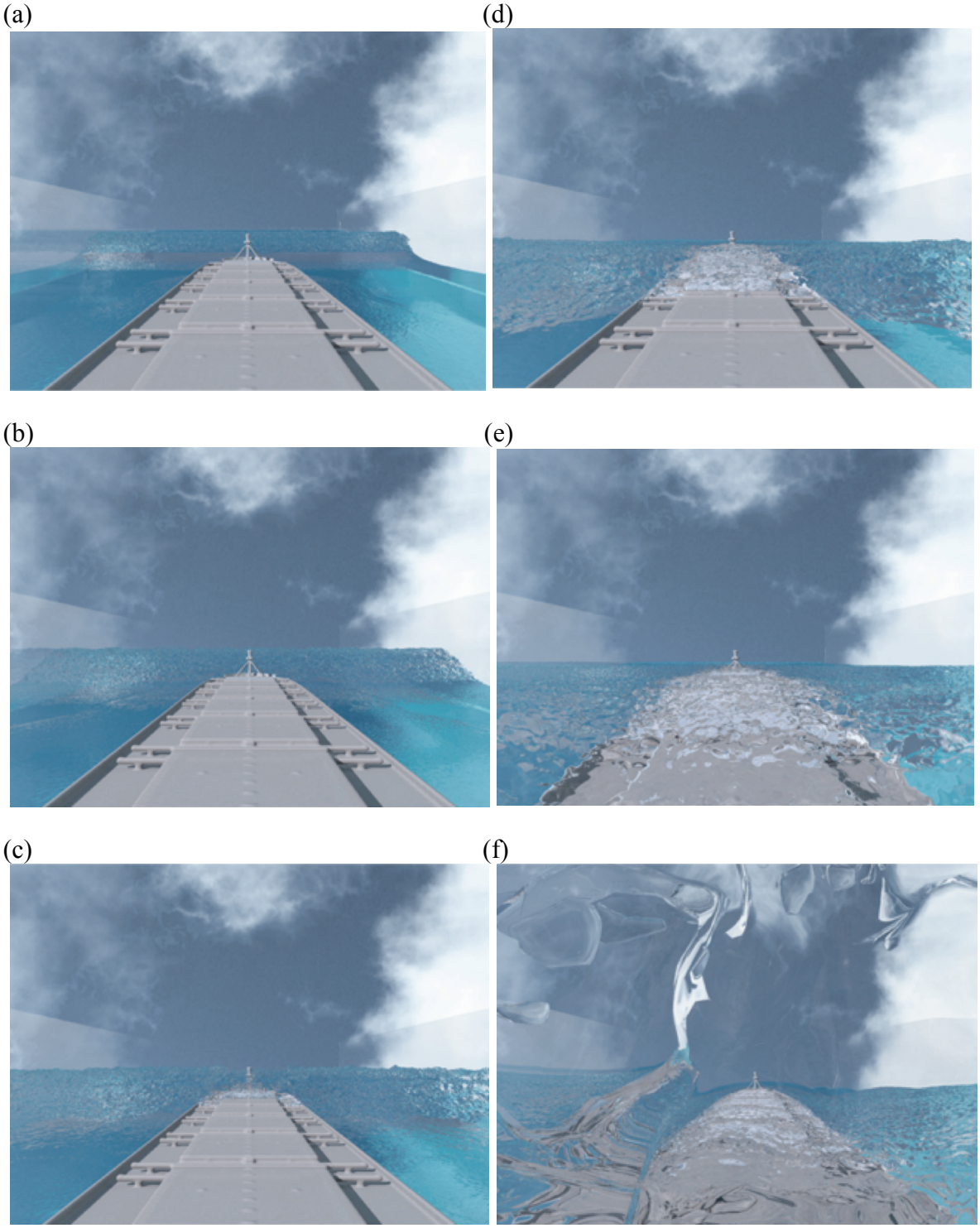


Fig. 10: The evolutionary process of the 3d-breaking dam with a bulk carrier data at time step (a) 6, (b) 9, (c) 12, (d) 15, (e) 18, and (f) 21 sec. (THINC-WLIC scheme : 2560 × 296 × 736 cells).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
Yuji Kobayashi, Kenji Yoshida, Itsuhei Kohri, Masaharu Sakai and Hideo Asano
Crete Island, Greece, 5–10 June 2016

LARGE-SCALE LES ANALYSIS OF AUTOMOTIVE ENGINE COOLING FAN

Yuji Kobayashi[1], Kenji Yoshida[2], Itsuhei Kohri[1], Masaharu Sakai[2], Hideo Asano[2]

[1] Tokyo City University
1-28-1, Tamazutsumi, Setagaya, Tokyo, 158-0087, Japan
e-mail: {yujikoba, ikohri}@tcu.ac.jp

[2] DENSO CORPORATION
1-1 Syouwacho, Kariya, Aichi, 448-8661, Japan
e-mail: {KENJI_K_YOSHIDA, MASAHARU_SAKAI, HIDEO_ASANO}@denso.co.jp

Keywords: Engine cooling fan; Computational Fluid Dynamics (CFD); Large Eddy Simulation (LES)

Abstract. *The purpose of this study is to develop a more efficient engine cooling fan with low noise by understanding the detail of flow structures around the fan. As in the first report, this study focused on the P–Q characteristic and revealed the typical flow structures primary influencing the performance of a ring fan using large-scale LES analysis with HPCS (high-performance computer systems) after experimental validation.*

1 INTRODUCTION

In recent years, there has been a trend to develop a smaller engine cooling fan because of reduced space of the engine compartment due to expansion of the cabin space. Since the P–Q (pressure gain–flow rate) characteristic, which determines the engine cooling performance, varies with the square of the diameter and the number of revolutions of the fan, the performance degradation should be minimized when downsizing. It was shown that a similar cooling performance can be attained by adopting an equal total surface area of the blade using a homothetic design with short cord length and increasing the number of blades under the region of high flow rate (low pressure loss) [1]. However, simply increasing the number of revolutions and/or the number of blades to improve the cooling performance will result in high aerodynamic noise and/or cacophony of the cooling fan. Therefore, the development of a high-performance but low-noise small fan is urgently required. To achieve a significant improvement, it is necessary to understand the relationship between the fan performance and the flow structures around the fan. Many CFD analyses have been conducted to analyse actual phenomena that are difficult to reproduce via experiment [2, 3]. Basically, P–Q characteristic is steady, so it can be calculated using steady state CFD simulation like what other authors conducted [4, 5]. Moreover, it is also important to clarify the relation between the fan performance and the flow structures in each frequency when we consider the aerodynamic noise to

attain a major improvement. However, a successful research concerning this has not been reported yet.

Considering the above background, this research aims to clarify the mechanism of the generated aerodynamic noise of the cooling fan by associating the flow structures with the source point of the turbulence and the Blade Passing Frequency (BPF) noise. As a first step of the research, the detail of the flow structures around the fan was determined by large eddy simulation (LES) analysis using high-performance computing systems (HPCS) of K computer and FX-10 super computing systems. In particular, three kinds of grid and time scale, i.e., small-, middle-, and large-scale LES analyses were conducted to reveal the dominant flow structures on the performance and the essential information for the accurate prediction after experimental validations.

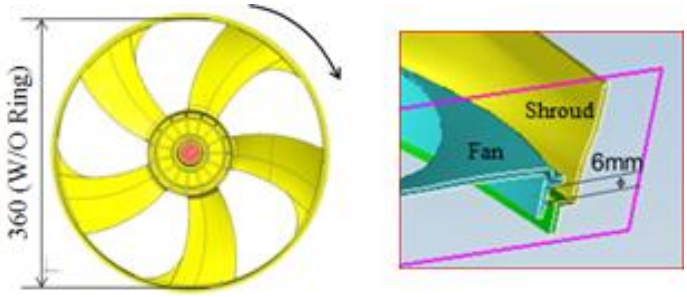


Figure 1: Test piece used in present study

2 METHODOLOGY

2.1 Experimental method

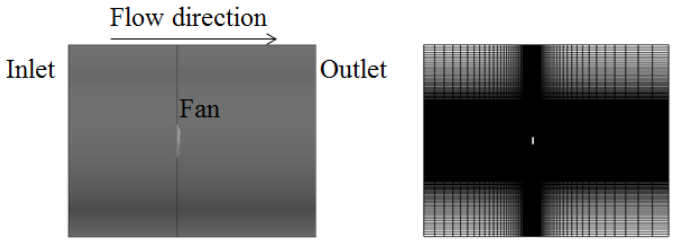
This study used a ring fan as a test piece as shown in Fig.1. The fan has five even-pitched blades with a ring at the tip of the blade. It is a real scale model of an actual engine cooling fan, which has a diameter of 360 mm. The performance of the fan was examined by measuring the flow rate and pressure gain at 2080 rpm in a test chamber. A shroud was only installed as peripheral equipment, and the clearance between outer wall of the ring and inner wall of the shroud was set at 6 mm. The experimental uncertainty of the measurement of the performance of the fan has been calculated to be within ±1%. In this study, the Reynolds number *Re* was approximately 9.4×10^5, based on the tip velocity (39.2 m/s) and diameter of the fan (360 mm).

In addition, the spectrum of the velocity magnitude measured by I type hot-wire 10 mm behind the trailing edge of the fan was calculated using fast Fourier transform (FFT) analysis for the experimental validation of the unsteady amount

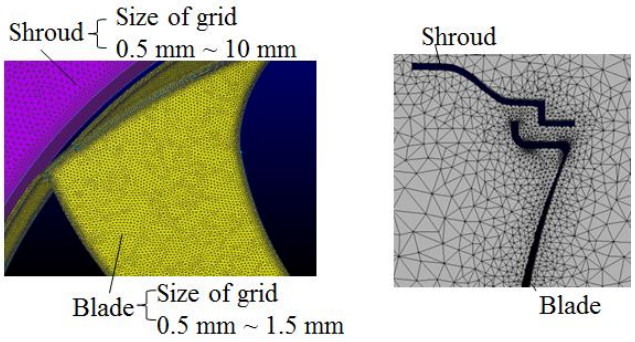
2.2 CFD setup

Fig.2 shows the computational model and grid used in this study. Three types of computational models with difference in grid and time resolution were used. For small-scale analysis, tetrahedral unstructured grids were adopted in the entire computational domain, and the minimum size of the grid at the blade surface was 0.5 mm. Thus, the total number of elements was approximately 14 million for this model. On the other hand, for middle-scale analysis, hexahedral grids were adopted in the entire computational domain, and the minimum size of the grid at the blade surface was 0.1 mm. Hence, the total number of elements was approximately 86 million. For large-scale analysis, the grid resolution was doubled in each coordinate

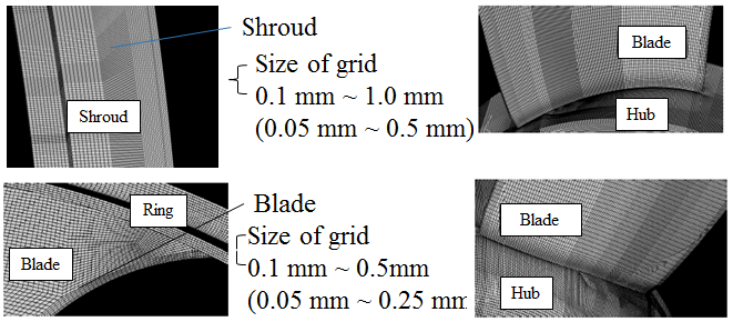
compared to the middle-scale model. Consequently, the total number of elements was approximately 690 million. The finite element method was adopted to solve the unsteady LES using low-Mach-number approximation. The dynamic Smagorinsky model (DSM) for the sub-grid-scale (SGS) model and the Crank–Nicolson method for time discretization were used. The overset method [6] was used for the space around the rotating blade. The calculation code for the FrontFlow/blue-Ver.8.1 was used in this study. Time step size was set to 7.67×10^{-6} s for the small- and middle-scale analyses and 3.83×10^{-6} s for the large-scale analysis to ensure that the Courant number is less than one in all elements. Table 1 summarizes the main features of the computational method and simulation parameters.



(a) Computational Domain and grid at section of the centre plane



(b) Computational grid used for small-scale analysis



(c) Computational grid used for middle- and large-scale analyses; w/o brackets: middle-scale, with brackets: large-scale

Figure 2: Computational model and grid used in this study

	Small-scale	Middle-scale	Large-scale
Governing Equation	NS Equations		
	Continuity Equation		
Fluid	Incompressible		
State	Unsteady		
Turbulence Model	LES (DSM)		
Treatment of Rotation	Overset method		
Time discretization	Crank-Nicolson		
Cell Structure	All Tetrahedral	All Hexahedral	
Total Number of Cells	14.0 Million	86.0 Million	690 Million
Time Step [s]	7.67×10^{-6}		3.83×10^{-6}

Table 1: Specifications for fan calculations

3 VALIDATION OF COMPUTATIONAL ACCURACY

The fan performance was evaluated using the $\varphi - \psi^*$ characteristics, where φ and ψ^* are the coefficients of the flow rate and pressure gain, respectively, as defined by the following formulas:

$$\varphi = \frac{Q}{u_t \cdot A} \tag{1}$$

$$\psi^* = \frac{2\Delta P}{\rho \cdot u_t^{\,2}} \tag{2}$$

where Q, A, u_t, ρ, and ΔP denote the flow rate, frontal area of the rotating region of the fan, tip velocity, air density, and pressure difference between the inside and the outside of the chamber, respectively.

Fig.3 shows the $\varphi - \psi^*$ performance obtained by calculation compared with the experimental results. Here, $\varphi = 0.1$ means the working condition and this study focused only on that condition. The numbers in Fig.3 indicate the ratio of the CFD result to the experimental result. As shown, the proportion is 82.5% for small-scale and 121.0% for middle-scale; therefore, the predictive accuracy is not good. On the other hand, large-scale result shows good agreement with the experimental value as the proportion is 99.0%. Finally, Fig.4 shows the velocity spectrum of the wake obtained by middle-scale calculation compared with the experimental result at the point 10 mm behind the trailing edge and for 170 mm radius of the fan. The result of the simulation also shows good agreement with the experimental value in the qualitative manner.

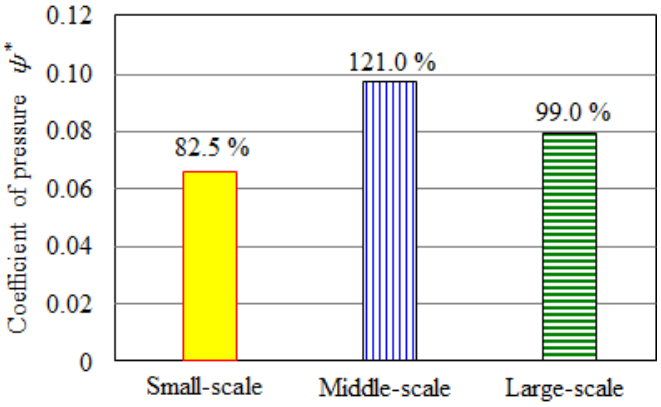


Figure 3: $\psi*$ characteristic calculated at the condition in $\varphi = 0.1$

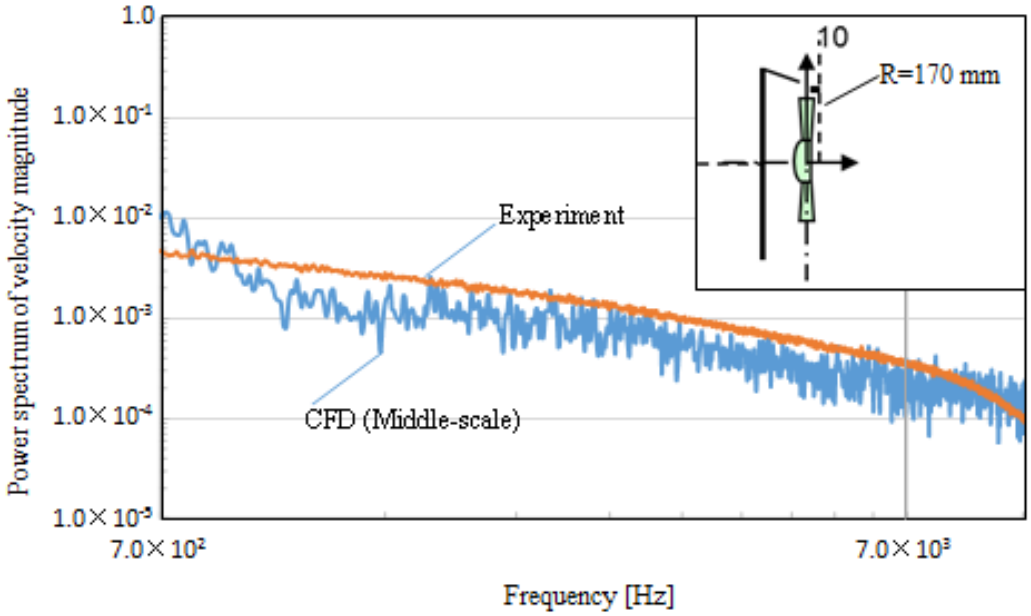


Figure 4: Power spectrum of velocity magnitude of wake flow

4 RELATION BETWEEN PERFORMANCE AND FLOW STRUCTURES

In this section, the essential point to the predictive accuracy is determined by comparing the flow structures of three kinds of simulations. Then, the relationship between the flow structures and the fan performance is discussed. Fig.5 shows the coefficient of pressure C_p of root mean square (RMS) value distribution, which denotes the time variation, while Fig.6 shows C_p distribution on the blade surface as defined by the following formula:

$$C_p = 2p/\left(\rho \cdot u_t^{\,2}\right) \tag{3}$$

Here, p denotes the local static pressure. It was found that the pressure fluctuation occurred locally at the suction surface of the leading edge and/or at the pressure surface of the leading edge near the tip for middle- and large-scale analyses. In contrast, the distribution feature was different particularly on the suction surface for small-scale analysis. This behaviour was also

confirmed by C_p distribution in perspective and the detail information was verified by C_p profile of concentric section of the blade surface at radius R = 110 and 150 mm as shown in Fig.7. Firstly, by comparing small-scale against middle- and large-scale, pressure recovery on the suction surface indicated different features. In addition, from the C_p profile of R = 150 mm at the leading edge, the prediction with small-scale of the attack angle of flow shows a large difference compared with the others. Next, by comparing middle- against large-scale, the difference of C_p was confined to the immediate vicinity of leading edge. This means that the difference of the $\varphi-\psi^*$ characteristic between middle- and large-scale was primarily caused by the flow structures between the blades (spatial flow structures). To confirm this theory, spatial vortex structures represented by isosurface of velocity gradient tensor of the second invariance Q were visualized as shown in Fig.8. As a result, for middle- and large-scale, it was observed that the turbulent vortex structures were generated from the leading and trailing edges. On the other hand, the vortex structures for small-scale were generated only at the leading edge, which corresponds to a separation vortex. It should be noted that the grid resolution was too poor to calculate the flow in boundary layer and/or turbulent wake structures. To clarify the influence of these turbulent structures on the spatial flow field around the fan, Q distribution and distribution of coefficient of velocity magnitude C_v normalized by tip velocity u_t at the section of R = 110 and 150 mm are shown in Fig.9 and Fig.10. For middle-scale analysis, it was found that the attack angle of flow was higher than large-scale because a larger turbulent structure was generated at the leading edge. In contrast, turbulent structures near the blade surface and wake for middle-scale were smaller than large-scale. Consequently, it should be noted that the attack angle of flow varied due to the difference of wake structures and finally changed the flow structures between the blades. This means that the essential point for the accurate prediction is the attack angle of flow in case of a ring fan because there is no large tip vortex that is always generated in case of a propeller fan. In other words, the flow in boundary layer and wake structures, which determines the flow field between the blades, dominates the performance of a ring fan. On the other hand, in a propeller fan used in previous studies [4], the essential point for the accurate prediction is the resolution of the spatial tip vortex because the vortex dominates the fan performance.

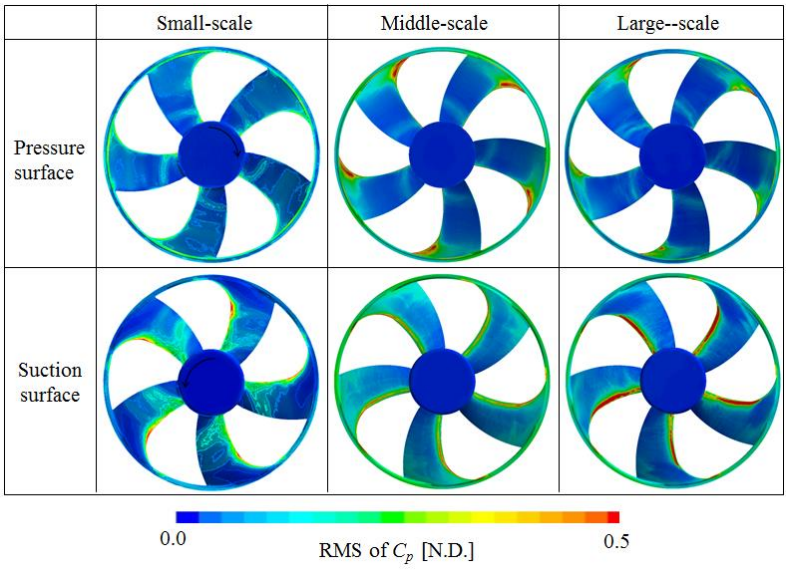


Figure 5: RMS distribution of C_p on the blade surface

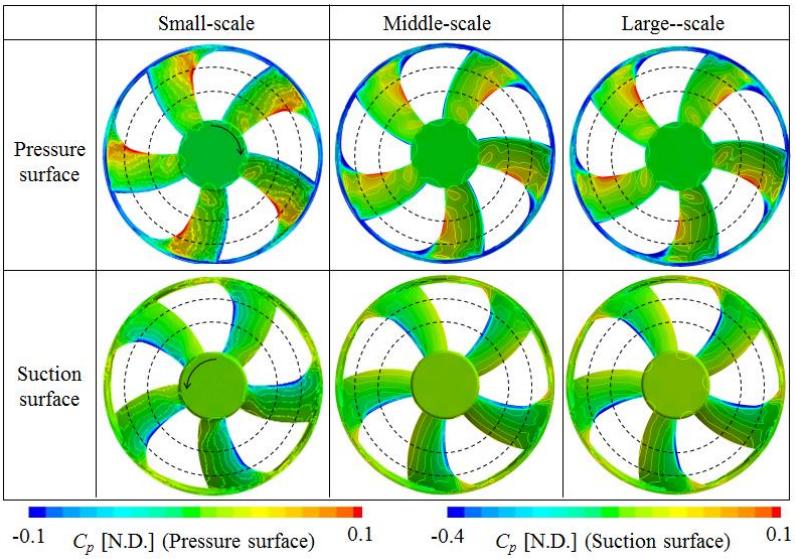
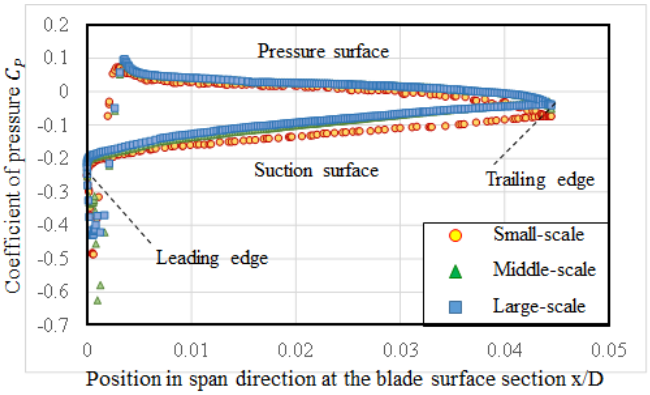
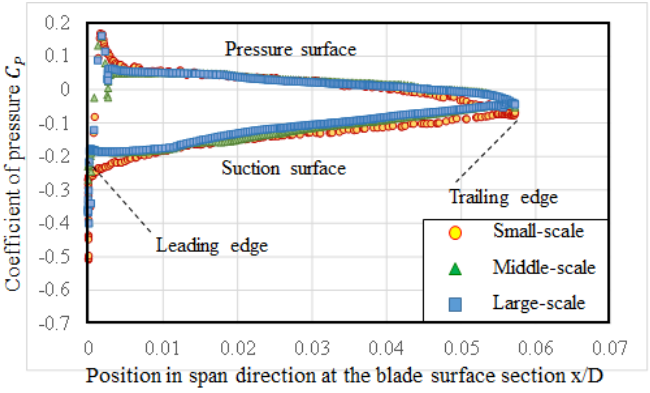


Figure 6: C_p distribution on the blade surface



(d) C_p profile at R = 110 mm section on the blade surface



(e) C_p profile at R = 150 mm section on the blade surface

Figure 7: C_p profile at R = 110 and 150 mm on the blade surface

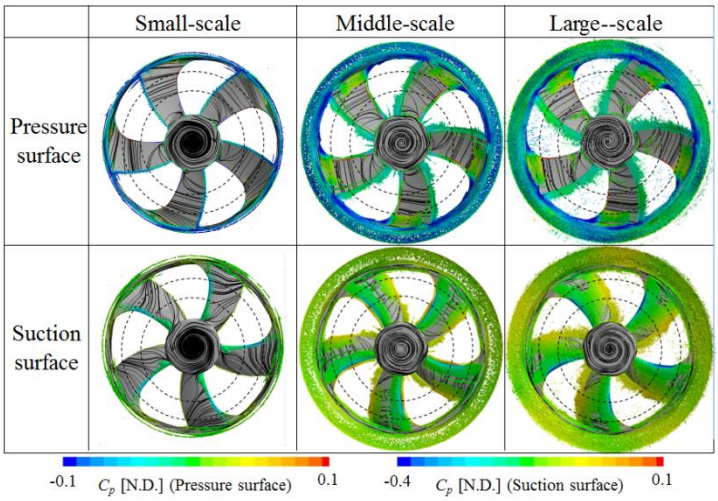


Figure 8: Isosurface of second invariance Q and streamline on the fan

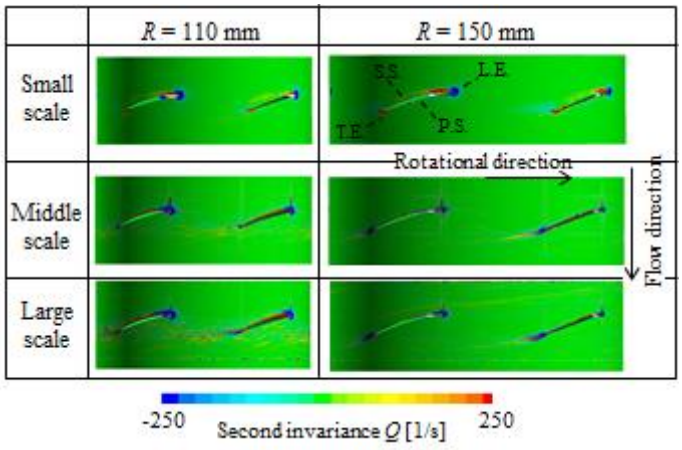


Figure 9: Q distribution at section of R = 110 and 150 mm section

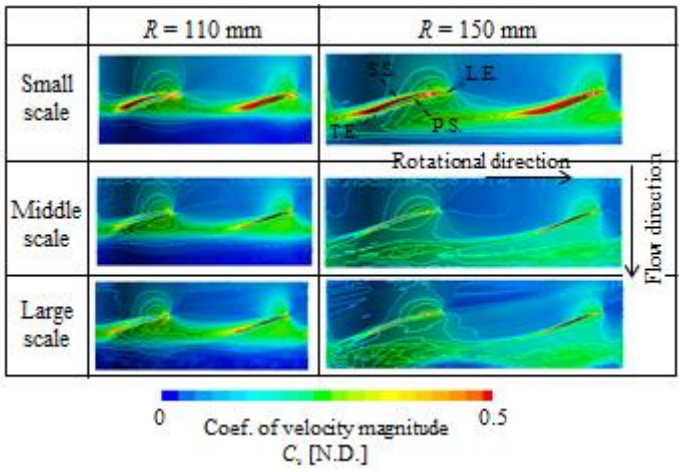


Figure 10: C_v distribution at section of R = 110 and 150 mm section

5 IMPORTANCE AND PROBLEM OF LARGE-SCALE ANALYSIS

We think that the clarification of the relationship between the fan performance and flow structures at each frequency is necessary to develop a high-performance and low-noise small cooling fan as mentioned above. For the cooling fan development, the study focused on very high-frequency region with such turbulent noise in the order of 10000 Hz and/or low-frequency region with the BPF noise for major improvement. It is difficult to study in these regions using conventional computing systems; therefore, HPCS is required. In addition, postprocessing of big data obtained from large-scale analysis remains a critical issue. In future work, we will focus on proper orthogonal decomposition (POD) [7, 8] for the analytical method of big data and classify the flow structures in each frequency to associate them with the fan performance.

6 CONCLUSIONS

The purpose of this study is to reveal the dominant flow structures on the fan performance and the essential information for the accurate prediction by conducting small-, medium-, and large-scale LES analyses using HPCS. For the case of the ring fan (such as that used in this study), it was found that accurate prediction is determined by the calculation result of the attack angle of flow because there is no large tip vortex that is absolutely generated in the case of a propeller fan. This means that the flow in boundary layer and wake structures dominate the fan performance. It should be noted that the resolution of longitudinal vortex, similar to the separation vortex, is most important when we predict the performance under the region of low flow rate conditions (region with high pressure loss condition).

REFERENCES

[1] Yuji Kobayashi et al., Influence of the Fundamental Parameters for the Blade Geometry on the Fan Performance, *Transactions of the Japan Society of Automotive Engineers, Vol.44, No.6, pp.1477-1482, November,* 2013.

[2] Kazuya Kusano et al., Three-dimensional Structures of Tip Vortex in a Half-ducted Propeller Fan, *Transactions of the Japan Society of Mechanical Engineers, Vol.80, No.810,* 2014.

[3] Tomoyoshi Sasajima et al., Numerical Analysis of Flow around Blades in Axial Flow Small Fan, *Transactions of the Japan Society of Mechanical Engineers, Series B, Vol.77, No.774, pp.255-263,* 2011.

[4] Itsuhei Kohri et al., Prediction of the Performance of the Engine Cooling Fan with CFD Simulation，*SAE International Journal Of Passenger Cars - Mechanical Systems, Vol. 7, No. 2, pp.728-738, August,* 2010.

[5] Yuji Kobayashi et al., Determination of Suitable MRF-Region for Practical Prediction of Engine Cooling Fan Performance, *AJK2011-FED, AJK2011-23005, July,* 2011.

[6] Chisachi Kato et al., Large Eddy Simulation of Unsteady Flow in a Mixed-Flow Pump : 1st Report, Numerical Method, *Transactions of the Japan Society of Mechanical Engineers, Series B, Vol.68, No.680, pp. 1729-1736,* 2002.

[7] Kunihiko Taira, Proper Orthogonal Decomposition in Fluid Flow Analysis: 1. Introduction, *Journal of Japan Society of Fluid Mechanics, Vol.30, pp.115-123,* 2011.

[8] Yuji Kobayashi et al., Proper Orthogonal Decomposition Analysis of Flow Structures Generated around Engine Cooling Fan, *SAE International Journal Of Passenger Cars - Mechanical Systems, Vol.3, Issue 1, pp.508-522, November,* 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DIRECT AEROACOUSTIC SIMULATION RELATED WITH MODE CHANGE IN A RECORDER

H. Yokoyama[1], R. Hamasuna[1], A. Miki[2], H. Onitsuka[2], and A. Iida[1]

[1] Department of Mechanical Engineering, Toyohashi University of Technology, Aichi 441-8580, Japan
e-mail: h-yokoyama@me.tut.ac.jp

[2] Research & Development Division, Yamaha Corporation, Shizuoka 438-0192, Japan

Keywords: Direct Aeroacoustic Simulation, Recorder, Musical Instruments, Mode Change

Abstract. *In musical instruments, the mode change of radiating sound is related with their quality. To clarify the flow and acoustic fields related with this mode change, we performed direct aeroacoustic simulations of flow and acoustic fields around two actual recorders with different shapes of windway. The computations are based on the compressible Navier-Stokes equations to predict the fluid-acoustic interactions. It is shown that the relationship between the mean velocity at the windway exit and the mode of the radiating sound for the two models is in good agreement with the measured results. The velocity of the mode change depends on the shape and configurations of the windway and edge. Based on the predicted flow fields, the asymmetry of the jet oscillations between in the upward and downward directions is observed in the recorder model for the earlier mode change.*

1 INTRODUCTION

Flue instruments, such as flutes, bamboo flutes, and recorders, consist of an edge ("labium") with which an air jet collides and a resonator (pipe). In the flue instruments, the fluid-acoustic interactions occur [1]. Since the pioneering works of Helmholtz [2] and Rayleigh [3], many researchers have investigated the self-sustained oscillations of flow and sound in flue instruments. Giordano [4] simulated the effects of the chamfers of the windway exit and clarified the position of the edge (with reference to the windway) on the tonal sound. However, the model was made by extruding a two-dimensional cross-section.

In the previous study [5], the authors conducted the aeroacoustic direct simulations for actual recorders with tone holes. The predicted flow and acoustic fields are in good agreement with the measured results by Particle Image Velocimetry (PIV) and a microphone. Those results also clarified the effects of the condition of tone holes on the sound and flow fields and the mechanism by which self-sustained oscillations (including fluid-acoustic interactions) are maintained.

Also, as the jet velocity increases, the mode of the tone jumps to the higher mode [4, 6]. This mode change of radiating sound is related to their quality. However, to the best knowledge of the authors, the effects of the shape and positions of edge and windway on that mode change have not been clarified sufficiently for actual recorders.

In this paper, the aeroacoustic direct simulations are conducted for two different types of Types S (YRA-28BIII : Straight windway) and A (YRA-38BIII : Arch windway) of recorders as shown in Figure 1. The effects of the shape on the flow and sound fields are discussed focusing on the mode change.

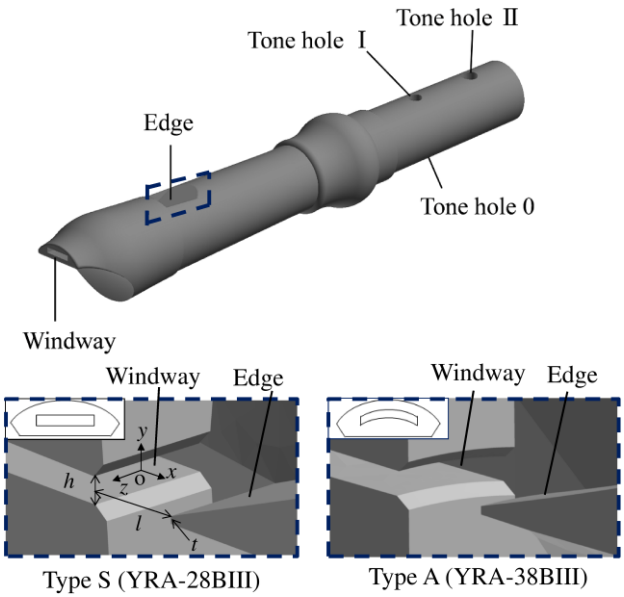


Figure 1: Geometries of recorders.

2 COMPUTATIONAL METHODOLOGIES

2.1 Flow conditions

The flow and acoustic fields around a recorder were investigated by using a short recorder with three tone holes (see Figure 1), namely, actual recorders (Types S and A) cut short after the third hole, as a model. The resonator is conical except in the region near the edge, where

the lower flat surface of the edge is smoothly connected with the circular surface into the resonator. In this paper, the simulations with all tone holes open are discussed.

The origin of the coordinate system of the full model is located at the center of the outlet of the windway. The streamwise direction in the windway is the *x*-axis, the vertical direction is the *y*-axis, and the spanwise direction intersecting with those two axes is the *z*-axis.

Type A and Type S have some different features as shown in Table 1. One difference is the shape of windway exit. The shape is straight and relatively wider (the height h/t = 3.7, where t is the height of the edge and 0.35 mm for both models) for Type S while that is slightly curved (the edge is also curved) and relatively narrower (h/t = 3.0) for Type A. Another difference is the configurations of the windway exit and edge. The distance between the windway exit and the edge, l, (distance of window) for Type S is relatively longer (l/t = 15.0) and the position of the edge with reference to the center of the windway exit, y_e , is relatively higher (y_e/t = -0.29) for Type S, while those are l/t = 13.9 and y_e/t = -0.43 for Type A.

	Windway height h/t	*Distance of window* l/t, l/h	*Edge position* y_e/t, y_e/h
Type S	3.7	15.0, 4.1	-0.29, -0.078
Type A	3.0	13.9, 4.6	-0.46, -0.151

Table 1: Parameters of recorders of Type S and Type A.

The cross-sectional average blow-jet velocity (U_0) is based on the area at the starting point of the chamfers around the exit of the windway. The effects of jet velocity (ranging from 8.0 to 65 m/s) were computationally investigated in the case that all tone holes were open.

2.2 Governing equations and finite-difference formulation

The governing equations used for simulating the interactions between flow and acoustic fields are based on the three-dimensional compressible Navier-Stokes equations. Spatial derivatives are solved by using the sixth-order-accuracy compact finite-difference scheme (fourth-order accuracy at boundaries) [7]. Time integration is performed using the third-order-accuracy Runge-Kutta method.

A volume-penalization (VP) method [8, 9], which is a kind of immersed-boundary method [10], is used to reproduce the flow and acoustic fields around a complex shape of the recorder on a rectangular grid. The penalization term, $\boldsymbol{V}$, is added to right-hand side of the governing equations of the three-dimensional compressible Navier-Stokes equations as follows:

$$\boldsymbol{Q}_t + (\boldsymbol{E} - \boldsymbol{E}_v)_x + (\boldsymbol{F} - \boldsymbol{F}_v)_y + (\boldsymbol{G} - \boldsymbol{G}_v)_z = \boldsymbol{V} \tag{1}$$

$$\boldsymbol{V} = -(1/\phi - 1)\chi \begin{pmatrix} \partial \rho u_i / \partial x_i \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \phi = 0.25, \tag{2}$$

where $\boldsymbol{Q}_t$ is the vector of the conservative variables, $\boldsymbol{E}$, $\boldsymbol{F}$, and $\boldsymbol{G}$ are inviscid flux vectors, $\boldsymbol{E}_v$ $\boldsymbol{F}_v$, and $\boldsymbol{G}_v$ are viscous flux vectors, and ϕ is the porosity of a porous medium, determined so that the sound wave can be reflected almost completely (reflectivity : 99%). The mask function χ is given as

$$\chi = \begin{cases} 1\,(\text{inside object}) \\ 0\,(\text{outside object}) \end{cases}. \tag{3}$$

To reduce computational cost, large-eddy simulations (LES) were performed for the reproduction of the flow fields in the recorder. In the simulations, no explicit subgrid-scale (SGS) model was used. The turbulent energy in the grid-scale (GS) that should be transferred to SGS eddies is dissipated by a 10th-order spatial filter as described below.

$$\alpha_{\mathrm{f}}\hat{\psi}_{i-1} + \hat{\psi}_i + \alpha_{\mathrm{f}}\hat{\psi}_{i+1} = \sum_{n=0}^{5} \frac{a_n}{2}\left(\psi_{i+n} + \psi_{i-n}\right), \tag{4}$$

where ψ is a conservative quantity and $\hat{\psi}$ is the filtered quantity, also removes numerical instabilities [11]. Coefficient a_n has the same value as that used by Gaitonde and Visbal [12], and the value of α_{f} is 0.45.

The predicted mean and instantaneous flow fields and acoustic fields based on the above-mentioned methodologies have been confirmed to agree with those measured by PIV and a microphone in the previous paper [5].

2.3 Computational domain

The computational domain (shown in Figure 2) is divided into three regions: a vortex region, a sound region, and a buffer region having different grid spacings, while the grid spacing is smoothly changed to prevent spurious acoustic reflections near the boundary of different regions. The computational grid is rectangular (as mentioned above). The total number of grid points is approximately 8.2×10^7. The computational grid is explained in detail in the previous paper [5].

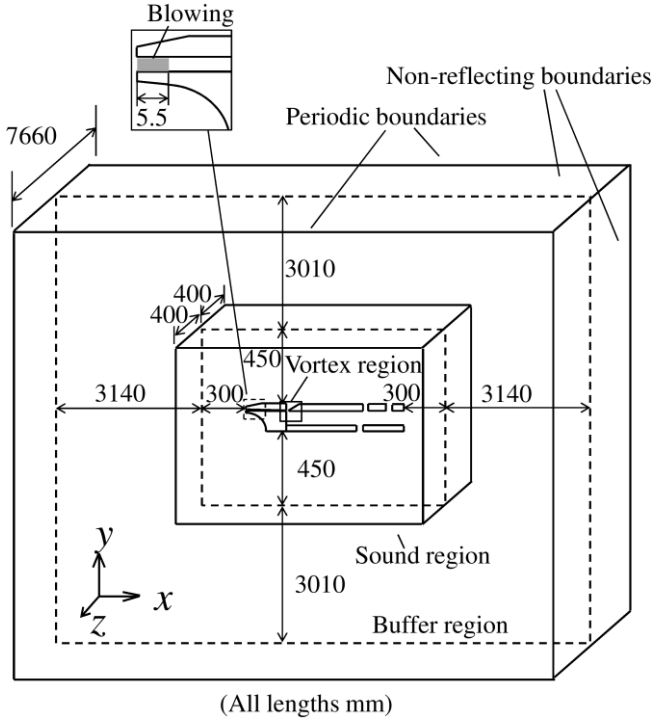


Figure 2: Computational domain and boundary conditions.

2.4 Boundary conditions and initial conditions

The boundary conditions of the computation are shown in Figure 2. Non-reflecting boundaries [13-15] are used at boundaries of the x and y directions, and periodic boundary conditions are adopted in the z direction. To reproduce the jet in the windway, the velocity is set to be a given blowing velocity in the inlet region of the windway, as shown in gray color in Figure 2. For the objects such as a recorder, the above-mentioned mask function is set to $\chi = 1$, and the velocity is set to zero.

In the computation, the velocity is set to zero in the initial field and the jet velocity is accelerated in the duration of computational development.

3 VALIDATION OF COMPUTATIONAL METHODS ABOUT MODE CHANGE

The predicted and measured effects of the jet velocity on the frequency of the dominant tone (fundamental frequency) for Types S and A are shown in Figure 3. As shown in the figure, the predicted frequencies are in good agreement with those measured. As the velocity becomes higher, the frequency of the tonal sound jumps to the second mode. The mode change occurs at a higher velocity for Type A with reference to that for Type S. This difference is also predicted in this simulation as shown in the figure.

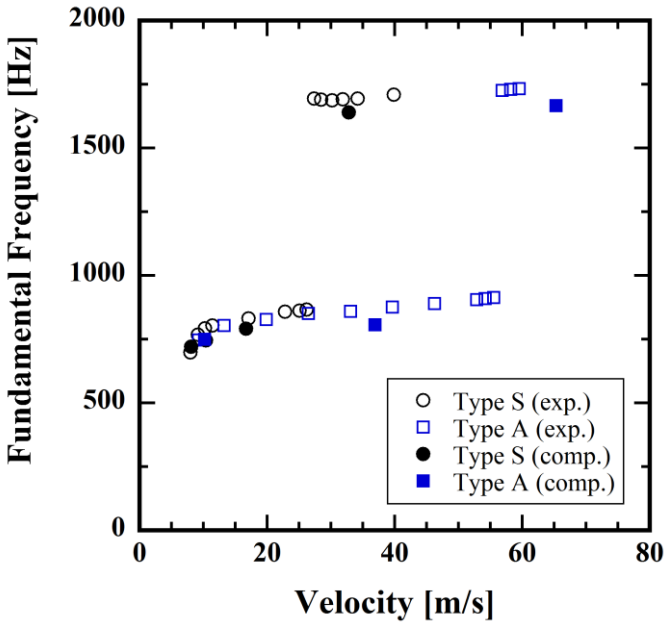


Figure 3: Relationship between the jet velocity U_0 at windway exit and fundamental frequency.

4 RESULTS AND DISCUSSIONS

4.1 Self-sustained oscillations for fundamental mode

Contours of pressure with a time-averaged component subtracted, p', (pressure-fluctuation) and velocity-fluctuation vectors around the exit of the windway and edge at U_0 = 10 m/s in Type S are shown in Figure 4. In the figure, the vectors are shown in the region where the effects of the fluid-dynamical structures (such as vortices) on the fields are small, and the intensity and direction of the acoustic particle velocity are indicated. The time origin (t = 0) is the time that the pressure becomes in the downward region of the edge in fundamental time period T.

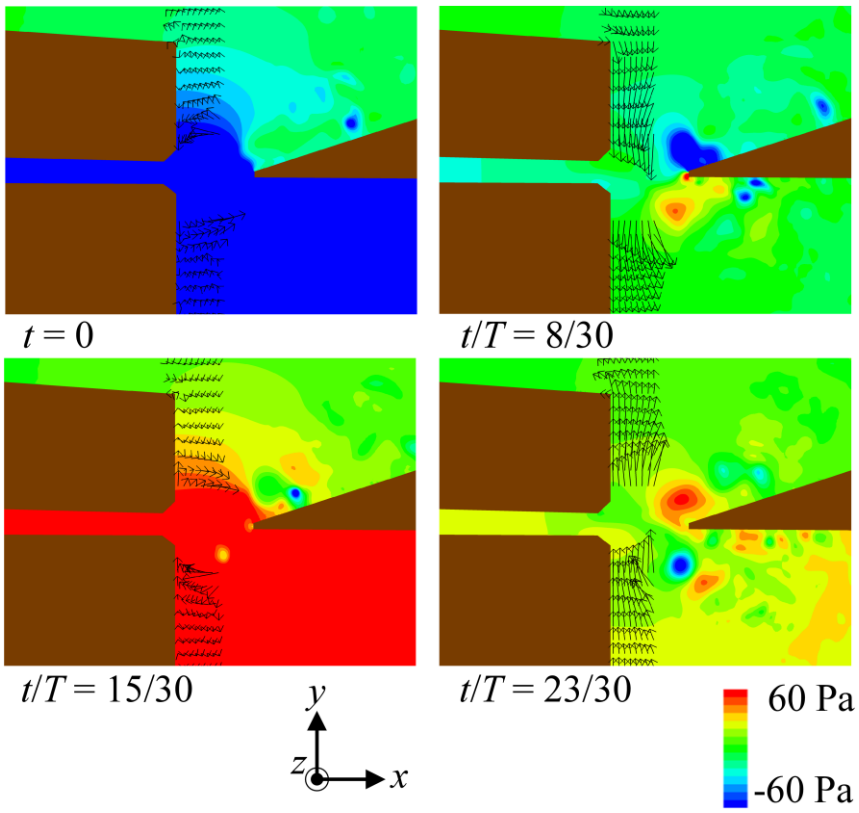


Figure 4 Contours of pressure-fluctuation and velocity-fluctuation vectors (z = 0) at U_0 =10 m/s in Type S, where T is the fundamental time period.

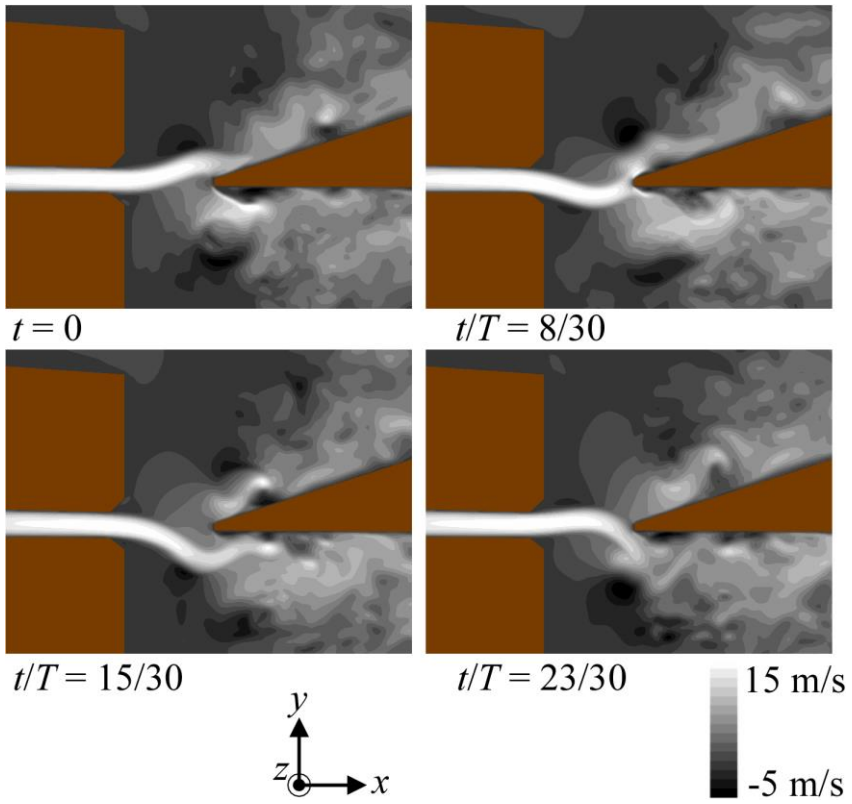


Figure 5 Contours of streamwise velocity (z = 0) at U_0 = 10 m/s in Type S, where T is the fundamental time period.

At t/T = 8/30 in Figure 4, the highest downward particle velocity is induced. Predicted contours of instantaneous streamwise velocity fields are shown in Figure 5. At t/T = 8/30, the jet starts to be deflected downward because the acoustic particle velocity is highest. This mechanism is consistent with experimental results reported by Coltman [16]. Moreover, these small deflections (disturbances) are developed and convected downstream, and as shown in Figure 5, the downward deflection of the jet becomes most intense at t/T = 15/30. As a result, air flows into the resonator on mass and compresses the air in it. In the opposite phase, namely, at t = 0, the air in the resonator expands when the upward deflection of the jet is largest.

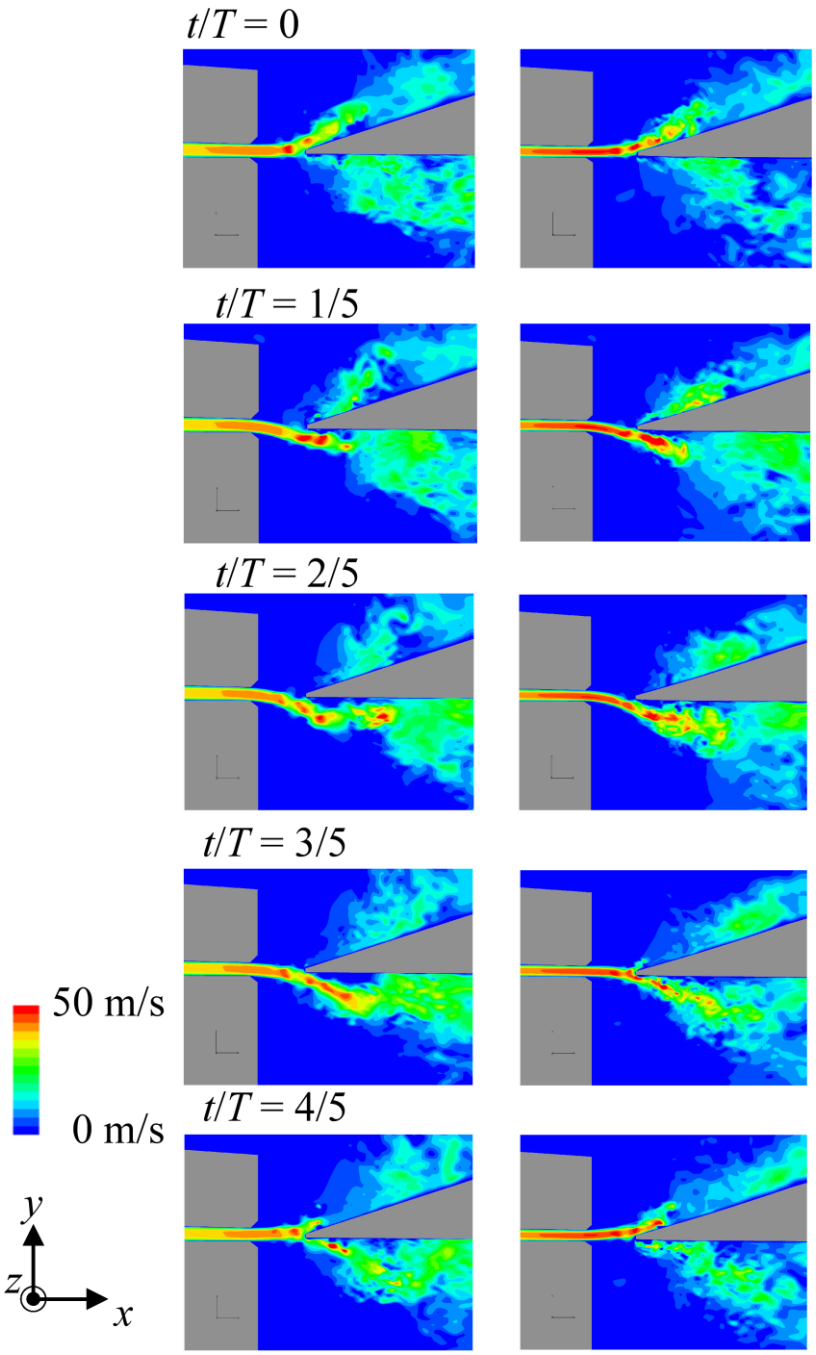


Figure 6 Contours of streamwise velocity (z = 0) at U_0 = 32 m/s in Type S (left) and at U_0 = 35 m/s in Type A (right), where T is fundamental time period.

4.2 Effects of shape on self-sustained oscillations around velocity of mode change

Predicted contours of instantaneous streamwise velocity fields at U_0 = 32 m/s in Type S (left) and at U_0 = 35 m/s in Type A (right) are shown in Figure 6, where the mode change occurs only in Type S.

As shown in the figures, the jet oscillates equally in the upward and downward directions for Type A. However, for Type S, the time in which the jet is inclined in the downward direction is longer than that in the upward direction. This asymmetry of jet oscillations distorts the sound generated by the oscillation of the jets of the fundamental mode, and possibly stimulates the higher mode and the mode change. The position of the edge is relatively higher for Type S as mentioned before, and this is one of the possible reasons why the jet is more likely to be inclined in the downward direction. Giordano [4] also shows that the position of the edge affects the ratios of the power of the fundamental and harmonic tones. Verge [6] also found that the minimum in the generation of the second harmonic was observed when the edge is slightly decentred with respect to the windway exit (y_e = -0.2h). This shift in effective zero transverse position of the edge could be associated with a result of an air entrainment by the jet which can create a static vacuum in the pipe (Coanda effect) and deflect the jet. However, the clear reason why this asymmetry of the jet oscillations is more likely to occur for Type S with comparison to Type A and the relationhip this asymmetry and the mode change are issues that needs further study.

Figure 7 shows the profiles of the mean streamwise velocity at U_0 = 10 and 32 m/s for Type S and at U_0 = 35 m/s for Type A between the windway exit and the edge, where the vertical axis is non-dimensionalized vertical position by the height of the edge, t, with the reference to the position of the center of the edge, y_e.

Comparing between the profiles at U_0 = 10 m/s and those at U_0 = 32 m/s for Type S, it is clarified that the velocity peak is relatively sharper at x/l = 0.5 and 0.75 at U_0 = 32 m/s. This indicates that the smaller amplitude of the oscillations of the jet makes the momentum diffusion smaller at the higher velocity of U_0 = 32 m/s. The variation of the position of the peak velocity along the streamwise position, x, in this figure also shows that the jet is inclined into the downward of the edge at U_0 = 32 m/s for Type S particularly in the profiles with $x/l \geq$ 0.5, while the jet approximately straightly collides with the edge and the jet oscillations are symmetry for Type A. This inclination is due to the asymmetry of the jet oscillations as mentioned above.

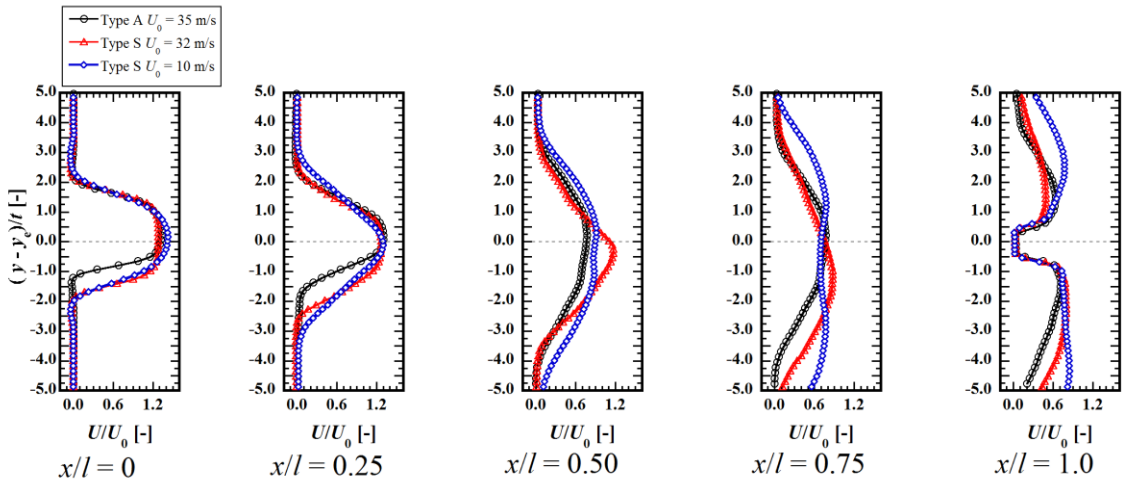


Figure 7 Profiles of mean streamwise velocity (x/l = 0.8) at U_0 = 10 and 32 m/s in Type S and at U_0 = 35 m/s in Type A.

The contours of the pressure-fluctuation, p', are shown in Figure 8. The right figures show that the fundamental mode (a half-wavelength acoustic mode in the pipe) is still dominant for Type A. Meanwhile, focusing on the figures at t/T = 4/5, it can be confirmed that the effects of the second mode (one-wavelength acoustic mode) on the sound fields are more intense for Type S.

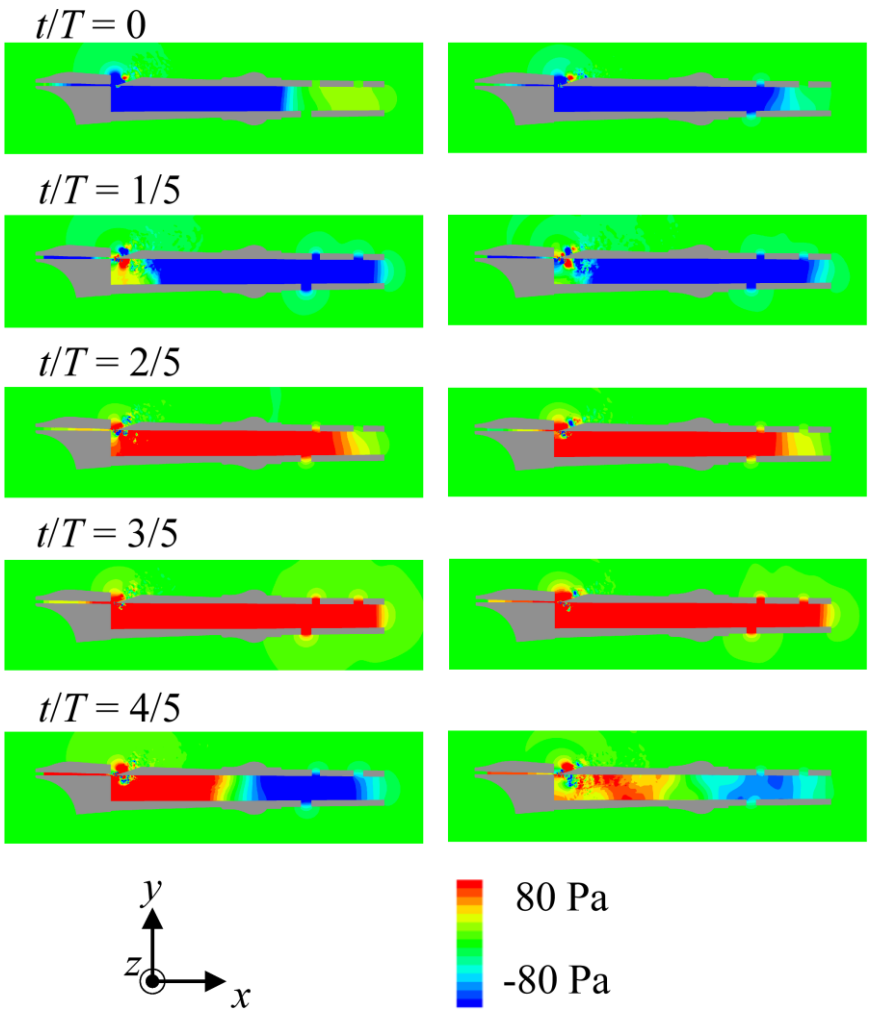


Figure 8 Contours of streamwise velocity (z = 0) at U_0 = 32 m/s in Type S (left) and at U_0 = 35 m/s in Type A (right), where T is the fundamental time period.

5 CONCLUSIONS

Focusing on the mode change of the sound radiating from the recorder, the flow and acoustic fields were predicted with direct aeroacoustic simulations around two actual recorders with different shapes. As a result, the following findings were acquired.

- The present simulations reproduce the mode change of both the recorders, where the velocity for the mode change depends on the configurations and shapes of the windway and edge.
- The asymmetry of the jet oscillations between in the upward direction and in the downward direction was observed only in the recorder model of the earlier mode change.

ACKNOWLEDGEMENTS

The present study was supported by JSPS KAKENHI Grant of No. 26820044 and through the Next-generation Supercomputer Strategy Program by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT).

REFERENCES

[1] B. Fabre, J. Gilbert, A. Hirschberg, and X. Pelorson, "Aeroacoustics of musical instruments", *Annual Review Fluid Mechanics*, **44**, 1-25, 2012.

[2] H. von Helmholtz, *On the Sensation of Tones*, translated by A. J. Ellis (Dover, New York, 1954), 88-94.

[3] L. Rayleigh (J. W. Strutt), *The Theory of Sound* (Dover, New York, 1945), Vol. 2, Chap. 21.

[4] N. Giordano, "Simulation studies of a recorder in three dimensions", *Journal of the Acoustical Society of America*, **135** (2), 906-916, 2014.

[5] Y. Hiroshi, A. Miki, H. Onitsuka, and A. Iida, *Direct numerical simulation of fluid-acoustic interactions in a recorder with tone holes, Journal of the Acoustical Society of America,* **138** (2), 858-873, 2015.

[6] M. P. Verge, B. Fabre, A. Hirschberg, A. P. J. Wijnands, "Sound production in recorder-like instrument. I. Dimensionless amplitude of the internal acoustic field," *Journal of the Acoustical Society of America*, **101** (5), 2914-2924, 1997.

[7] S. K. Lele. "Compact finite difference schemes with spectral-like resolution," *Journal of Computational Physics*, **103**, 16-42, 1992.

[8] P. Angot, CH Bruneau, and P. Frabrie, "A penalization method to take into account obstacles in viscous flows," Numerische Mathematik **81** (4), 497–520, 1999.

[9] Q. Liu and O. V. Vasilyev, "A Brinkman penalization method for compressible flows in complex geometries", *Journal of Computational Physics*, **227**, 946-966, 2007.

[10] R. Mittal and G. Iaccarino, "Immersed boundary methods," *Annual Review Fluid Mechanics*, **37**, 239-261, 2005.

[11] K. Matsuura and C. Kato, "Large-eddy simulation of compressible transitional flows in a low-pressure turbine cascade," *AIAA Journal*, **45**, 442-457, 2007.

[12] D. V. Gaitonde and M. R. Visbal, "Pade-type higher-order boundary filters for the Navier-Stokes equations," *AIAA Journal*, **38**, 2103-2112, 2000.

[13] K. W. Thompson, "Time dependent boundary conditions for hyperbolic systems," *Journal of Computational Physics*, **68**, 1-24, 1987.

[14] T. J. Poinsot and S. K. Lele, "Boundary conditions for direct simulations of compressible viscous flows," *Journal of Computational Physics*, **101**, 104-129, 1992.

[15] J. W. Kim and D. J. Lee, "Generalized characteristic boundary conditions for computational aeroacoustics," *AIAA Journal*, **38** (11), 2040-2049, 2000.

[16] J. W. Coltman, "Jet drive mechanism in edge tones and organ pipes," *Journal of the Acoustical Society of America*, **60**, 725-733, 1976.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL ANALYSIS OF RADIO FREQUENCY BLACKOUT FOR ATMOSPHERIC REENTRY VEHICLE USING CFD-CEM COMBINED METHOD

Yusuke Takahashi[1], Reo Nakasato[2], and Nobuyuki Oshima[3]

[1]Hokkaido University
Kita 13 Nishi 8, Kita-ku, Sapporo, Hokkaido, 060-8628 Japan
e-mail: ytakahashi@eng.hokudai.ac.jp
and
[2]Hokkaido University
Kita 13 Nishi 8, Kita-ku, Sapporo, Hokkaido, 060-8628 Japan
e-mail: reopon717@eis.hokudai.ac.jp
and
[3]Hokkaido University
Kita 13 Nishi 8, Kita-ku, Sapporo, Hokkaido, 060-8628 Japan
e-mail: oshima@eng.hokudai.ac.jp

Keywords: Radio Frequency Blackout, Atmospheric Reentry Capsule, Hypersonic Flow Solver, FD2TD Method

Abstract. *Numerical simulations of electromagnetic waves around the atmospheric reentry demonstrator (ARD) of the European Space Agency (ESA) in an atmospheric reentry mission were conducted with a combined method of computational fluid dynamics (CFD) and computational electromagnetics (CEM). During the reentry mission, radio frequency blackout and plasma attenuation of radio waves in communications with a data-relay satellite were observed. Cut-off of waves propagation was caused by highly dense plasma formed by a strong shock wave formed in front of the vehicle due to orbital speed. In this study, the physical properties of the plasma flow in the shock layer and wake region of the ESA ARD were obtained using a high-enthalpy flow solver applicable unstructured grids. Moreover, electromagnetic waves were calculated using a frequency-dependent finite-difference time-domain method using the plasma properties. The present analysis model was validated based on experimental flight data of ESA ARD. Comparisons of the measured and predicted results showed good agreements. Attenuation and reflection by reentry plasma were clarified in detail using the present model. It was suggested that the analysis model could be an effective tool for investigating radio frequency blackout and plasma attenuation in radio wave communication. Behaviors of plasma and electromagnetic waves around ESA ARD for several altitudes were investigated using the proposed model in detail.*

1 INTRODUCTION

One of the most important issues in the entry, descent, and landing approach of spacecraft is radio frequency (RF) blackout; that is, the cutoff of communication between the reentry vehicle and ground stations and/or data-relay satellites. During reentry from a low Earth orbit, a vehicle flies at orbital speed of more than 7 km/s, and a strong shock wave is formed at its front. Freestream air is rapidly dissociated by aerodynamic heating in the shock wave layer. Charged species, i.e., electrons and ions, are created, and these species then cover the vehicle's surface. RF blackout is caused by a highly dense plasma layer on the vehicle surface, attenuating or reflecting electromagnetic waves used for telecommunication.

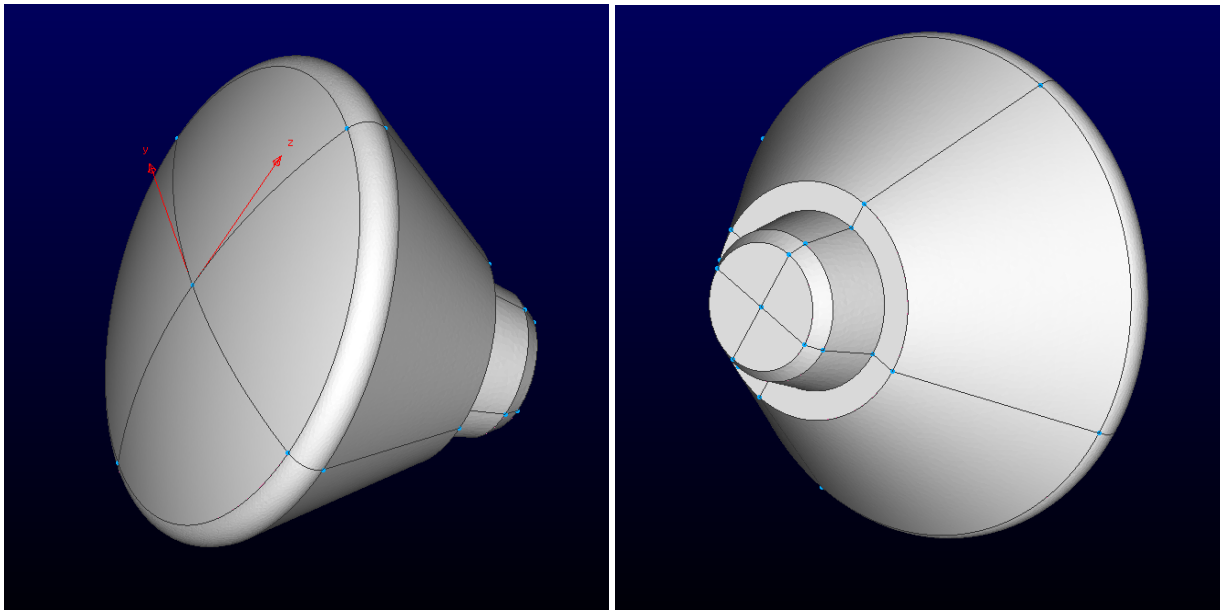
Because the RF blackout phenomenon is affected by the flow properties around the vehicle, specifically the number density of electrons, clarifying the distribution of charged particles is a key issue. In particular, determination of the plasma distribution around a blunt-body reentry capsule is expected to be important from the point of view of their frequent use in reentry missions. Recently, remarkable progress has been made in the development of high-performance computers and algorithms for numerical simulation. Computational fluid dynamics (CFD) techniques have been used widely to predict flow fields around reentry vehicles. In addition, computational electromagnetics (CEM) to describe electromagnetic waves in reentry plasma are also important approach. In particular, the frequency-dependent finite-difference time-domain (FD2TD) method [1] has also been an effective tool for simulating electromagnetic waves in a frequency-dependent medium [2]. An analysis approach to predict RF blackout and plasma attenuation for a reentry vehicle, based on a method combining CFD and CEM, has been suggested by Takahashi et al. [3, 4, 5].

Flight experiment for the atmospheric reentry demonstrator (ARD) [6] of the European Space Agency (ESA) was conducted. In the ESA ARD mission, reentry plasma attenuation of electromagnetic waves in communication with the tracking and data-relay satellite (TDRS) was measured. ESA ARD had typical blunt-body capsule shape and typical reentry trajectory, which are widely used in a reentry mission. Thus, it is worth investigating the distributions of plasma and electromagnetic waves for design and development of new reentry vehicles and for research and development concerning RF blackout mitigation techniques. Takahashi [5] recently developed an RF blackout prediction tool combining RG-FaSTAR, which is a high-enthalpy flow solver applicable to unstructured grid systems, and FD2TD simulation code. Using the analysis models, it is possible to investigate the behavior of electromagnetic waves in reentry plasma. The present research objective is to reproduce the actual reentry environment during the ESA ARD flight and to clarify the plasma flow field and electromagnetic wave behavior.

2 ATMOSPHERIC REENTRY DEMONSTRATOR

The atmospheric reentry demonstrator (ARD) [6] of the European Space Agency (ESA) was launched by the Ariane-5 V503 rocket on October 21, 1998. The reentry capsule was separated from the Ariane-5 at an altitude of 218 km. ESA ARD reached its apogee at an altitude 830 km, and then reentered the Earth's atmosphere. It was reported that the reentry velocity at an altitude of 120 km was approximately 7452 m/s. Because of its high reentry velocity, a shock wave was generated in front of the vehicle, and the vehicle was surrounded by plasma. After deceleration by the aerodynamic force, ESA ARD began its recovery sequence. Finally, ESA ARD splashed down in the Pacific Ocean. The reentry trajectory (position, velocity, and attitude reconstructed by acceleration measurements) was obtained using on-board inertial navigation system measurements and a GPS receiver.

The analytical shape of the ESA ARD is shown in Fig. 1. Its shape was basically the same as the Apollo command module scaled down to 70%. It comprised the front shield section, rear cone section, and back cover section. The capsule was essentially a spherical cone with a nose radius of 3,360 mm and a base radius of 2,800 mm for the front shield section with a half-angle of 33° for the rear section. To protect it from strong aerodynamic heating, an ablator material was coated onto the front heat shield.



(a) Front side view (b) Rear side view

Figure 1: Analysis model of the ESA ARD.

ESA ARD employed eight dedicated skin antennas for a telemetry link and communication with TDRS. Its link frequencies were approximately 2.267 GHz. Measurements were made of the RF blackout and plasma attenuation caused by the plasma around the vehicle generated by the shock wave. No RF blackout for the TDRS link was observed, even though strong plasma attenuation due to reentry plasma around ESA ARD was measured from an altitude of 86 km to an altitude of 44 km.

3 PLASMA FLOW MODELING

Physical model and numerical method used in this plasma flow simulation are basically the same as those used in Takahashi's work [5], except for part numerical scheme and calculation conditions. Here, brief explanation about the present plasma flow model are presented.

3.1 ANALYSIS MODEL

The flow field around the reentry capsule is expressed by the Navier-Stokes equations extended to a multi-temperature model and the equation of state. The governing equations include total mass, momentum, total energy, species mass, rotation energy, vibration energy, and electron energy conservations.

Transport properties such as the viscosity, thermal conductivity, and binary diffusion coefficients for a gas mixture are evaluated by Yos's model [7], which is based on the first Chapman-Enskog approximation [8]. The collision cross sections are given by Gupta [9] and Fertig

[10, 11]. The diffusion coefficients are expressed by Curtiss and Hirschfelder's formula [12]. Ambipolar diffusion is assumed for charged species.

For chemical reactions, the inflow gas is assumed to consist of 11 chemical species (N_2, O_2, NO, N_2^+, O_2^+, NO^+, N, O, N^+, O^+ and e^-) and 49 reactions are assumed to occur. The chemical reaction rate is determined using an Arrhenius-type form, and the reaction rate coefficients are obtained from Park [13]. The backward reaction rates are evaluated from the equilibrium constants which, being functions of temperature only, are calculated by the curve-fitting formula given in [14, 9].

The energy transfer between each of the internal energy modes are consider as follows: translation-rotation (T-R) [15], translation-vibration (T-V) [16, 17], translation-electron (T-e) [18, 19, 20], rotation-vibration (R-V) [21], rotation-electron (R-e) [22, 23], and vibration-electron (V-e) [24]. The energy losses/releases for vibrations and rotations associated with chemical reactions reveal the dissociation energies for heavy-particle impact reactions. These can be obtained using a non-preferential dissociation model [20]. In addition, the electron energy losses/releases due to electron-impact dissociation and ionization are also considered.

3.2 NUMERICAL SCHEMES

The governing equations for the flow field are transformed into the delta form and solved using a finite-volume approach. The numerical fluxes of the advection term are calculated using the Harten, Lax, van Leer, and Einfeldt (HLLE) scheme [25] using the monotonic upstream-centered scheme for interpolation of the conservation laws (MUSCL). On the other hand, the viscous terms of the equation are calculated using the second-order central differencing method. The spatial gradients of the flow variables are evaluated with the Green-Gauss method. A time integration is performed implicitly using the lower-upper symmetric Gauss-Seidel (LU-SGS) method [26] combined with the point-implicit method [27] for stable computation.

3.3 SOFTWARE PACKAGE

These physical and numerical models are implemented in the high-enthalpy flow solver RG-FaSTAR [5]. RG-FaSTAR is a version of the fast unstructured CFD code FaSTAR [28] originally developed by JAXA that incorporates real gas effects. These solvers can handle unstructured grid systems such as tetrahedral, triangular prism, square pyramid, and hexahedral meshes, and have high flexibility in generating the computational grid. For massive parallel computation on high-performance computers, the message-passing interface (MPI) technique with a domain partition approach is used.

3.4 CALCULATION CONDITIONS

Numerical analyses of flow fields by RG-FaSTAR are performed for cases as listed in Table 1. Note that AOA denotes the angle of attack of the ESA ARD. The freestream velocity is determined according to Tran's report [6].

Computational grids for ESA ARD in RG-FaSTAR at an altitude of 85 km are composed of tetrahedral meshes, prismatic layer meshes near the wall, hexahedral meshes in the front region, and squared pyramids. Figure 2 shows a computational grids at an altitude of 85 km. To finely resolve the shock layer near ESA ARD, hexahedral meshes are inserted into its front region. In addition, prismatic meshes near walls are used to solve the boundary layer near the wall. Square pyramid meshes are used to connect the hexahedral meshes and the prism meshes. The numbers of cells and nodes of the grid system at altitude of 85 km are 4,244,601 and 2,434,920,

Table 1: Freestream parameters.

Altitude, km	Density, kg/m^3	Temperature, K	Velocity, m/s	AOA, degree
85.0	8.18×10^{-6}	191.0	7577	20.0
80.0	1.85×10^{-5}	195.8	7609	20.0
75.0	4.07×10^{-5}	201.7	7593	19.2
70.0	8.83×10^{-5}	210.9	7542	19.2
65.0	1.78×10^{-4}	225.3	7145	20.0
60.0	3.40×10^{-4}	256.5	6105	19.4
55.0	6.31×10^{-4}	242.0	5393	19.8
50.0	1.15×10^{-3}	265.3	4567	20.0
45.0	2.25×10^{-3}	255.6	3323	20.8
40.0	4.36×10^{-3}	250.3	2218	21.2

respectively. At other altitudes, the number of cells and nodes are different but the composition of mesh types is basically similar.

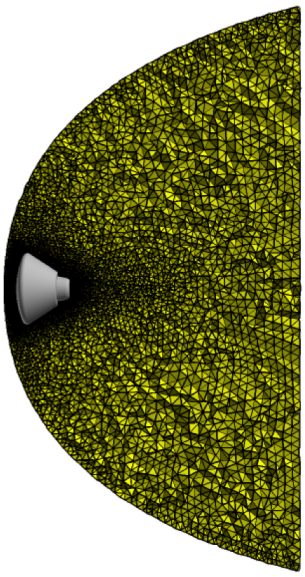


Figure 2: Computational grids system for plasma flow around ESA ARD at altitude of 85 km.

For the boundary conditions, inlet, outlet, and wall conditions are considered, as shown in Fig. 3. Freestream density, temperature, and velocity, listed in Table 1, are given at the inflow boundary. Mass fractions of chemical species are always set to 0.765 for N_2 and 0.235 for O_2 at the inflow. The degree of ionization at the inflow condition is always set to 10^{-10}. At the outlet boundary, all flow variables are determined by the zeroth-order extrapolation. On the wall surface, non-pressure gradient in the wall normal direction and no-slip wall are given. In addition, the wall temperature is fixed at 273K. Moreover, the wall surface is assumed to be non catalytic for chemical species.

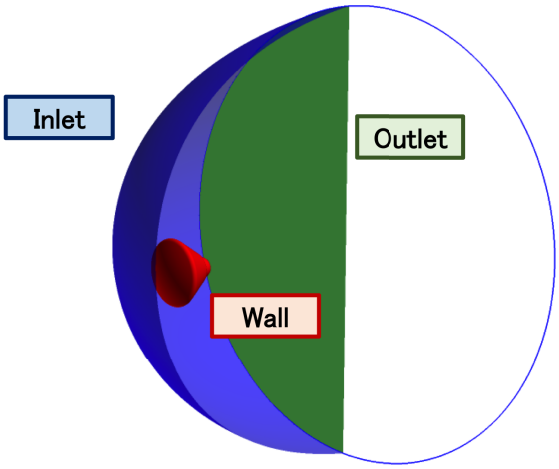


Figure 3: Boundary conditions for plasma flow around ESA ARD at altitude of 85 km.

4 ELECTROMAGNETIC WAVE MODELING

4.1 GOVERNING EQUATION

In order to evaluate RF blackout and plasma attenuation in communication between ESA ARD and TDRS, we clarify the behavior of electromagnetic waves in the plasma region during reentry. Here, the Maxwell's equations for describing the behavior of electromagnetic waves are adopted.

4.2 COMPUTATION METHOD

So far, the finite-difference time-domain (FDTD) method has been widely used for simulating electromagnetic wave propagation. On the other hand, plasmas are generally frequency-dependent media. Dielectric constants and electrical conductivities of plasma vary over a frequency range. It is possible that the FDTD formulation leads to a risk to predict incorrect results when applying it to frequency-dependent media, because the FDTD method assumes that these medium properties are independent of frequency. On the other hand, the FD2TD method is effective for discretizing the Maxwell's equations in such frequency-dependent media.

Figure 4 shows the computational domain in the FD2TD approach for cases of altitudes from 55 to 40 km. it should be noted that different computational domains are used at the other altitudes. The computational domain for electromagnetic waves is set to be a cuboid 10.5 m in length in the x direction, 5.5 m in width in the y direction, and 5.0 m in height in the z direction. All of the computational grid is composed of hexahedral meshes and the grid nodes are set to 800 (x-direction), 420 (y-direction), and 380 (z-direction), i.e., 127,680,000 nodes. The origin of the computational domain used in the CFD approach corresponds to that of the simulation of the electromagnetic waves. The physical properties, ω_{p} and ν_{c}, required in the present FD2TD method are obtained by using the computational results obtained by RG-FaSTAR. To reduce computational cost, the magnetic permeability, μ is set to be the same as that of free space, μ_0.

As a boundary condition, Mur's first-order absorption boundary [29] is imposed at all faces of the computational domain. A conduction boundary condition is imposed on the surface of the reentry vehicle.

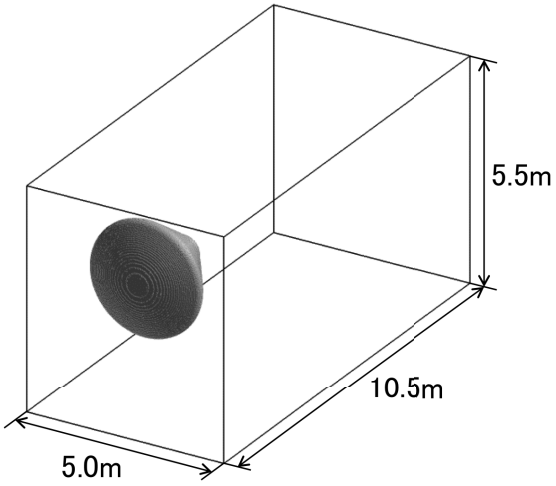


Figure 4: Computational domain for electromagnetic wave around ARD.

4.3 ANALYSIS CODE

The physical and numerical models have been implemented in the electromagnetic wave simulation code "Arcflow/Arcwave" developed by Hokkaido University. For fast computation and efficient allocation of adequate memory, the simulation code is parallelized using the MPI and OpenMP techniques. In addition, a data transfer tool between the electromagnetic wave simulation code and RG-FaSTAR is also implemented. Both the simulation codes using the CFD and FD2TD methods are packaged.

5 RESULTS AND DISCUSSION

5.1 SURFACE PRESSURE

First of all, to validate the plasma flow analysis model used here, stagnation pressure coefficients in front of the ESA ARD between the flight experiment and numerical analysis are compared. The pressure obtained in the ESA ARD flight was reconstructed based on the flight data on velocity and density deduced by the on-board acceleration measurement. The computed pressure coefficient at stagnation, C_p, is defined as the peak pressure on the front surface of the vehicle, which is obtained by the following expression:

$$C_p = \frac{p_0}{\frac{1}{2}\rho_\infty U_\infty^2}, \tag{1}$$

where p_0 is the peak pressure on the surface. In addition, ρ_∞ and U_∞ are the freestream density and velocity, respectively. Figure 5 shows a comparison of the stagnation pressure coefficient profiles with altitude from 85 km to 40 km. Error between the flight experiment results and the numerical analysis is several percentages at most except for an altitude of 85 km. This indicates that the plasma flow analysis correctly reproduces the aerodynamic tendencies in the shock layer.

5.2 SIGNAL LOSS

Gain of the electromagnetic waves is attenuated by plasma around the reentry vehicle, which was measured in the ESA ARD mission. To validate the CFD and CEM combined method,

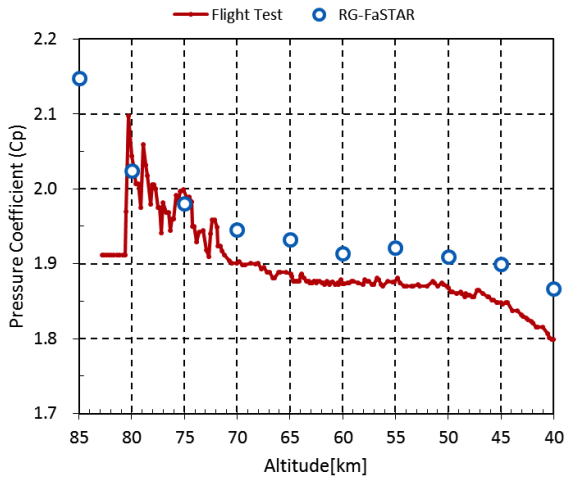


Figure 5: Comparison of the pressure coefficient at stagnation at the vehicle's front between the flight experiment and numerical analysis.

signal loss of the electromagnetic waves that TDRS received are compared between the experiments by ESA ARD mission and the present computed results. When the vehicle is completely surrounded by highly dense plasma whose electron number density exceeds the critical value, no signal is detected, and then RF blackout is observed. In the present study, the signal loss is calculated by

$$G_{\mathrm{SL}} = -10\log\left(\frac{|\mathbf{E}_{bd}|^2}{|\mathbf{E}_{bd,free}|^2}\right). \tag{2}$$

where $\mathbf{E}_{bd}$ and $\mathbf{E}_{bd,free}$ are the electric fields on certain far field boundary at an altitude and in free space where there is no plasma, respectively. The location of the TDRS from the viewpoint of ESA ARD during reentry can be obtained from its flight path. In addition, it is deduced that the windows where the TDRS is visible at each altitude shown in Fig. 6 are on the boundaries of the FD2TD computational domain.

Comparison of the signal loss profile along an altitude is shown in Fig. 7. The measured attenuation of the communication waves increases with small fluctuation to altitude of 70 km, and then, the communication recovers with decrease in attenuation. No RF blackout is observed in the flight experiment as an important event. The computed result relatively shows good agreement with the measurement at altitude of 85 km and below 60 km. However, at altitudes between 80 km and 65 km, electromagnetic waves near ESA ARD are completely blocked by the reentry plasma, and then the signal loss almost becomes zero. This is possibly because the plasma flow simulation overestimates number density of electron in the rear region of ESA ARD. In particular, reaction models of associative ionizations which are primary ionization and recombination reactions of electrons in the altitude range can attribute the discrepancy.

5.3 ELECTRON NUMBER DENSITY

The distribution of electron number density around ESA ARD at an altitude of 85 km is shown in Fig. 8. In these figures, "y-z plane" shows a plane cutting a region where antenna for communication with TDRS exists. A region appears to exist in which the electron number density generated by the shock layer in front of ESA ARD remains high at the altitude. When the electromagnetic waves pass through the regions where the electron density exceeds the

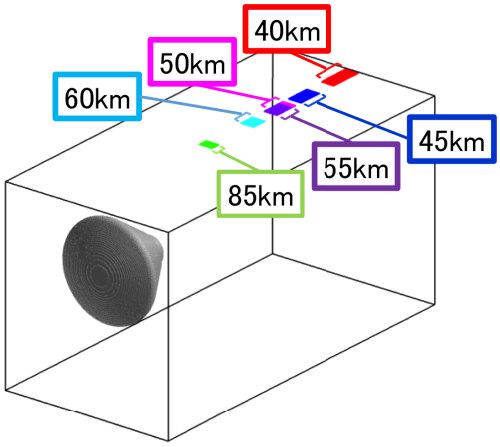


Figure 6: Windows in which the TDRS is visible at each altitude on the boundaries of the FD2TD domain.

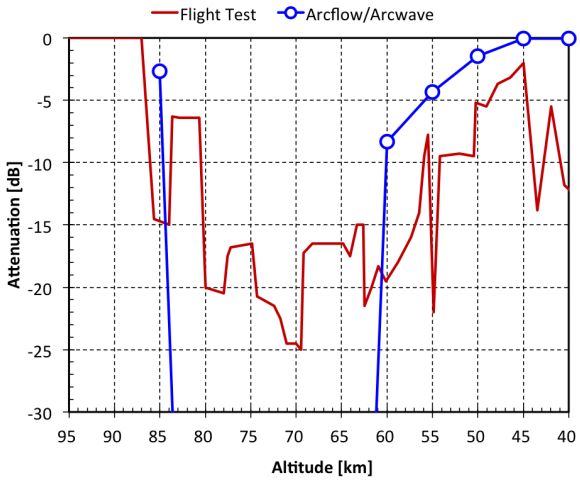
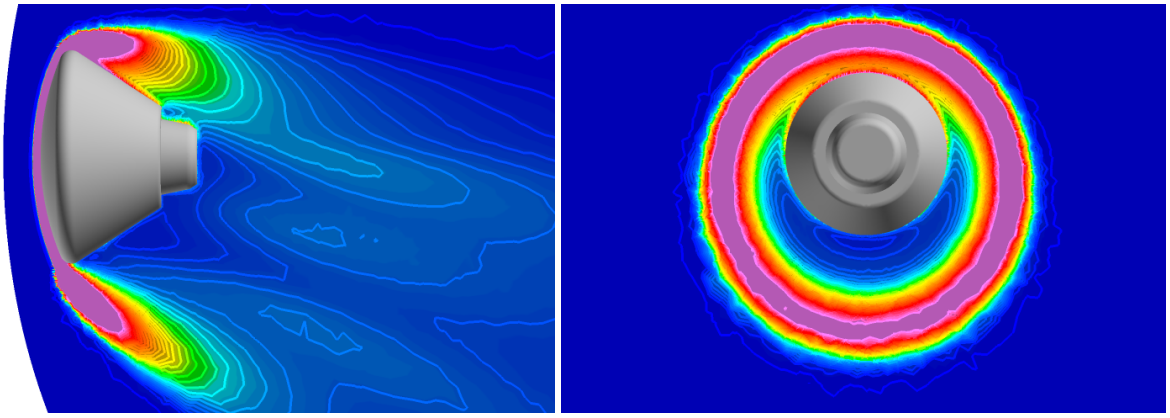


Figure 7: Comparison of the signal loss between the flight experiment and numerical analysis.

critical value (6.4×10^{16} 1/m^3) for the link frequency (2.267 GHz), attenuation, refraction, or reflection will occur. On the other hand, compressed gas in the shock layer expands when passing near the shoulder of the vehicle inflowing in the wake region. The plasma density decreases to well below the critical value at the rear of the vehicle because of its supersonic expansion.

5.4 ELECTROMAGNETIC WAVES

Number density of electron and collision frequency required in the electromagnetic wave calculations are computed by RG-FaSTAR and then, are transposed in the FD2TD computational domain with the data transfer tool. Figures 9, 10, 11, and 12 show electric field magnitudes of the electromagnetic waves for the cases of free space, an altitude of 85 km, an altitude of 60 km, and an altitude of 50 km, respectively. The magnitude of the electric field is normalized to 1.0 on the antenna in the calculation. All values of the electric field in the figures are expressed on a logarithmic scale.



(a) x-z plane (b) y-z plane

Figure 8: Number density of electrons around the ESA ARD at an altitude of 85 km.

The electric field distribution in free space where there is no plasma is presented to compare with results for cases of other altitudes. For the free-space condition, the electromagnetic waves can propagate in all directions, diffracting and reflecting on the ESA ARD surface. On the other hand, below an altitude of 85 km, electromagnetic waves cannot propagate on the front side of ESA ARD along the shock layer surface, because electromagnetic waves are completely reflected. As mentioned above, moreover, electromagnetic waves cannot propagate in any direction between altitudes of 80 km and 65 km due to highly-dense plasma. Thus, a complete RF blackout occurs, although this is not shown here. As mentioned above, however, this RF blackout was not observed in communication between ESA ARD and the TDRS using a link frequency of 2.267 GHz during atmospheric reentry. Uncertainties in chemical reaction models used in this analysis model causes this discrepancy, as pointed out by Takahashi et al. [4]. At an altitude of 60 km, wave propagation recovers from the RF blackout, while strong attenuation and reflection in the plasma wake are caused by the recompression gas region. The number density of electrons becomes low in the ESA ARD's rear region at an altitude of 50 km and electromagnetic waves propagate in a similar fashion to the computed results for the free-space condition.

6 CONCLUSIONS

Numerical analyses of plasma flows and communication waves around the Atmospheric Reentry Demonstrator (ARD) were conducted for prediction of radio frequency blackout and plasma attenuation. The ESA ARD capsule shape and its reentry trajectory were typical for reentry missions with blunt-body reentry capsules. The plasma flow models were implemented with an unstructured grid high-enthalpy flow solver "RG-FaSTAR" based on the computational fluid dynamics (CFD). Moreover, electromagnetic waves behaviors were expressed using a simulation code "Arcflow/Arcwave," which uses the frequency-dependent finite-difference time-domain (FD2TD) method. Computed results were validated with experimental data for stagnation pressure and signal loss profiles during the atmospheric reentry. It was indicated that the present analysis model is effective for predicting reentry plasma and signal loss of radio waves for this kind of reentry capsule mission, except for a strong attenuation altitude. As an important parameter for electromagnetic wave simulation in reentry plasma, distributions of the electron number density were investigated. In addition, radio frequency blackout and plasma attenuation were discussed using information. It was found that reflection, diffraction, and attenuation on

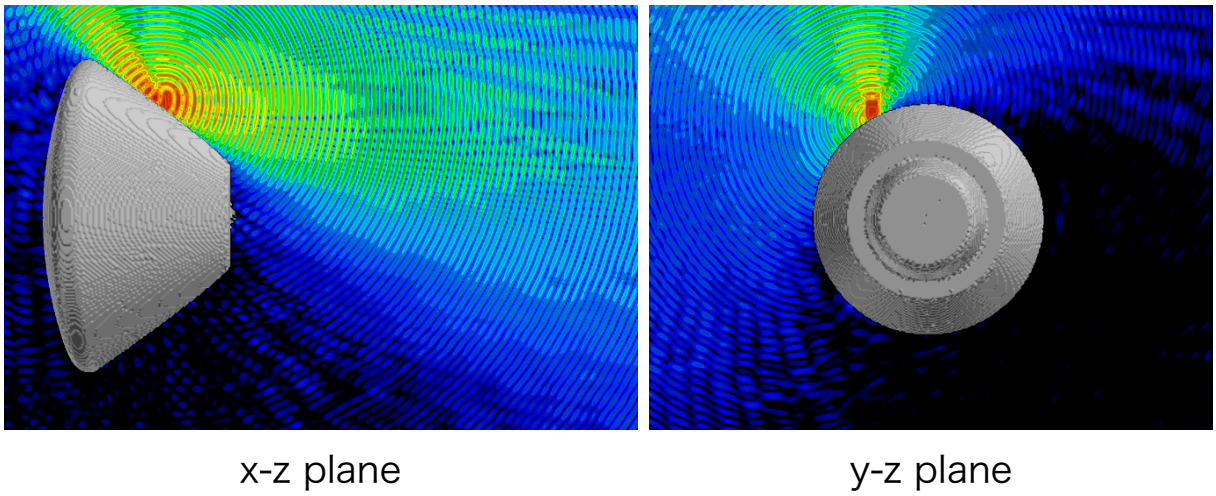


Figure 9: Electromagnetic waves behavior (electric field magnitude in common logarithm scale) around ESA ARD in free space.

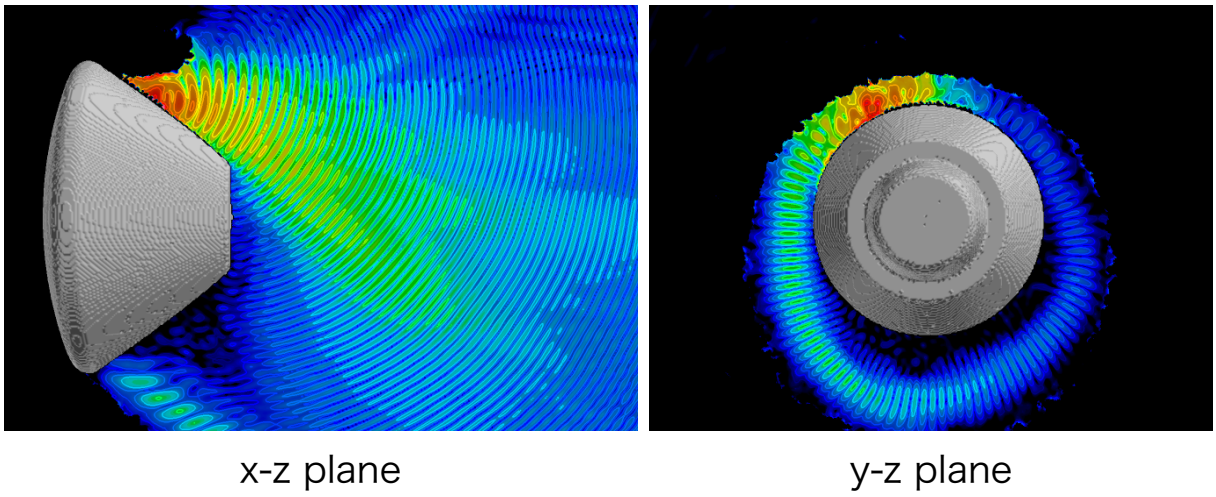


Figure 10: Electromagnetic waves behavior (electric field magnitude in common logarithm scale) around ESA ARD at an altitude of 85 km.

the strong shock wave surface in front of the vehicle occur, while the electromagnetic waves propagate in wake region in the rear of vehicle.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26820366. The present computational results were obtained using the fast unstructured CFD code "FaSTAR," which was developed by the Japan Aerospace Exploration Agency (JAXA). This research used computational resources of the HPCI system provided by the Information Initiative Center, Hokkaido University through the HPCI System Research Project (Project ID:hp150100). The computations were performed mainly using the computational facilities (HITACHI HA8000-tc/HT210 and HITACHI SR16000 VM1) at the Research Institute for Information Technology, Kyushu University, and the supercomputer system (HITACHI SR16000 model M1) at the Information

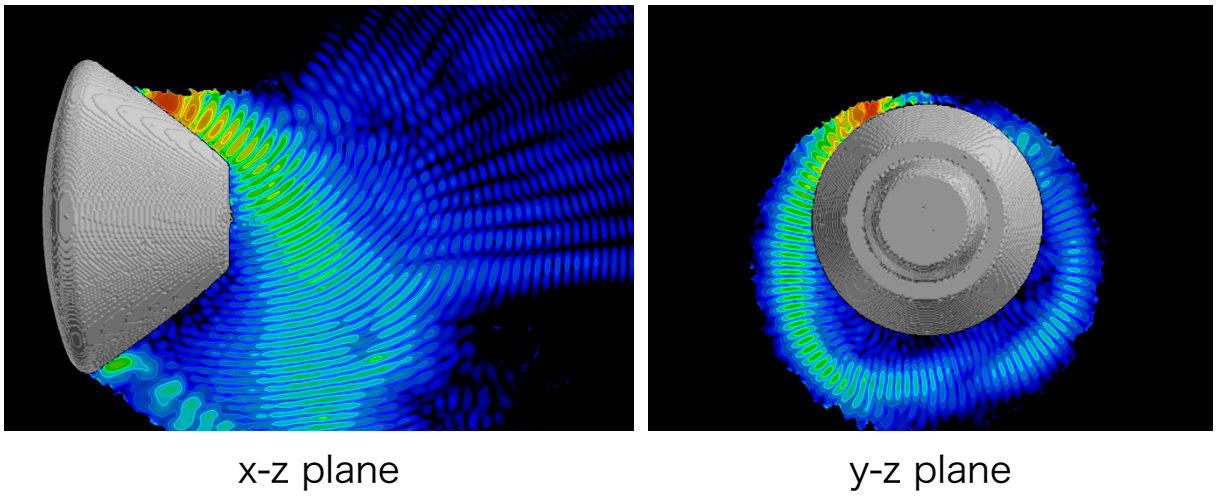


Figure 11: Electromagnetic waves behavior (electric field magnitude in common logarithm scale) around ESA ARD at an altitude of 60 km.

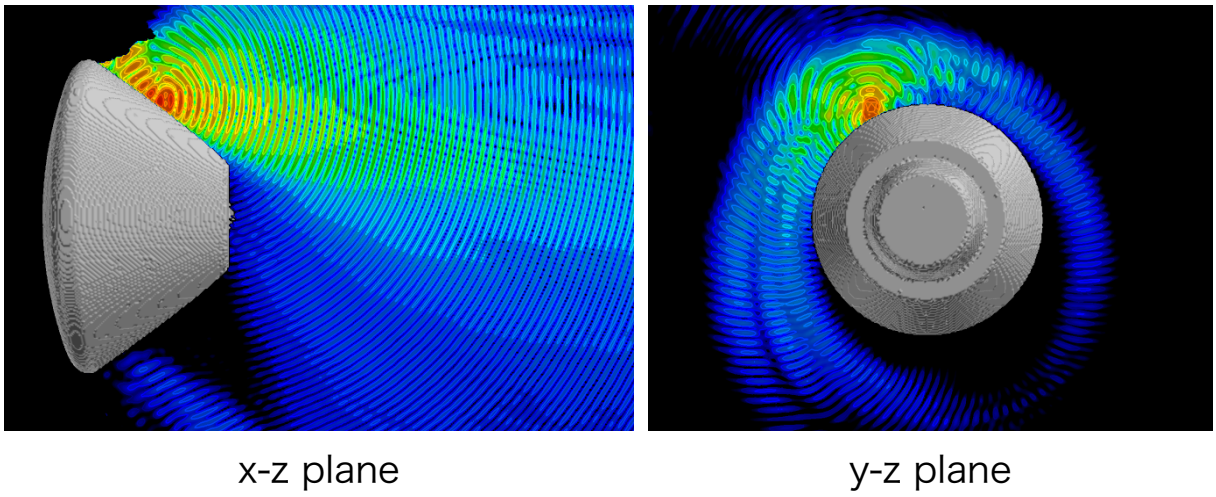


Figure 12: Electromagnetic waves behavior (electric field magnitude in common logarithm scale) around ESA ARD at an altitude of 50 km.

Initiative Center, Hokkaido University.

REFERENCES

[1] K.S. Kunz and R.J. Luebbers. *The Finite Difference Time Domain Method for Electromagnetics*. CRC Press, Boca Raton, 1993.

[2] K. Kinefuchi, I. Funaki, and T. Abe. "Frequency-Dependent FDTD Simulation of the Interaction of Microwaves With Rocket-Plume". *IEEE Transactions on Antennas and Propagation*, 58(10):3282–3288, October 2010.

[3] Y. Takahashi, K. Yamada, and T. Abe. "Examination of Radio Frequency Blackout for an Inflatable Vehicle during Atmospheric Reentry". *Journal of Spacecraft and Rockets*, 51(2):430–441, March 2014.

[4] Y. Takahashi, K. Yamada, and T. Abe. "Prediction Performance of Blackout and Plasma Attenuation in Atmospheric Reentry Demonstrator Mission". *Journal of Spacecraft and Rockets*, 51(6):1954–1964, Novermber 2014.

[5] Y. Takahashi. "Advanced validation of CFD-FDTD combined method using highly applicable solver for reentry blackout prediction". *Journal of Physics D: Applied Physics*, 49(1):015201, 2016.

[6] P. Tran, J.C. Paulat, and P. Boukhobza. "Re-entry Flight Experiments Lessons Learned - The Atmospheric Reentry Demonstrator ARD". *Nato Report, RTO-EN-AVT-130-10*, pages 10–1 – 10–46, June 2007.

[7] J.M. Yos. "Transport Properties of Nitrogen, Hydrogen Oxygen and Air to 30,000 K". *TRAD-TM-63-7,* Research and Advanced Development Division, AVCO Corp., 1963.

[8] J.O. Hirschfelder, C.F. Curtiss, and R.B. Bird. *Molecular Theory of Gases and Liquids*. Wiley, New York, 1954.

[9] R.N. Gupta, J.M. Yos, R.A. Thompson, and K.P. Lee. "A Review of Reaction Rates and Thermodynamic and Transport Properties for an 11-Species Air Model for Chemical and Thermal Nonequilibrium Calculations to 30000 K". *NASA RP-1232*, Aug. 1990.

[10] M. Fertig, A. Dohr, and H.H. Frühaufu. "Transport Coefficients for High-Temperature Nonequilibrium Air Flows". *AIAA Paper 98-2937*, 1998.

[11] M. Fertig, A. Dohr, and H.H. Frühauf. "Transport Coefficients for High-Temperature Nonequilibrium Air Flows". *Journal of Thermophysics and Heat Transfer*, 15(2):148–156, April-June 2001.

[12] C.F. Curtiss and J.O. Hirschfelder. "Transport Properties of Multicomponent Gas Mixture". *Journal of Chemical Physics*, 17(6):550–555, June 1949.

[13] C. Park. "Assessment of a Two-Temperature Kinetic Model for Dissociating and Weakly Ionizing Nitrogen". *Journal of Thermophysics and Heat Transfer*, 2(1):8–16, Jan.-March 1988.

[14] C. Park. *Nonequilibrium Hypersonic Aerothermodynamics*. Wiley, New York, 1990.

[15] J.G. Parker. "Rotational and Vibrational Relaxation in Diatomic Gases". *The Physics of Fluids*, 2(4):449–462, July-Aug. 1959.

[16] R.C. Millikan and D.R. White. "Systematics of Vibrational Relaxation". *The Journal of Chemical Physics*, 39(12):3209–3213, Dec. 1963.

[17] C. Park. "Problems of Rate Chemistry in the Flight Regimes of Aeroassisted Orbital Transfer Vehicles". *AIAA Paper 84-1730*, 1984.

[18] J.P. Appleton and K.N.C. Bray. "The Conservation Equations for a Nonequilibrium Plasma". *Journal of Fluid Mechanics*, 20(4):659–672, June 1964.

[19] M. Mitchner and C.H. Kruger Jr. *Partially Ionized Gases*. Wiley, New York, 1973.

[20] P.A. Gnoffo, R.N. Gupta, and J.L. Shinn. "Conservation Equations and Physical Models for Hypersonic Air Flows in Thermal and Chemical Nonequilibrium". *NASA TP-2867*, Feb. 1989.

[21] C. Park. "Rotational Relaxation of N_2 Behind a Strong Shock Wave". *Journal of Thermophysics and Heat Transfer*, 18(4):527–533, Oct.-Dec. 2004.

[22] M. Nishida and M. Matsumoto. "Thermochemical Nonequilibrium in Rapidly Expanding Flows of High-Temperature Air". *Zeitschrift für Naturforschung, Teil A: Physik, Physikalische Chemie, Kosmophysik*, 52(4):358–368, 1997.

[23] S.S. Lazdinis and S.L. Petrie. "Free Electron and Vibrational Temperature Nonequilibrium in High Temperature Nitrogen". *Physics of Fluids*, 17(8):1539–1546, Aug. 1974.

[24] J.H. Lee. "Electron-Impact Vibrational Relaxation in High-Temperature Nitrogen". *Journal of Thermophysics and Heat Transfer*, 7(3):399–405, July-Sept. 1993.

[25] B. Einfeldt. "On Godunov-type methods for gas dynamics". *SIAM Journal on Numerical Analysis*, 25(2):294–318, 1988.

[26] A. Jameson and S. Yoon. "Lower-Upper Implicit Schemes with Multiple Grids for the Euler Equations". *AIAA Journal*, 25(7):929–935, July 1987.

[27] T.R.A. Bussing and E.M. Murman. "Finite-Volume Method for the Calculation of Compressible Chemically Reacting Flows". *AIAA Journal*, 26(9):1070–1078, 1988.

[28] A. Hashimoto, K. Murakami, T. Aoyama, K. Ishiko, M. Hishida, M. Sakashita, and P.R. Lahur. "Toward the Fastest Unstructured CFD". *AIAA Paper 2012-1075*, 2012.

[29] G. Mur. "Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic-Field Equations". *IEEE Transactions on Electromagnetic Compatibility*, EMC-23(4):377–382, 1981.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DIRECT FEM LARGE SCALE COMPUTATION OF TURBULENT MULTIPHASE FLOW IN URBAN WATER SYSTEMS AND MARINE ENERGY

Ezhilmathi Krishnasamy[1], **Johan Hoffman**[1 2], **and Johan Jansson**[1 2]

[1]BCAM - Basque Center for Applied Mathematics
Mazarredo 14, 48009 Bilbao, Spain
e-mail: ekrishnasamy@bcamath.org

[2] Department of High Performance Computing and Visualization KTH Royal Institute of Technology
SE-10044, Stockholm, Sweden
e-mail: {jjan,jhoffman}@kth.se

Keywords: Adaptive, FEM, Turbulence, Marine Engineering.

Abstract. *High-Reynolds number turbulent incompressible multiphase flow represents a large class of engineering problems of key relevance to society. Here we describe our work on modeling two such problems:*
1. The Consorcio de Aguas Bilbao Bizkaia is constructing a new storm tank system with an automatic cleaning system, based on periodically flushing tank water out in a tunnel.
2. In the framework of the collaboration between BCAM - Basque Center for Applied Mathematics and Tecnalia R & I, the interaction of the sea flow with a semi submersible floating offshore wind platform is computationally investigated. We study the MARIN benchmark modeling breaking waves over objects in marine environments.
Both of these problems are modeled in the the Direct FEM/General Galerkin methodology for turbulent incompressible variable-density flow [1, 2] in the FEniCS software framework.

1 INTRODUCTION

The world is facing increased climate instability and extreme weather phenomena, resulting in severe floods or drought. In the 20th century, ca. 1 million people were killed and 1.4 billion people were affected by floods [3]. Floods not only kill and affect people, they also affect the eco-systems, agriculture production, infrastructure and create economical instability. According to The International Disaster Database [4], from January 1975 to June 2002, flash floods (due to heavy rain) in Europe had 5.6 % mortality rate [3]. In Spain from 1900 to 2016, flash floods killed 987 people, 1350 people were affected and caused damage of 642 million USD [4]. It is predicted that the weather instabilities are going to worsen over the coming years [5].

Bilbao is located in the north of Spain and has Oceanic/Atlantic climate; its annual precipitation is 1200-2000 mm [6]. Consorcio de Aguas Bilbao Bizkaia (BWC) is constructing a new storm tank (detention tank) system with an automatic cleaning system, based on periodically flushing the tank out in a tunnel located in the Galindo area. This would prevent rain water flowing into the river and also minimize the hydraulic load on the existing sewer infrastructure. Later this water can be treated in the wastewater treatment plant (WWTP) in Galindo to produce potable water, or for other purposes.

The excessive water from the detention tank overflows through the tunnel, where sediments and floating objects might permanently settle on the bottom of the tunnel, which will eventually give rise to a foul smell and over time will also affect the downstream flow in the tunnel. It is not feasible to send human workers to clean the tunnel. Instead BWC wants to clean the tunnel with periodic flushing using water from the detention tank.

Our research involvement in the project has been to computationally model the water-air system, and predict the velocity, pressure and flow rate in the initial part of the tunnel. Where this velocity, pressure and flow rate values will be used as a input parameter for the shallow water modelling; and also velocity at the door section will be used to design the stronger gate to open a water from the detention tank.

We are also involved in research projects in Marine Renewable Energy (MRE) which involves similar mathematical models of water-air interaction. In the framework of the collaboration between the Basque Centre for Applied Mathematics (BCAM) and Tecnalia R&I, the interaction of the sea flow with a semisubmersible floating offshore wind platform is of key importance and high complexity.

We here take an initial step in the computational modeling in the phenomena of the platform, where we test our computational methodology on the established MARIN marine engineering benchmark.

We apply the Direct FEM/General Galerkin methodology [1, 2] to the primitive equations: variable-density incompressible Navier-Stokes, in the FEniCS-HPC software framework.

2 MATHEMATICAL MODEL

We model the problem by the primitive incompressible Navier-Stokes equations with variable density ρ:

$$
\begin{aligned}
R(\hat{u}) &= \begin{cases} \rho(\partial_t u + (u \cdot \nabla)u) + \nabla p - \nu \Delta u - \rho g = 0 \\ \partial_t \rho + (u \cdot \nabla)\rho = 0 \\ \nabla \cdot u = 0 \end{cases} \\
\hat{u} &= (u, p, \rho)
\end{aligned}
$$

By using a parameter-free stabilized finite element method (FEM) we are not introducing any explicit parametrization or modeling, aside from the slip model of the boundary layer, and we thus expect the simulations to be *predictive* if the mesh is fine enough to control the computational error.

2.1 Turbulent boundary layers

In our work on high Reynolds number turbulent flow [7, 8, 9, 10, 11], we have chosen to apply a skin friction stress as wall layer model. That is, we append the Navier-Stokes equations (NSE) with the following boundary conditions:

$$
\begin{aligned}
&u \cdot n = 0, \\
&\beta u \cdot \tau_k + n^T \sigma \tau_k = 0, \quad k = 1, 2,
\end{aligned}
$$

for $(x,t) \in \Gamma_{solid} \times I$, with $n = n(x)$ an outward unit normal vector, and $\tau_k = \tau_k(x)$ orthogonal unit tangent vectors of the solid boundary Γ_{solid}. We use matrix notation with all vectors v being column vectors and the corresponding row vector is denoted v^T.

With skin friction boundary conditions, the rate of kinetic energy dissipation in cG(1)cG(1) has a contribution of the form (for constant density):

$$
\sum_{k=1}^{2} \int_0^T \int_{\Gamma_{solid}} |\beta^{1/2} \bar{U} \cdot \tau_k|^2 \, ds \, dt,
$$

from the kinetic energy which is dissipated as friction in the boundary layer. For high Re, we model $Re \to \infty$ by $\beta \to 0$, so that the dissipative effect of the boundary layer vanishes with large Re. In particular, we have found that a small β does not influence the solution [9]. For the present simulations we used the approximation $\beta = 0$, which can be expected to be a good approximation for the high Reynolds numbers expected for the present setting.

3 COMPUTATIONAL METHODOLOGY AND SOFTWARE

The mathematical framework for the simulation method is functional analysis and the concept of weak solutions to the NSE, introduced by the mathematician Jean Leray in 1934. Leray proved that there exist weak solutions (or turbulent solutions in the terminology of Leray) that satisfy NSE in variational form, that is NSE integrated against a family of test functions.

A finite element method (FEM) is based on the variational form of NSE, and one can show that, if the formulation of the method satisfies certain conditions on stability and consistency, the approximate FEM solutions converge towards a weak solution of the NSE as the finite element mesh is refined [12]. We refer to such FEM as General Galerkin (G2) methods.

The test functions in G2 are defined over the mesh, and thus the finest scales of a G2 approximation are set by the mesh size. In contrast to RANS or LES (Large eddy simulation), no averaging operator or filter is applied to NSE, and thus no Reynolds or subgrid stresses that need modeling are introduced. Dissipation of turbulent kinetic energy in under-resolved parts of the flow is provided by the numerical stabilization of G2 in the form of a weighted least squares method based on the residual of NSE. Thus, the method is purely based on the NSE mathematical model, and no other modeling assumptions are made.

In G2, the mesh is adaptively constructed based on *a posteriori* estimation of the error in chosen goal or target functionals, such as drag and lift forces for example. Using duality in

a variational framework, *a posteriori* error estimates can be derived in terms of the residual, the mesh size, and the solution of a "dual" (or "adjoint") problem [13]. We initiate the adaptive mesh refinement algorithm from a coarse mesh, fine enough to capture the geometry, but without any further assumptions on the solution (i.e., no boundary layer meshes or *ad hoc* mesh design based on expected separation and wake structures are needed).

To model the effect of unresolved turbulent boundary layers we use a simple parametrization of the wall shear stress in terms of the skin friction [14]. In particular, for the very high Re for this problem we approximate the small skin friction by zero skin friction, which corresponds to a free slip boundary condition without boundary layer resolution.

This methodology is validated for a number of standard benchmark problems in the literature [15, 16, 17, 18], and in the following sections we describe the basic elements of the G2 method, also referred to as Adaptive DNS/LES, or simply Direct Finite element Simulation (DFS).

We have used a low order finite element discretization on unstructured tetrahedral meshes, which we refer to as cG(1)cG(1), i.e., continuous piecewise linear approximations in space and time. In this project we have not applied adaptive mesh refinement, since it currently only is functioning robustly for production simulations for constant density.

3.1 Direct FEM cG(1)cG(1) for variable-density

In a cG(1)cG(1) method [12] we seek an approximate space-time solution $\hat{U} = (D, U, P)$ (with D the discrete density ρ) which is continuous piecewise linear in space and time (equivalent to the implicit Crank-Nicolson method). With I a time interval with subintervals $I_n = (t_{n-1}, t_n)$, W^n a standard spatial finite element space of continuous piecewise linear functions, and W_0^n the functions in W^n which are zero on the boundary Γ, the cG(1)cG(1) method for variable-density incompressible flow with homogeneous Dirichlet boundary conditions for the velocity takes the form: for $n = 1, ..., N$, find $(D^n, U^n, P^n) \equiv (D(t_n), U(t_n), P(t_n))$ with $D^n \in W^n$, $U^n \in V_0^n \equiv [W_0^n]^3$ and $P^n \in W^n$, such that

$$
\begin{aligned}
r(\hat{U}, \hat{v}) &= (D((U^n - U^{n-1})k_n^{-1} + (\bar{U}^n \cdot \nabla)\bar{U}^n), v) + (2\nu\epsilon(\bar{U}^n), \epsilon(v)) \\
&- (P, \nabla \cdot v) - (Dg, v) + (\nabla \cdot \bar{U}^n, q) + (D^n - D^{n-1})k_n^{-1} + (\bar{U}^n \cdot \nabla)\bar{D}^n), v) \\
&+ LS(D^n, \bar{U}^n, P^n) + SC(D^n, \bar{U}^n, P^n) = 0, \ \forall \hat{v} = (z, v, q) \in W^n \times V_0^n \times W^n
\end{aligned} \tag{1}
$$

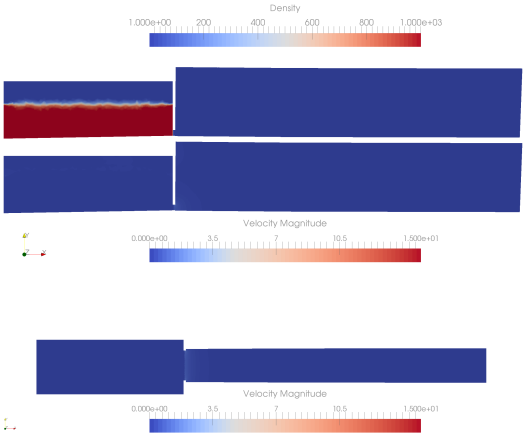
where $\bar{U}^n = 1/2(U^n + U^{n-1})$ is piecewise constant in time over I_n and LS and SC are least-squares and shock-capturing stabilizing term described in [12].

Here we also add an experimental phase separation term of the form $(\rho_{air} - \rho)(\rho_{water} - \rho)$. The term has only limited effect for short times. We will investigate the sensitivity of the parametrization of this term in future work for long-time simulations.

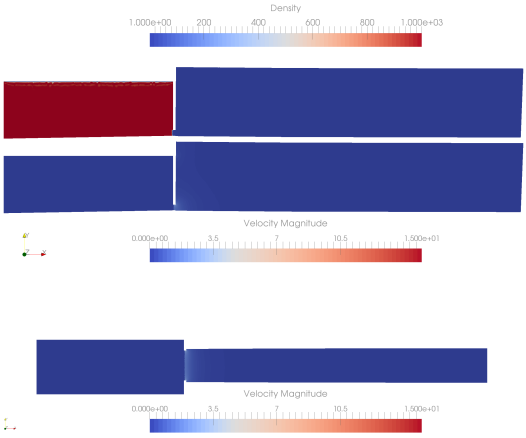
3.2 The FEniCS-HPC finite element computational framework

The simulations in this paper have been computed using the Unicorn solver in the FEniCS-HPC automated FEM software framework.

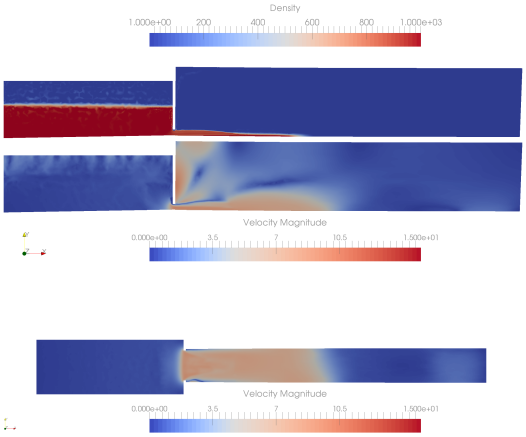
FEniCS-HPC is an open source framework for automated solution of PDE on massively parallel architectures, providing automated evaluation of variational forms given a high-level description in mathematical notation, duality-based adaptive error control, implicit parameter-free turbulence modeling by use of stabilized FEM and strong linear scaling up to thousands of cores [19, 20, 21, 22, 23, 24]. FEniCS-HPC is a branch of the FEniCS [25, 26] framework focusing on high performance on massively parallel computer architectures.



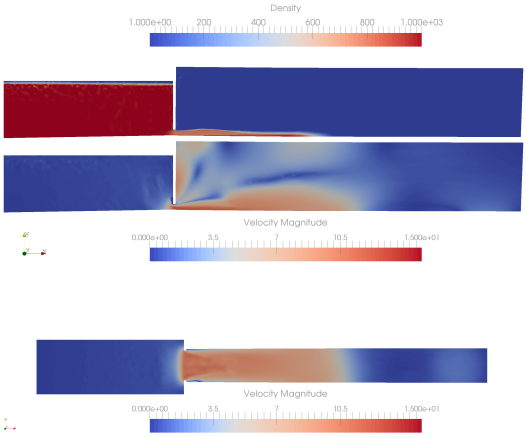
(a) Density and velocity x-y and x-z slice T_D={5s,10s}, H=6m, t=0.



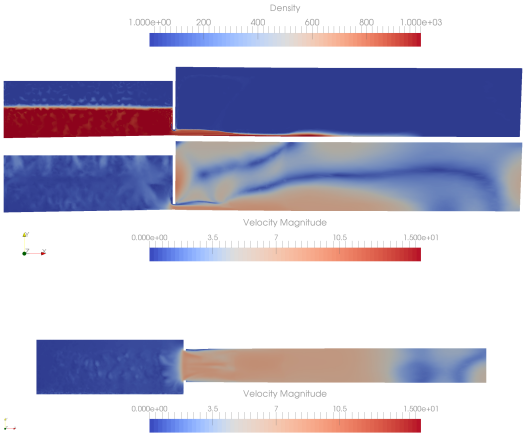
(d) Density and velocity x-y and x-z slice T_D={5s,10s}, H=10m, t=0.



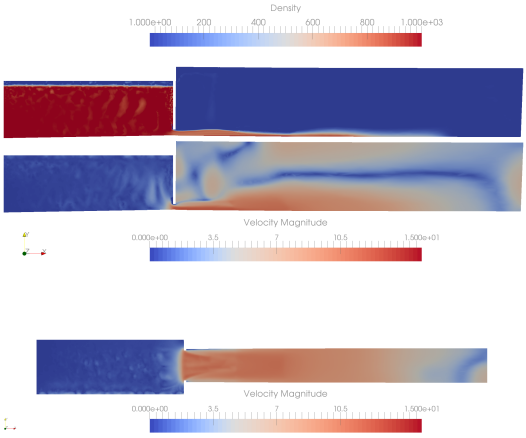
(b) Density and velocity x-y and x-z slice T_D=5s, H=6m, t=5s.



(e) Density and velocity x-y and x-z slice T_D=5s, H=10m, t=5s.



(c) Density and velocity x-y and x-z slice T_D=10s, H=6m, t=10s.

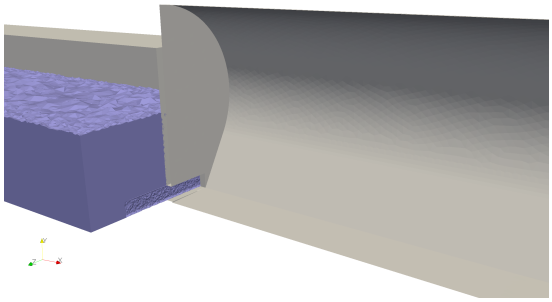


(f) Density and velocity x-y and x-z slice T_D=10s, H=10m, t=10s.

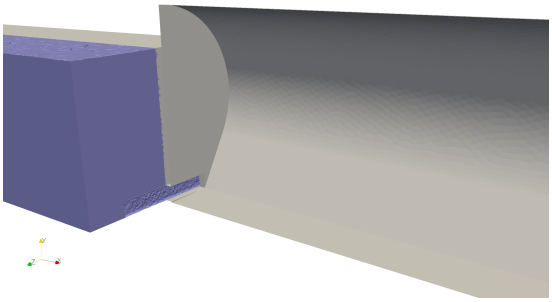
Figure 6: Density and velocity x-y and x-z at different heights with different door opening times.

In this section we plot slice plots of the density (showing the evolution of the water surface), the velocity, 3D plots of the isovolume of the density (showing the evolution of the water surface). Additionally we plot the flow rate through the door over time, and the average velocity in the door section and in the first 10m section of the tunnel.

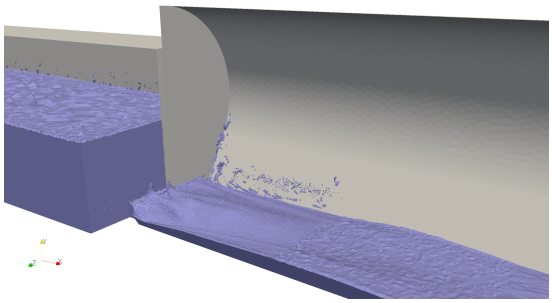
Figure 6 shows the density and velocity at different times and height settings. As we can see in figure 6 (b) and (e), the turbulent mixing is slightly higher for figure 6 (e); and the velocity at the bottom of the tunnel is also higher in both cases (T_D={5s and 10s}) for H=10m.



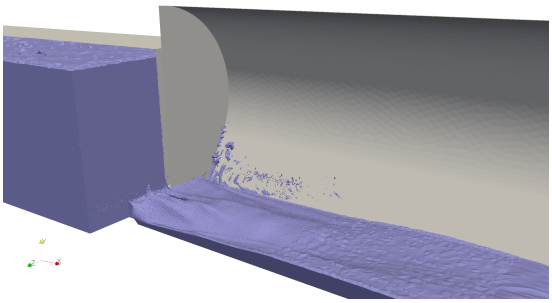
(a) Water isovolume T_D={5s,10s}, H=6m, t=0s



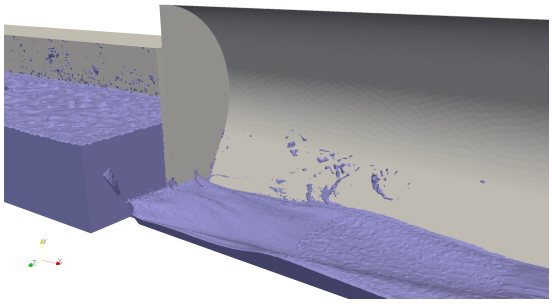
(d) Water isovolume T_D={5s,10s}, H=10m, t=0s



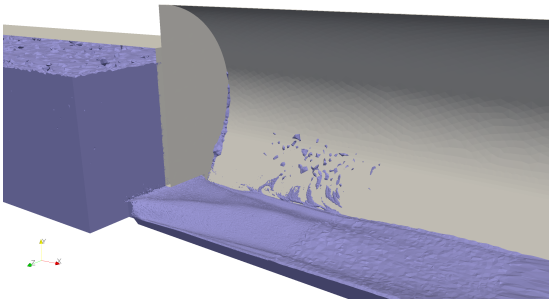
(b) Water isovolume T_D=5s, H=6m, t=5s



(e) Water isovolume T_D=5s, H=10m, t=5s



(c) Water isovolume T_D=10s, H=6m, t=10s



(f) Water isovolume T_D=10s, H=10m, t=10s

Figure 7: Water isovolume at different heights with different door opening times.

Water isovolume in the simulation can be seen in figure 7, where we can see water splashes closer to the door as the gate opens; it is due to pressure difference between tank and tunnel [29] and variable density model which is mentioned in section 3.1. Figure 8 shows the flow rate through the gate. As we can see here, the gate opening time and water height do have a

influence over the water flow rate through the opening gate.

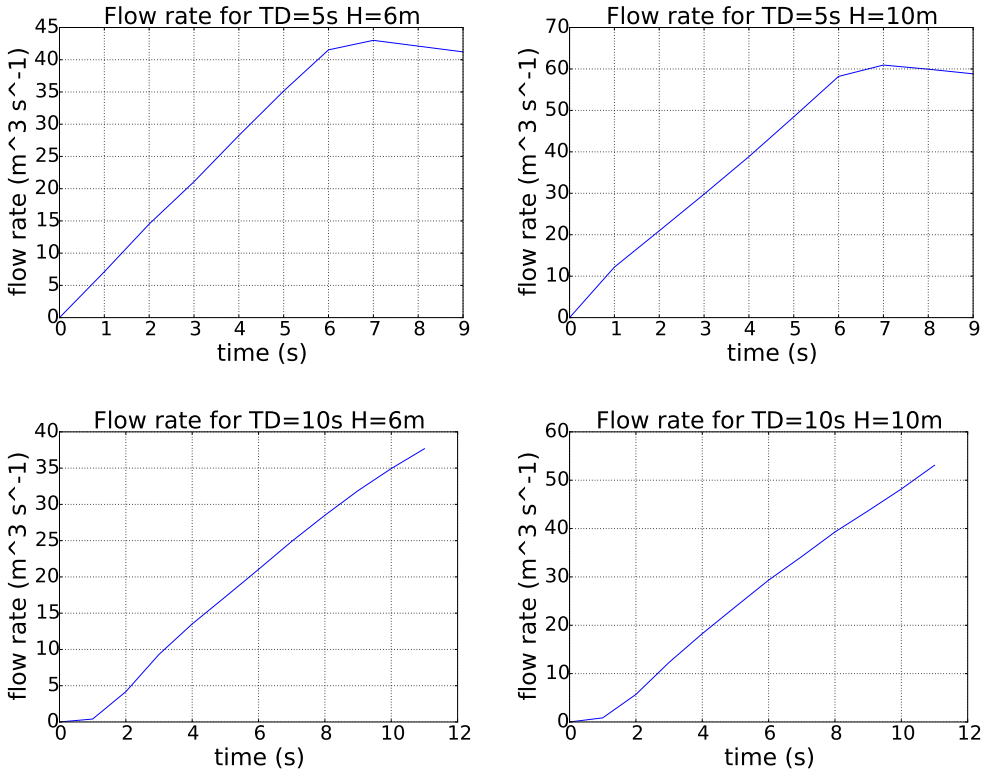


Figure 8: "Spending" flow rate through the door.

Figure 9 shows the average velocity at the door section. It is clear that the velocity is higher at the door section for H=10m compared to H=6m (for the same opening time T_D=5s).

The average velocity at the downstream side of the tunnel is plotted in figure 10. It shows that for T_D=5s and H=6m, the velocity increases gradually up to $6\,\mathrm{m\,s^{-1}}$ until $t = 4s$, after that the velocity goes down. On the other hand, for T_D=5s and H=10m, the velocity is higher, up to ca. $8\,\mathrm{m\,s^{-1}}$. This behaviour is similar to the case of T_D=10s. Thus we can conclude that the water height influences the average velocity at the bottom of the tunnel.

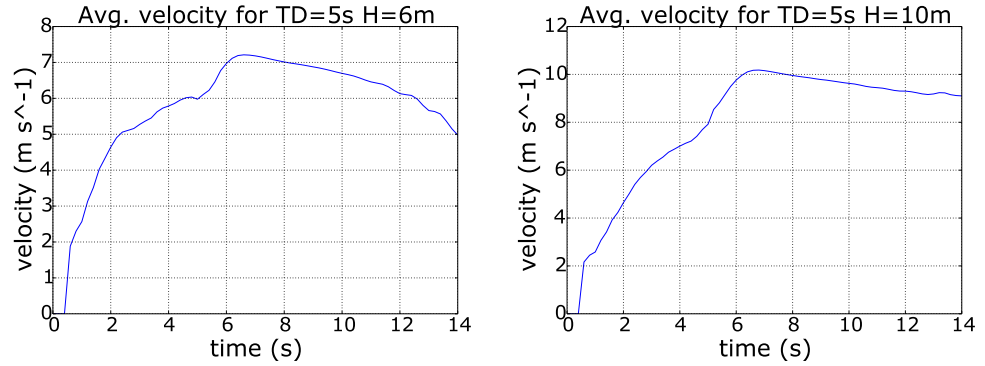


Figure 9: Average x-velocity in the door section.

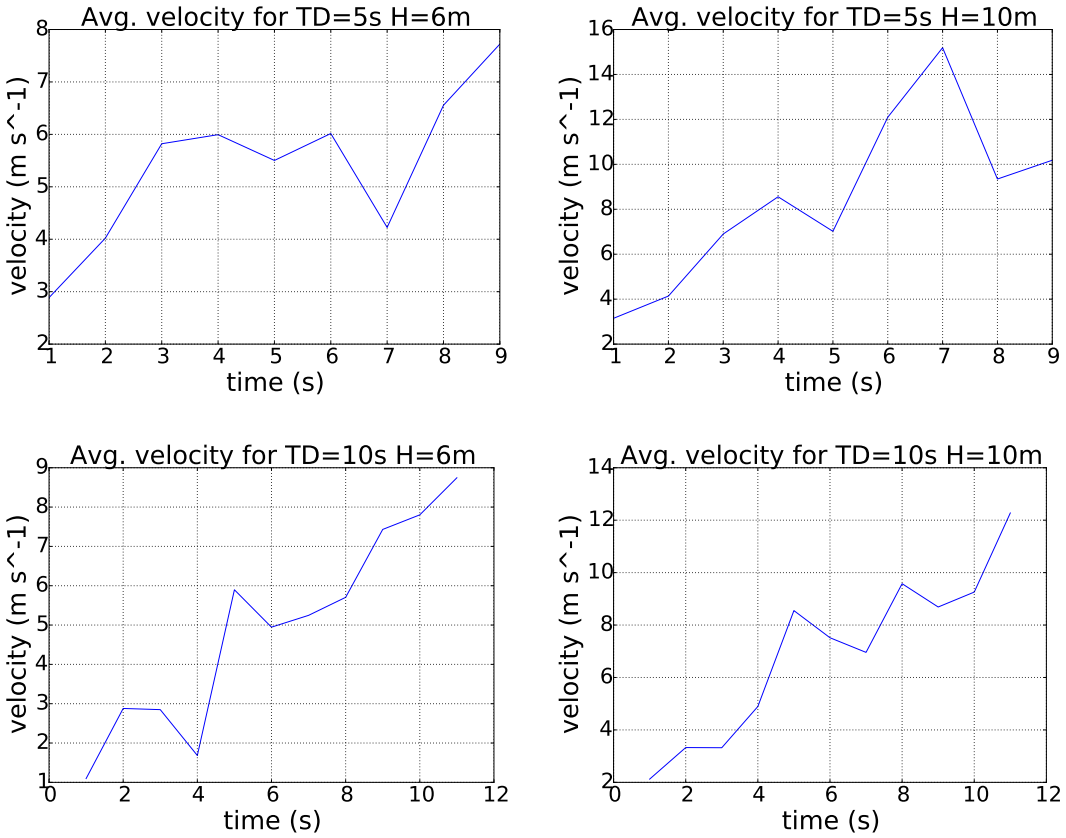


Figure 10: Average flushing x-velocity in the first 10m-section of the tunnel.

6 CONCLUSIONS

In this paper we provide computational results for Direct FEM simulations of the primitive 3D variable-density incompressible Navier-Stokes equations. We investigate the MARIN marine engineering benchmark and a storm drain system.

The simulation results for the MARIN benchmark qualitatively capture the evolution of the pressure signals in the sensors. The discrepancies are quantitatively of approximately similar magnitudes as the state of the art published in the literature, e.g. in [30, 28].

For the storm drain problem, the density and velocity fields have a 3D structure, a triangular jet shape, at the exit of the door. The door opening time does not appear to have a large influence on the structure or magnitude of the velocity. The water height in the tank has a significant influence on the magnitude of the velocity in the flushing section at the beginning of the tunnel.

In future work we would like to study the mesh sensitivity of the method and implement a robust adaptive method, similar to what we have developed for constant density turbulent flow, to enable adaptive error control.

REFERENCES

[1] J Hoffman, J Jansson, RV de Abreu, NC Degirmenci, Niclas Jansson, Kaspar Mller, Murtazo Nazarov, JHiromi Sphler, Unicorn: Parallel adaptive finite element simulation of tur-

bulent flow and fluid-structure interaction for deforming domains and complex geometry. *Computers & Fluids*, 2013.

[2] J Hoffman, J Jansson, N Jansson, FEniCS-HPC: Automated predictive high-performance finite element computing with applications in aerodynamics, PPAM 2015. *Lecture Notes in Computer Science*, 2015.

[3] Jonkman, Sebastiaan N. Global perspectives on loss of human life caused by floods , Springer, 2005.

[4] The International Disaster Database. *http://www.emdat.be/country_profile/index.html*.

[5] Füssel, Hans-Martin and Jol, André and others. *Climate change, impacts and vulnerability in Europe 2012 an indicator-based report*, Luxembourg: Publications Office of the European Union, 2012.

[6] Ezcurra, A and Areitio, J and Herrero, I. Relationships between cloud-to-ground lightning and surface rainfall during 1992–1996 in the Spanish Basque Country area. *Atmospheric research*, Elsevier, 2002.

[7] Johan Hoffman, Johan Jansson, Niclas Jansson, and Rodrigo Vilela De Abreu. To- wards a parameter-free method for high reynolds number turbulent flow simulation based on adaptive finite element approximation. *Computer Methods in Applied Mechanics and Engineering*, 288(0):60-74, 2015. Error Estimation and Adaptivity for Nonlinear and Time-Dependent Problems.

[8] Johan Hoffman, Johan Jansson, and Claes Johnson. New theory of flight. *Journal of Mathematical Fluid Mechanics*, 2015.

[9] Johan Hoffman and Niclas Jansson. *A computational study of turbulent flow separation for a circular cylinder using skin friction boundary conditions*. Ercoftac, series Vol.16, Springer, 2010.

[10] Johan Hoffman and Claes Johnson. *Resolution of dalemberts paradox. J. Math. Fluid Mech*., Published Online First at www.springerlink.com: 10 December 2008.

[11] Rodrigo Vilela de Abreu, Niclas Jansson, and Johan Hoffman. Adaptive computation of aeroacoustic sources for a rudimentary landing gear. *Int. J. Numer. Meth. Fluids*, 74(6):406-421, 2014.

[12] Johan Hoffman and Claes Johnson. *Computational Turbulent Incompressible Flow: Applied Mathematics Body and Soul Vol 4*. Springer-Verlag Publishing, 2006.

[13] Johan Hoffman and Claes Johnson. *Computational Turbulent Incompressible Flow, volume 4 of Applied Mathematics: Body and Soul*. Springer, 2007.

[14] U. Schumann. Subgrid-scale model for finite difference simulation of turbulent flows in plane channels and annuli. *J. Comput. Phys*., 18:376-404, 1975.

[15] Johan Hoffman. Computation of mean drag for bluff body problems using adaptive DNS/LES. *SIAM J. Sci. Comput*., 27(1):184-207, 2005.

[16] Johan Hoffman. Adaptive simulation of the turbulent flow past a sphere. *J. Fluid Mech*., 568:77-88, 2006.

[17] Johan Hoffman. Efficient computation of mean drag for the subcritical flow past a circular cylinder using general galerkin g2. Int. *J. Numer. Meth. Fluids*, 59(11):1241-1258, 2009.

[18] Johan Hoffman and Claes Johnson. A new approach to computational turbulence modeling. *Comput. Methods Appl. Mech. Engrg*., 195:2865-2880, 2006.

[19] Johan Hoffman, Johan Jansson, Rodrigo Vilela de Abreu, Niyazi Cem Degirmenci, Niclas Jansson, Kaspar Mller, Murtazo Nazarov, and Jeannette Hiromi Sphler. Unicorn: Parallel adaptive finite element simulation of turbulent flow and fluid- structure interaction for deforming domains and complex geometry. *Computers and Fluids*, 2012.

[20] Johan Hoffman, Johan Jansson, Cem Degirmenci, Niclas Jansson, and Murtazo Nazarov. Unicorn: a Unified Continuum Mechanics Solver, chapter 18. Springer, 2012.

[21] Niclas Jansson, Johan Hoffman, and Johan Jansson. Framework for Massively Parallel Adaptive Finite Element Computational Fluid Dynamics on Tetrahedral Meshes. *SIAM J. Sci. Comput*., 34(1):C24-C41, 2012.

[22] Robert C. Kirby. *FIAT: Numerical Construction of Finite Element Basis Functions*, chapter 13. Springer, 2012.

[23] Johan Hoffman, Johan Jansson, Niclas Jansson, C. Johnson, and Rodrigo V. de Abreu. Turbulent flow and fluid-structure interaction. In *Automated Solutions of Differential Equations by the Finite Element Method*. Springer, 2011.

[24] Johan Hoffman, Johan Jansson, Niclas Jansson, and Murtazo Nazarov. Unicorn: A unified continuum mechanics solver. In *Automated Solutions of Differential Equations by the Finite Element Method*. Springer, 2011.

[25] FEniCS. Fenics project. *http://www.fenicsproject.org*, 2003.

[26] Anders Logg, Kent-Andre Mardal, Garth N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.

[27] J. Hoffman, J. Jansson, and M. Stckli. Unified continuum modeling of fluid-structure interaction. *Mathematical Models and Methods in Applied Sciences*, 2011.

[28] K.M.T. Kleefsman, G. Fekken, A.E.P. Veldman, B. Iwanowski and B. Buchner: A Volume-of-Fluid based simulation method for wave impact problems. *J. Comp. Phys*, 206, (2005), 363-393.

[29] Xu, Lei and Zhang, Wendy W and Nagel, Sidney R. Drop splashing on a dry smooth surface. *Physical review letters*, APS, 94, 2005.

[30] Kees, C. E., Akkerman, I., Farthing, M. W., Bazilevs, Y. (2011). A conservative level set method suitable for variable-order approximations and unstructured meshes. *Journal of Computational Physics*, 230(12), 4536-4558.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

KINETIC MODELS AND ALGORITHMS FOR SOLUTION OF THE MAGNETOGASDYNAMIC PROBLEMS ON THE MODERN SUPERCOMPUTING SYSTEMS

B. Chetverushkin[1], N. D'Ascenzo[1,2], A. Saveliev[3] and V. Saveliev[1,4]

[1] Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Russia
e-mail: office@keldysh.ru

[2] Deutsche Electronen Synchrotron, Hamburg, Germany
e-mail: ndasc@mail.desy.de

[3] Hamburg University, Hamburg, Germany
e-mail: andrey.saveliev@desy.de

[4] Immanuel Kant Baltic Federal University, Kaliningrad, Russia
e-mail: saveliev@mail.desy.de

Keywords: kinetic schemes, magneto gas dynamics, high performance computing

Abstract. *The impressive progress of the kinetic schemes in the solution of gas dynamics problems and the development of effective parallel algorithms for modern high performance parallel computing systems led to the development of advanced methods for the solution of the magnetohydrodynamics problem in the important area of plasma physics. The novel feature of the method is the formulation of the complex Boltzmann-like distribution function of kinetic method with the implementation of electromagnetic interaction terms. The numerical method is based on the explicit schemes. Due to logical simplicity and its efficiency, the algorithm is easily adapted to modern high performance parallel computer systems including hybrid computing systems with graphic processors.*

1 INTRODUCTION

The tremendous progress in the development of high performance computing systems, especially expecting drastically new exascale computing systems, including the challenges in architecture, scale, power and reliability, gives new opportunities for the mathematical modelling of important physical phenomena in the present and future. Nevertheless the complexity of the challenges in science and engineering continues to outpace our ability to adequately address them through impressively growing computational power.

A feature of the present is that the development of technologies and computer systems architecture are well ahead of software development. The software problems are primarily associated with the complexity of the algorithms adaptation for the differential equations of mathematical physics to high performance computing systems architecture. In particular they refer to one of the important requirements as the accuracy in combination with the correctness of the initial mathematical models. Another requirement for the methods is their logical simplicity and high efficiency at the same time. The numerical algorithms should be simple and transparent from a logical point of view.

One of the important directions to overcome these problems is the development of a nontraditional approach to initial mathematical models and computational algorithms. In the present study for the solution of the multidimensional gas dynamics and magnetohydrodynamics problems kinetic difference scheme is proposed. It is convenient from the physics point of view to define the gas dynamics and magnetohydrodynamics quantities from close relations between the kinetic and gas dynamics description of physics processes [1, 2, 3, 4, 5, 6].

Another aspect is the study of the explicit finite difference schemes, which seem to be preferable for future high performance parallel computing, especially in terms of their simplicity and well adaptability to parallel program realization, including hybrid high performance parallel computing systems. The weakness of explicit schemes is a strictly limited time step that ensures computational stability. This restriction becomes critical with the growing number of nodes and the reduction in the step of a spatial mesh. The advanced explicit kinetic finite difference schemes have a soft stability condition giving the opportunity to enhance the stability and to use very fine meshes [7].

The mentioned aspects are used for the development of the framework for the study of the dynamics of the conducting gas media in strong magnetic fields at high performance parallel computing systems.

2 THEORETICAL ISSUES

2.1 Gas Dynamics Processes

The kinetic theory describes the gas dynamics by the Boltzmann differential equation through the evolution of the distribution function $f(\mathbf{x}, \boldsymbol{\xi}, t)$ [8]:

$$\frac{\partial f(\mathbf{x}, \boldsymbol{\xi}, t)}{\partial t} + \boldsymbol{\xi} \cdot \nabla f(\mathbf{x}, \boldsymbol{\xi}, t) = C(f) \tag{1}$$

where $C(f)$ is a nonlinear integral operator which describes the collisions between gas molecules.

This evolution equation follows naturally from the relations between the kinetic and the gas dynamics description of continuous media. The macroscopic observables such as density, momentum, energy flux as a function of $\mathbf{x}$ and t are obtained from the moments of the distribution

function with respect to the macroscopic velocity. The evolution equations for these gas dynamics quantities are obtained by integrating Eq. (1) over molecular velocities $\boldsymbol{\xi}$ with summational invariants m,$m\boldsymbol{\xi}$,$\frac{1}{2}m\boldsymbol{\xi}^2$. The computational interest in kinetic formulations of the gas dynamics is high due to the linearity of the differential operator on the left side of Eq. (1). Nonlinearity is confined by the collision term, which is generally local in $\mathbf{x}$ and t.

An important feature is that the collision integral vanishes in the equilibrium state when the local Boltzmann distribution function f is a Maxwellian:

$$f(\mathbf{x},\boldsymbol{\xi},t) = \frac{\rho(\mathbf{x},t)m^{1/2}}{\left(2\pi kT\left(\mathbf{x},t\right)\right)^{3/2}} exp\left\{-\frac{m}{2kT(t,\mathbf{x})}(\boldsymbol{\xi}-\mathbf{u}(t,\mathbf{x}))^2\right\} \tag{2}$$

This leads to the use of this model for numerical methods and possible generalisations in order provide a natural kinetic description of the system of conservation laws. This approximation is sufficient for the gas dynamics processes and is called the kinetic approach [8].

2.2 Magneto Gas Dynamic Processes

In [9] it was shown that electromagnetic fields do not destroy the validity of the Boltzmann equation and this opened the way for the implementation of the electromagnetic term in the Boltzmann-like distribution function. From the vector nature of the electromagnetic interaction, the distribution function should taking into account the vector behaviour and provide correct formulation for the evolution of the magnetic field, i. e. the magnetic field should be generally defined as the momentum of the Boltzmann-like distribution function.

A few useful attempts to formulate the vector Boltzmann-like distribution function can be found in [10, 11, 12], but physical meaning was not clearly defined.

We propose an evaluation of the electromagnetic processes in the context of the distribution function, taking into account the axial nature of the magnetic field. The electromagnetic interaction processes contributions are considered as a complex velocities vector, described by kinetic motion of charged particles in magnetic field [13]. Using the above definitions we define the local complex Maxwellian distribution function of magnetohydrodynamics with drift velocity $\mathbf{u}$ in magnetic field $\mathbf{B}$ at the equilibrium:

$$f_M = \frac{\rho m^{1/2}}{(2\pi kT)^{3/2}} \exp\left\{-\frac{m}{2kT}\left|(\boldsymbol{\xi}-\mathbf{u}) - i\frac{\mathbf{B}}{\sqrt{\mu_m\rho}}\right|^2\right\}, \tag{3}$$

The first term on the right-hand side of (3) includes the kinetic energy and the second term is contribution from electromagnetic interactions. The hydrodynamics observables are real scalars and vectors. The complex components include the dynamics of the macroscopic observables introduced by the evolution of the magnetic field, keeping their specific pseudo-vectorial nature.

The hydrodynamic and electromagnetic observables are obtained respectively as the real and imaginary part of the integral of the distribution function (3) with the summational invariants $\left(m, m\boldsymbol{\xi}, \frac{1}{2}m\boldsymbol{\xi}^2, m\boldsymbol{\xi}^*\right)$ along the line L parallel to the real axis and shifted by $i\mathbf{B}/\sqrt{\mu\rho}$ in the

imaginary axis direction.

$$\rho = \int_L m f_M d^3\boldsymbol{\xi} \tag{4}$$

$$\mathbf{u} = \frac{1}{\rho}\int_L m\boldsymbol{\xi} f_M d^3\boldsymbol{\xi} \tag{5}$$

$$E = \int_L \frac{1}{2} m\boldsymbol{\xi}^2 f_M d^3\boldsymbol{\xi} \tag{6}$$

$$\mathbf{B} = -\sqrt{\frac{\mu_m}{\rho}}\int_L m\boldsymbol{\xi}^* f_M d^3\boldsymbol{\xi} \tag{7}$$

For the compressible media, we use the linear approximation of the magnetic permeability:

$$\mu_m = \mu_\rho \rho \tag{8}$$

where μ_ρ is a constant representing the magnetic permeability per unit density.

The proposed complex Boltzmann Maxwell like distribution function contains the hydrodynamics terms and the electromagnetic terms. Thus by using this distribution function to calculate the mass, momentum, energy and magnetic field fluxes, most of the electromagnetic contributions are calculated directly, i.e. one does not have to solve the hydrodynamics and magnetic force components separately or differently, as will be shown below.

3 THE IDEAL MAGNETO GAS DYNAMICS EQUATIONS

To provide the first step of the formulation of the MHD conservation laws equation, the equilibrium state is considered with the proposed distribution function. The MHD system of equations is obtained by the integration of (1) with vanishing collision integral with the summational invariants following the definition in (7):

$$\int_L m\frac{\partial f}{\partial t} + \int_L m\boldsymbol{\xi}\cdot\nabla f d^3\boldsymbol{\xi} = 0 \tag{9}$$

$$\int_L m\boldsymbol{\xi}\frac{\partial f}{\partial t} + \int_L m\boldsymbol{\xi}\boldsymbol{\xi}\cdot\nabla f d^3\boldsymbol{\xi} = 0 \tag{10}$$

$$\int_L \frac{1}{2} m\boldsymbol{\xi}^2\frac{\partial f}{\partial t} + \int_L \frac{1}{2} m\boldsymbol{\xi}^2\boldsymbol{\xi}\cdot\nabla f d^3\boldsymbol{\xi} = 0 \tag{11}$$

$$\int_L m\boldsymbol{\xi}^*\frac{\partial f}{\partial t} + \int_L m\boldsymbol{\xi}^*\boldsymbol{\xi}\cdot\nabla f d^3\boldsymbol{\xi} = 0 \tag{12}$$

The result obtained, set of Eq. (12), is the ideal magnetohydrodynamics system of equations:

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x_i}\rho u_i = 0 \tag{13}$$

$$\frac{\partial}{\partial t}\rho u_i + \frac{\partial}{\partial x_k}\left[\left(p + \frac{B^2}{2\mu_M}\right)\delta_{ik} + \rho u_i u_k - \frac{B_i B_k}{\mu_M}\right] = 0 \tag{14}$$

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x_i}\left[u_i\left(E + p + \frac{B^2}{2\mu_M}\right) - \frac{B_i u_k B_k}{\mu_M}\right] = 0 \tag{15}$$

$$\frac{\partial B_i}{\partial t} + \frac{\partial}{\partial x_k}\left[u_k B_i - u_i B_k\right] = 0 \tag{16}$$

In addition an equation for $\nabla \cdot \mathbf{B}$ is obtained as the imaginary part of the integral of the summational invariant (m) respect to the velocities $\boldsymbol{\xi}$:

$$\int_L m\frac{\partial f}{\partial t} + \int_L m\boldsymbol{\xi}\cdot\nabla f d^3\boldsymbol{\xi} = 0 \tag{17}$$

$$\frac{\partial B_i}{\partial x_i} = 0 \tag{18}$$

4 THE QUASI MAGNETO GAS DYNAMIC SYSTEM OF EQUATIONS

The quasi MHD system of equations for the real gases is obtained using the integration technique outlined in section 2 and the balance relation based on the Taylor expansion of the variation of the local distribution function in spatial variables up to third order infinitesimal [2]:

$$\frac{f^{j+1}-f^j}{\Delta t} + \xi_i\frac{\partial f^j}{\partial x_i} = \xi_i\xi_k\frac{\partial}{\partial x_i}\frac{\tau}{2}\frac{\partial f^j}{\partial x_k} \tag{19}$$

The evaluation equation for the magneto gas dynamic variables (quasi magneto gas dynamic equations) obtained as:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}\rho u_i = \frac{\partial}{\partial x_i}\left(\frac{\tau}{2}\frac{\partial}{\partial x_k}\Pi_{ik}\right) \tag{20}$$

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial}{\partial x_k}\Pi_{ik} = \frac{\partial}{\partial x_k}\Pi^D_{ik} + \frac{\partial}{\partial x_k}\left[\left(\frac{\tau}{2}\frac{\partial}{\partial x_k}\Pi_{ik}\right)u_k\right] \tag{21}$$

$$\frac{\partial E}{\partial t} + \frac{\partial F_i}{\partial x_i} = \frac{\partial Q_i}{\partial x_i} + \frac{\partial}{\partial x_i}\Pi^D_{ik}u_k + \\ + \frac{\partial}{\partial x_i}\left[\left(\frac{E+p}{\rho} + \frac{1}{\mu_m}\frac{B^2}{2\rho}\right)\frac{\tau}{2}\frac{\partial}{\partial x_k}\Pi_{ik}\right] \tag{22}$$

$$\frac{\partial B_i}{\partial t} + \frac{\partial}{\partial x_k}M^B_{ik} = \frac{\partial}{\partial x_k}\Pi^{DB}_{ik} \tag{23}$$

The right-hand of the kinetic magneto gas dynamic Eq. (20-23) contains dissipative terms. In comparison with other methods, the dissipative terms are obtained not by phenomenology with some assumptions about magneto gas dynamics processes but in consistency with the difference scheme of the Boltzmann equation.

Π_{ik} is the momentum flux density tensor for a perfect gas in magnetic field:

$$\Pi_{ik} = \left(p + \frac{B^2}{2\mu_m}\right)\delta_{ik} + \rho u_i u_k - \frac{1}{\mu_m}B_iB_k \tag{24}$$

F_i is the heat transfer flux of a perfect gas in magnetic field:

$$F_i = \left[\left(E + p + \frac{B^2}{2\mu_m}\right)u_i - \frac{1}{\mu_m}B_iu_kB_k\right] \tag{25}$$

M^B_{ik} is the asymmetric product between velocity $\mathbf{u}$ and magnetic field flux $\mathbf{B}$:

$$M^B_{ik} = u_kB_i - u_iB_k \tag{26}$$

The dissipative terms:

$$
\begin{aligned}
\Pi^D_{ik} =& \frac{\tau}{2}\left(p\frac{\partial u_i}{\partial x_k} + p\frac{\partial u_k}{\partial x_i} - \frac{2}{3}p\frac{\partial u_l}{\partial x_l}\delta_{ik}\right) \\
&+ \frac{\tau}{2}\left[\left(\frac{1}{\mu_m}\frac{B^2}{2}\delta_{lk} - \frac{1}{\mu_m}B_lB_k\right)\frac{\partial u_i}{\partial x_l} + \left(\frac{1}{\mu_m}\frac{B^2}{2}\delta_{il} - \frac{1}{\mu_m}B_iB_l\right)\frac{\partial u_k}{\partial x_l}\right. \\
&\left. - \left(\frac{1}{\mu_m}\frac{B^2}{2}\delta_{ik} - \frac{1}{\mu_m}B_iB_k\right)\frac{\partial u_l}{\partial x_l}\right] \\
&+ \frac{\tau}{2}\left[\frac{1}{\mu_m}B_l\left(-B_k\frac{\partial u_i}{\partial x_l} - B_i\frac{\partial u_k}{\partial x_l} + B_n\frac{\partial u_n}{\partial x_l}\delta_{ik}\right)\right] \\
&+ \frac{\tau}{2}\left[\rho u_iu_l\frac{\partial u_k}{\partial x_l} + u_i\frac{\partial p}{\partial x_k} + u_i\frac{\partial}{\partial x_k}\frac{1}{\mu_m}\frac{B^2}{2} - u_i\frac{\partial}{\partial x_l}\frac{1}{\mu_m}B_lB_k\right] \\
&+ \frac{\tau}{2}\left[u_l\frac{\partial p}{\partial x_l} + \gamma p\frac{\partial u_l}{\partial x_l}\right]\delta_{ik} + \frac{\tau}{2}\left[\frac{1}{\mu_m}B_n^2\frac{\partial u_l}{\partial x_l} - \frac{B_nB_l}{\mu_m}\frac{\partial u_n}{\partial x_l}\right. \\
&\left. + \frac{1}{\mu_m}B_nu_l\frac{\partial B_n}{\partial x_l}\right]\delta_{ik} + \frac{\tau}{2}\left[-\frac{B_iB_k}{\mu_m}\frac{\partial u_l}{\partial x_l} + \frac{B_iB_l}{\mu_m}\frac{\partial u_k}{\partial x_l} - \frac{B_i}{\mu_m}u_l\frac{\partial B_k}{\partial x_l}\right] \\
&+ \frac{\tau}{2}\left[-\frac{B_kB_i}{\mu_m}\frac{\partial u_l}{\partial x_l} + \frac{B_kB_l}{\mu_m}\frac{\partial u_i}{\partial x_l} - \frac{B_k}{\mu_m}u_l\frac{\partial B_i}{\partial x_l}\right]
\end{aligned}
\tag{27}
$$

$$
\begin{aligned}
Q^D_i =& \frac{\tau}{2}\left(\frac{5}{2}p\frac{\partial}{\partial x_i}\frac{p}{\rho}\right) \\
&+ \frac{\tau}{2}\left[\frac{5}{2}\left(\frac{B^2}{2\mu_m}\delta_{ik} - \frac{B_iB_k}{\mu_m}\right)\frac{\partial}{\partial x_k}\frac{p}{\rho}\right] \\
&+ \frac{\tau}{2}\left[\frac{3}{2}\left(p\delta_{ik} + \frac{B^2}{2\mu_m}\delta_{ik} - \frac{B_iB_k}{\mu_m}\right)\frac{\partial}{\partial x_k}\frac{B^2}{2\mu_m\rho} - \left(p + \frac{B^2}{2\mu_m}\right)\frac{\partial}{\partial x_k}\frac{B_iB_k}{\mu_m\rho}\right. \\
&\left. - \frac{B_iB_k}{\mu_m\rho}\frac{\partial}{\partial x_k}\frac{B^2}{2\mu_m}\right] + \frac{\tau}{2}\left[\rho u_iu_k\frac{\partial}{\partial x_k}\frac{3}{2}\frac{p}{\rho}\right] \\
&+ \frac{\tau}{2}\left[\rho u_iu_k\left(p + \frac{B^2}{\mu_m}\right)\frac{\partial}{\partial x_k}\frac{1}{\rho} - u_i\frac{B^2}{\mu_m}\frac{\partial u_k}{\partial x_k}\right] \\
&+ \frac{\tau}{2}\left[u_i\frac{B_l}{\mu_m}\left(B_l\frac{\partial u_k}{\partial x_k} - B_k\frac{\partial u_l}{\partial x_k} + u_k\frac{\partial B_l}{\partial x_k}\right)\right] \\
&+ \frac{\tau}{2}\left[\frac{1}{2}\rho u_iu_k\left(\frac{B^2}{2\mu_m}\frac{\partial}{\partial x_k}\frac{1}{\rho} - \frac{1}{\rho}\frac{\partial}{\partial x_k}\frac{B^2}{2\mu_m}\right)\right] \\
&+ \frac{\tau}{2}\left[B_iB_l\left(-u_k\frac{\partial u_l}{\partial x_k} - \frac{1}{\rho}\frac{\partial p}{\partial x_l} - \frac{1}{\rho}\frac{\partial}{\partial x_l}\frac{B^2}{2\mu_m} + \frac{1}{\rho}\frac{\partial}{\partial x_k}\frac{B_lB_k}{\mu_m}\right)\right]
\end{aligned}
\tag{28}
$$

$$
\begin{aligned}
\Pi_{ik}^{DB} =& \frac{\tau_l}{2}\left[\frac{1}{\rho}\left(p+\frac{B^2}{2\mu_m}\right)\left(\frac{\partial B_i}{\partial x_k}-\frac{\partial B_k}{\partial x_i}\right)\right] \\
&+\frac{\tau}{2}\left[\left(p+\frac{B^2}{2\mu_m}\right)\left(B_i\frac{\partial}{\partial x_k}\frac{1}{\rho}-B_k\frac{\partial}{\partial x_i}\frac{1}{\rho}\right)\right] \\
&+\frac{\tau}{2}\left[u_k B_l\frac{\partial u_i}{\partial x_l}-u_i B_l\frac{\partial u_k}{\partial x_l}\right]+\frac{\tau}{2}\left[\frac{1}{\rho}\frac{B_i B_l}{\mu_m}\frac{\partial B_k}{\partial x_l}-\frac{1}{\rho}\frac{B_k B_l}{\mu_m}\frac{\partial B_i}{\partial x_l}\right] \\
&+\frac{\tau}{2}\left[u_k B_i\frac{\partial u_l}{\partial x_l}-u_k B_l\frac{\partial u_i}{\partial x_l}+u_k u_l\frac{\partial B_i}{\partial x_l}\right] \\
&+\frac{\tau}{2}\left[B_i u_l\frac{\partial u_k}{\partial x_l}+\frac{B_i}{\rho}\frac{\partial p}{\partial x_k}+\frac{B_i}{\rho}\frac{\partial}{\partial x_k}\frac{B^2}{2\mu_m}-\frac{B_i}{\rho}\frac{\partial}{\partial x_l}\frac{B_k B_l}{\mu_m}\right] \\
&+\frac{\tau}{2}\left[-u_i B_k\frac{\partial u_l}{\partial x_l}+u_i B_l\frac{\partial u_k}{\partial x_l}-u_i u_l\frac{\partial B_k}{\partial x_l}\right] \\
&+\frac{\tau}{2}\left[-B_k u_l\frac{\partial u_i}{\partial x_l}-\frac{B_k}{\rho}\frac{\partial p}{\partial x_i}-\frac{B_k}{\rho}\frac{\partial}{\partial x_i}\frac{B^2}{2\mu_m}+\frac{B_k}{\rho}\frac{\partial}{\partial x_l}\frac{B_i B_l}{\mu_m}\right]
\end{aligned}
\tag{29}
$$

It was shown in [2] that the dissipative terms of the gas dynamics system of equations are small in comparison with the convective terms. The corresponding dissipative terms are associated with the real physics processes and an important remark is that in this case the gas dynamics parameters such as viscosity and heat conductivity are obtained from the kinetic theory.

The Navier-Stokes viscosity is identified as the first term of Eq. (27):

$$
\Pi_{ik}^{NS}=\frac{\tau}{2}\left(p\frac{\partial u_i}{\partial x_k}+p\frac{\partial u_k}{\partial x_i}-\frac{2}{3}p\frac{\partial u_l}{\partial x_l}\delta_{ik}\right)=\mu\left(\frac{\partial u_i}{\partial x_k}+\frac{\partial u_k}{\partial x_i}-\frac{2}{3}\frac{\partial u_l}{\partial x_l}\delta_{ik}\right) \tag{30}
$$

where the bulk viscosity component is neglected and the shear viscosity μ is related to the gas pressure p and the characteristic time τ as $\mu=\frac{\tau}{2}p$.

The Navier-Stokes thermal flux vector is identified as the first term of Eq. (28):

$$
Q_i^{NS}=\frac{\tau}{2}\left(\frac{5}{2}p\frac{\partial}{\partial x_i}\frac{p}{\rho}\right)=k\frac{\partial T}{\partial x_i} \tag{31}
$$

with T gas temperature and k thermal coefficient expressed as $k=\frac{1}{Pr}\frac{5}{2}R\frac{\tau}{2}p$, with Pr Prandtl number.

A similar analysis of the dissipative terms of the electromagnetic processes gives the estimation of their smallness. The correct magnetic viscosity is represented as part of dissepative term. The gas resistivity is identified as the first term of Eq. (29) and also appears as a result of the kinetic formulation:

$$
\Pi_{ik}^{B}=\frac{\tau_m}{2}\left[\left(p+\frac{B^2}{2}\right)\left(\frac{\partial B_i}{\partial x_k}-\frac{\partial B_k}{\partial x_i}\right)\right]=\eta\left(\frac{\partial B_i}{\partial x_k}-\frac{\partial B_k}{\partial x_i}\right) \tag{32}
$$

Where the $\eta=\frac{\tau_m}{2}\left(p+\frac{B^2}{2}\right)$ represents the resistivity.

5 THE COMPUTATIONAL ALGORITHM

The computational task consists of the solution of the kinetic consistent MHD system of equations (20 - 23).

The computational algorithm used in this paper is build upon the Finite Volume method applied to the gas dynamic equations and on the Constrained Transport method applied to the magnetic induction equation [14, 15].

In the finite volume method the conserved gas dynamics variables, density, gas momentum and energy, are averaged above the volume of the computational cell. The contrained transport method treatment is based on the area-averaging of the magnetic field through the surfaces of the grid cells.

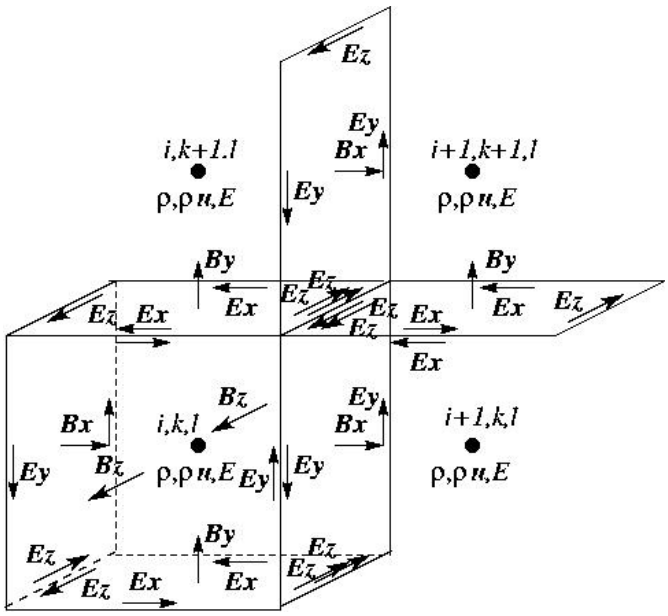


Figure 1: 3D Computational Domain

Fig. 1 shows the four neighbour to the cell (i, k, l) used in evaluations of the hydrodynamics and electromagnetic variables.

The gas dynamics observables - mass density, momentum and energy density are defined at the cell centre. The components of the magnetic field are defined at the face centers of the cells. A duality is established between the electric field and the fluxes. This duality is utilised to obtain the electric field at the edges of the computational cell through a reconstruction process that is applied directly to the properly up-winded fluxes. The electric field is then utilised to make an update of the magnetic fields that preserves the solenoidal nature of the magnetic field and ensures that the magnetic field in a magnetohydrodynamics model remains strictly solenoidal up to discretisation errors.

The explicit finite volume computational algorithm for the magneto gas gas dynamics are used. This method has the convenient stability conditions and very important for adaptation on the massive parallel computing systems.

6 RESULTS OF NUMERICAL MODELLING

The parallel implementation of the numerical algorithm is obtained through the domain decomposition in distributed memory approach. The communication between processors is performed via MPI protocol. The computational method needs only one layer of ghost cells at the boundary and along the edges of each subdomain which is exchanged between the domains with MPI calls. The algorithm is well suited to distributed memory clusters as it is time explicit and

requires only one MPI call per time step, reducing significantly the processor communication time contribution to the total computational time.

The demonstration of the method is performed on the basis of the solution of the spherical expansion problem of ionised gas and the solution of the expansion of an ionised gas in strong magnetic field. The initial conditions are the same as proposed in [16] and are expressed in normalised units in order to provide a direct comparison. They consist of a sphere with radius 0.1 placed in the centre of the physical region with pressure of 100 in comparison to the overall represented area with pressure 1. The density is uniform and equal to 1 in the full computational domain. For the study of ionised gas in a strong magnetic field the uniform magnetic field $B_z = 5/\sqrt{\pi}$ aligned with the z coordinate is added to the initial conditions. The adiabatic coefficient of the gas is chosen as $\gamma = 5/3$. Reflective border conditions are used throughout the simulation.

The simulations are performed for a Cartesian rectangular mesh $100 \times 100 \times 100$ in the physics domain [0,1].

Fig. 2 present the state of the 3D simulation of the processes for relative time 0.03. On the 3D pictures the arrows represent the velocities of the ionised gas and the colour represents the density of gas. 3D figures clearly show the confinement of the ionized gas in the cylindrical area along z due to the magnetic field.

Fig. 3,4 represent the $2D$ projections of the density, pressure and kinetic energy of the gas expansion without magnetic field and the $1D$ density profile for these condition at time $t = 0.03$.

Fig. 5,6 represents the $2D$ projections of the density, pressure, magnetic pressure and kinetic energy for the gas expansion problem of the ionized gas with initial magnetic field and Fig. 7 shows the $1D$ profile of the density for these conditions at time $t = 0.03$.

This test demonstrates the ability of the computational algorithm in preserving perfect symmetry. The comparison of the results in [16] shows a reasonable agreement and will be analysed further.

Another significant numerical test is the implosion. It consists of a 3-dimensional generalisation of the Sod shock tube problem [17]. The initial condition corresponds to hydrogen gas presenting a discontinuity perpendicular to the diagonal of the computational domain. The left state has initial density $\rho_l = 0.719$ Kg/m^3 and pressure $P_r = 722680$ Pa. The right state has initial density $\rho_r = 0.125 \times \rho_l$ and pressure $P_r = 0.14 \times P_l$. The initial magnetic field of 0.1 T is aligned with the y-axis. Reflective boundary conditions are used.

The simulations are performed for a Cartesian rectangular mesh $2000 \times 2000 \times 2000$ in the physics domain [0,10] μm.

Fig. 8 represents the $2D$ projections of the density at time t $= 1.76 \times 10^{-8}$ for the two cases with and without magnetic field.

The reflection of the shock generated by the initial contact discontinuity drives consequent vortices and instabilities. The result is characterised by complex shock reflections and rarefactions. In absence of magnetic field the symmetry along the diagonal is preserved. The presence of the magnetic field changes dramatically the picture. The deflection of the charged media flow in the direction of the magnetic field is clearly visible. Another observed effect is the strong smoothing of the instabilities of the gas dynamics, which corresponds to a physical experimental evidence. The possibility of such observation is the basis of the importance of the very detailed space discretisation, which is available with the proposed method.

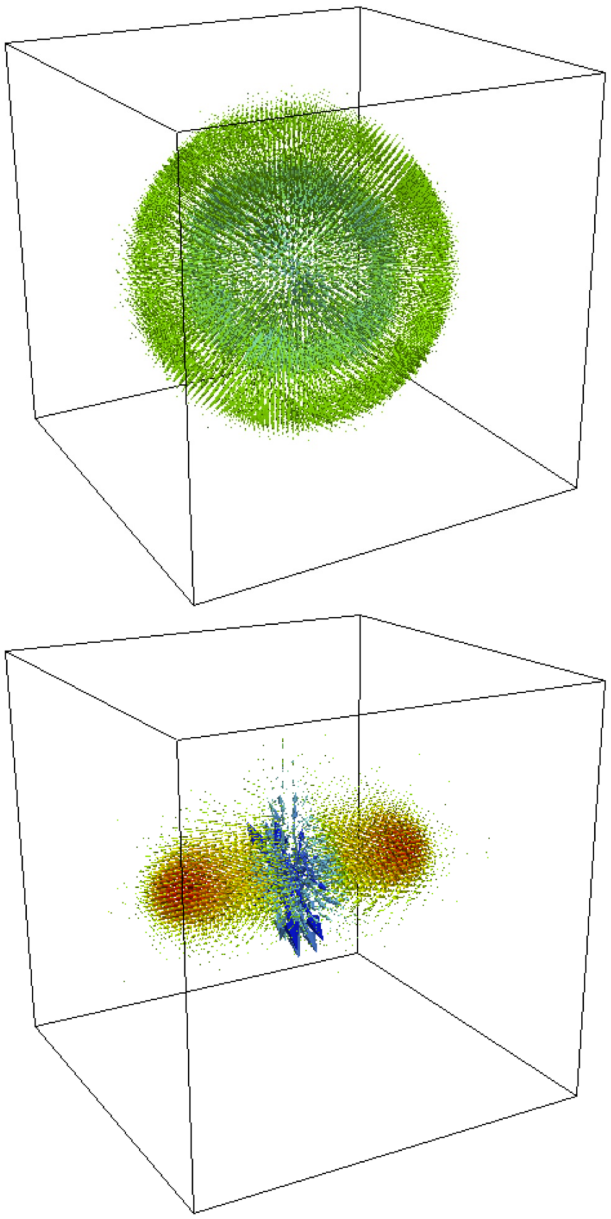


Figure 2: $3D$ view of the conductive gas expansion in strong magnetic field

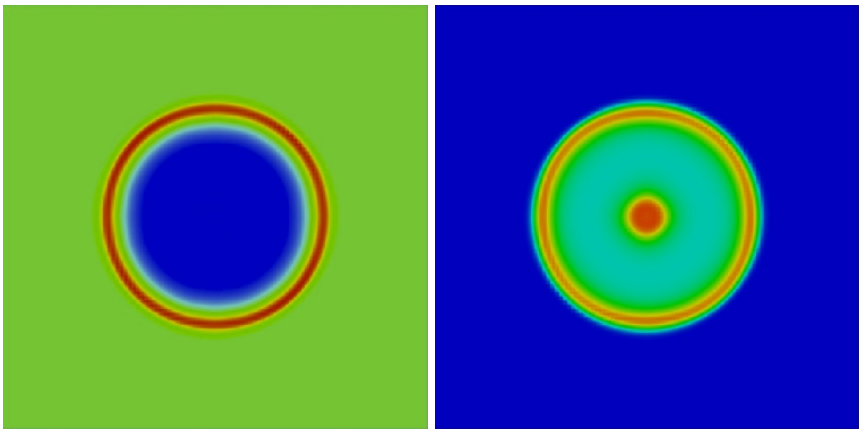


Figure 3: $2D$ gas density and 2D gas pressure projections

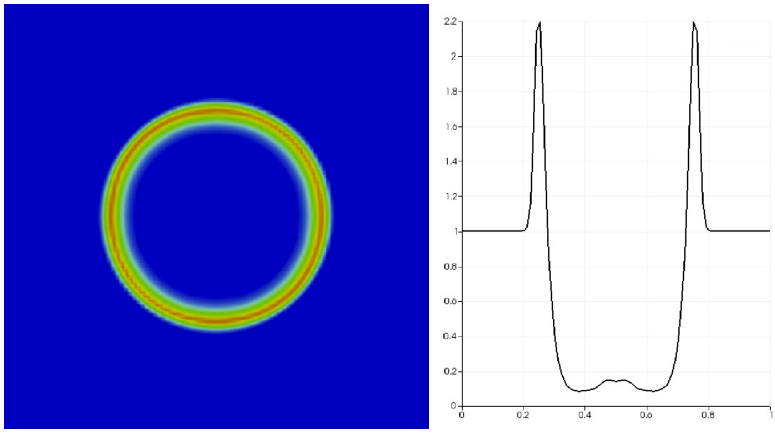


Figure 4: $2D$ kinetic energy and 1D density profile

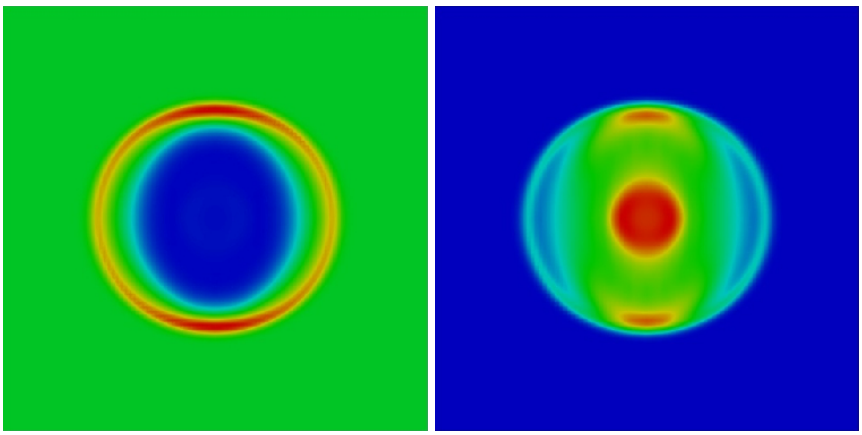


Figure 5: $2D$ gas density and $2D$ gas pressure projections in the magnetic field $5/\sqrt{\pi}$

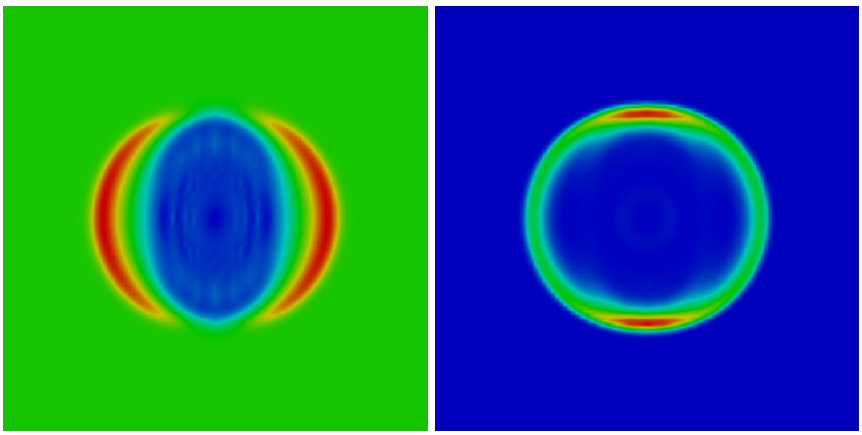


Figure 6: $2D$ magnetic pressure and $2D$ kinetic energy in the magnetic field $5/\sqrt{\pi}$

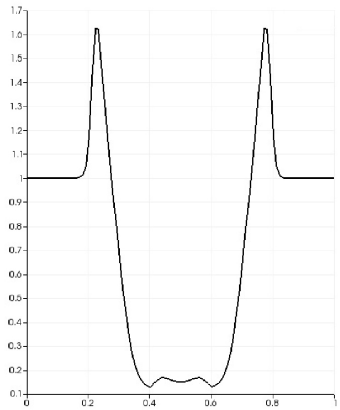


Figure 7: $1D$ density profile in the magnetic field $5/\sqrt{\pi}$

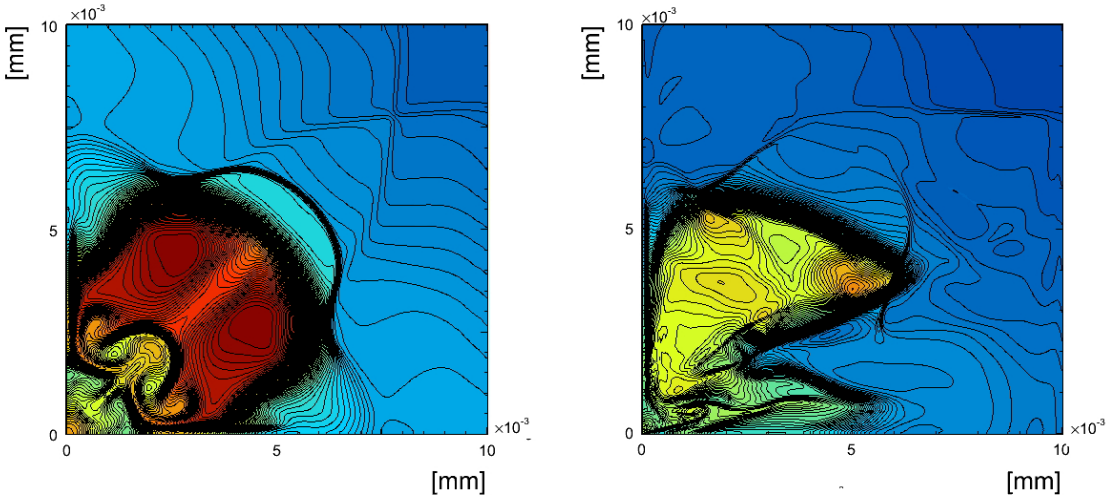


Figure 8: Sod shocktube problem: 2D gas density projections at time 1.76×10^{-8} s without (a) and with (b) magnetic field.

7 CONCLUSIONS

A new 3D kinetic algorithm has been developed for the solution of the magnetohydrodynamics problems. The novel feature of the method is that the local complex Boltzmann-like distribution function incorporated most of the electromagnetic processes terms. The fluxes of mass, momentum and energy across the cell interface as well as the magnetic field are calculated by integrating a local complex Boltzmann-like distribution function over the velocity space. Thus by using this distribution function to calculate the mass, momentum and energy fluxes, most of the electromagnetic contributions are calculated directly, i.e. one does not have to solve the hydrodynamics and magnetic force components separately or differently.

A staggered, divergence free mesh configuration is used for the evaluation of the electromagnetic behaviour.

Numerical examples demonstrate that the proposed method can achieve high numerical accuracy and resolve strong shock waves of the magnetohydrodynamics problems.

The explicit method is chosen with respect to optimal adaptation on the large scale parallel computing systems. The improvement of the stability conditions which is one of the limiting factor for the explicit method will be studied by the implementation of the hyperbolic type of the magneto gas dynamic equations.

8 ACNOWLEDGEMENT

This work was supported by the Russian Science Foundation, project 14-11-00170

REFERENCES

[1] B.Chetverushin, A.Pavlov, M.Volchinskaya, An Integration Scheme for Gas Dynamic Equations, Keldysh Institute of Applied Mathematics, RAS Preprint 113 (1983).

[2] B.Chetverushin, Kinetic Schemes and Quasi-Gasdynamic System of Equation, CIMNE, 2008.

[3] B.Chetverushkin, Resolution limits of continuous media models and their mathematical formulations, Matem. Mod. 24 (2012) 33-52.

[4] B.N. Chetverushkin, N. D'Ascenzo, V.I. Saveliev, Kinetically Consistent Magnetogasdynamic Equations and Their Use in Supercomputer Computations, Dokl. Math. 90 (2014) 495-498.

[5] B.N. Chetverushkin, N. D'Ascenzo, V.I. Saveliev, Three Level Scheme for Solving Parabolic and Elliptic Equations, Dokl. Math. 91 (2015) 341-343.

[6] B.N. Chetverushkin, N. D'Ascenzo, V.I. Saveliev, On an Algorithm for solving parabolic and elliptic equations, Comput. Math. Math. Phys. 55 (2015) 1290-1297.

[7] B.N. Chetverushkin, N. D'Ascenzo, V.I. Saveliev, Hyperbolic type explicit kinetic scheme of magneto gas dynamics for high performance computing systems, Russ. J. Num. Anal. Math. Model. 30 (2015) 27-36.

[8] L.Boltzmann, Lectures on Gas Theory, Dover, 1964.

[9] L. Tonks, Plasma Electron Drift in a Magnetic Field with a Velocity Distribution Function, Phys.Rev. 52 (1937) 710-713.

[10] J.-P.Croisille, R.Khanfir, G.Chanteur, Numerical Simulation of the MHD Equations by a Kinetic-Type Method, J.Sci.Comput. 10 (1995) 81-92.

[11] J.D.Huba, J.G.Lyon, A new 3D MHD algorithm: the distribution function method, J. Plasma Phys. 61 (1999) 391-405.

[12] P.-J.Dellar, Lattice Kinetic Schemes for Magnetohydrodynamics, J. Comput. Phys. 179 (2002) 95-126.

[13] L.Landau, E.Lifshitz, The Classical Theory of Fields. Vol. 2., Nauka, 1988.

[14] D.Balsara, D.Spicer, A Staggered Mesh Algorithm Using High Order Godunov Fluxes to Ensure Solenoidal Msgnetic Fields in Magnetohydrodynamics Simulation, J. Comput. Physics 149 (1999) 270-292.

[15] D.Balsara, Divergence-Free Adaptive Mesh Refinement for Magnetohydrodynamics, J. Comput. Phys. 174 (2001) 614-648.

[16] H.Tang, K.Xu and C.Cai, Gas-Kinetic BGK Scheme for Three Dimentional Magnetohydros, Numer. Math. Theor. Meth. Appl. 3 (2010) 387-404.

[17] Sod, G. A., A Survey of Several Finite Difference Methods for Systems of Nonlinear Hyperbolic Conservation Laws., J. Comput. Phys., (1978)

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ERROR ANALYSIS AND QUANTIFICATION IN NEURON SIMULATIONS

Francesco Casalegno*[1], Francesco Cremonesi[1], Stuart Yates[1], Michael L. Hines[2], Felix Schürmann[1], and Fabien Delalondre[1]

[1]Blue Brain Project
École Polytechnique Fédérale de Lausanne
Campus Biotech, Chemin des Mines 9, 1202 Genève, Switzerland
{francesco.casalegno,francesco.cremonesi,sam.yates,felix.schuermann,fabien.delalondre}@epfl.ch

[2]Yale University
P.O.Box 208001, New Haven, CT 06520-8001, United States
michael.hines@yale.edu

Keywords: Computational Neuroscience, NEURON Software, Error Analysis, Splitting Methods, Stiff Systems, Nonlinear PDEs.

Abstract. *The Blue Brain Project (BBP) uses the NEURON simulator to model the electrical activity of large networks of morphologically detailed neurons. Each individual neuron is typically described by a model that couples the actions of localized membrane mechanisms with an electrical system described by the cable equation on a topology derived from the neuron's dendritic tree. NEURON discretizes the electrical problem in space with finite differences. The system is then solved in time with an implicit Euler or Crank–Nicolson scheme, using Strang splitting to decouple the evolution of the mechanisms from the membrane potential. A typical spatial discretization of a single neuron will have hundreds of elements, and the resulting linear system for the implicit solver is almost tridiagonal.*

In this paper, we present a detailed analysis of the different sources of error arising from the biological and mathematical models underlying NEURON simulations. We provide a detailed account of the mathematical model, identify and evaluate the sources of uncertainty in the biological and numerical models and quantify the errors resulting from the model discretization and from the choice of the integrator. Post-processing based techniques are applied to assess the numerical error in the solution. Of particular interest is the analysis of the discontinuities arising from the pointwise synaptic processes that constitute part of the model, including the effects of voltage-proportional synaptic conductance.

We validate our analysis through a series of numerical experiments on a branched dendrite model incorporating Hodgkin–Huxley distributed ion channels with simple alpha-synapes and biologically realistic AMPA/NMDA activated synapses. This model exhibits the action-potential behaviour with fast dynamics characteristic of a typical simulation of a network of neurons.

1 INTRODUCTION

The assessment of the errors and uncertainties associated with the numerical simulation of mathematical models describing physical phenomena represents a topic of crucial importance in both scientific and engineering research. Several techniques have been developed to provide information on the different sources of error and thereby provide assessments of the accuracy of simulation results.

The motivations for this error analysis are twofold: firstly, confidence in the accuracy and reliability of simulation results requires bounds on the errors of those simulations; secondly, the choice of discretization in space and time should be chosen such that the global error is constrained to lie within certain user-specified bounds. The latter is particularly pertinent when considering the implementation of numerical methods that involve mesh adaptivity and variable step-size integrators.

The results of a numerical simulation are subject to four main classes of error.

- *Modelling errors* are produced when translating the real world physical phenomena into the mathematical model consisting of a set of differential equations.
- *Parameter errors* are due to the uncertainties in the measurements of the quantities necessary for the determination of the parameters of the mathematical model.
- *Discretization errors* arise from the choice of the numerical discretization methods utilized for the numerical solution of the differential equations.
- *Round-off errors* are introduced by the finite-precision arithmetic of the floating point representation used by the computer.

We aim in this paper to address a perceived gap in the literature concerning error bounds and estimates for the results obtained using NEURON simulator[3], the primary tool employed by the BBP for the analysis of the electrical behaviour of morphologically detailed neural networks.

The outline of the paper is as follows. The modelling and parameter errors for this problem are discussed in Section 2. As a prelude to the numerical error analysis, we present in Section 3 a formal statement of the mathematical model. In Section 4 we discuss the space discretization of the differential equations using the finite difference method and we consider the implementation of an error estimator based on the Zienkiewicz–Zhu post-processing technique. In Section 5 we examine the numerical integration methods currently implemented in the NEURON simulator and propose different techniques that allow an improvement in the accuracy of the simulation results. We provide results of numerical experiments in Section 6, using a simplified dendrite model with a single branching point. While consideration of round-off error is important, we do not address it in this paper.

2 MODELLING AND PARAMETER ERRORS

Three simplifying assumptions underlie the derivation of the mathematical model of neuronal electrical behaviour used by NEURON: the Maxwell quasi-static approximation, the 1-D approximation, and the distribution of ion channels and point processes.

The quasi-electrostatic approximation to Maxwell's equations assumes that inductive effects and wave propagation delays are negligible (see for example [13]). In the NEURON model, it is assumed the only significant capacitive phenomena are those caused by the difference in extra- and intra-cellular potentials at the membrane of the dendrite, and that the frequencies involved are sufficiently low that the behaviour lies in the quasi-electrostatic domain. This simplification was first justified by the research on electro-physiological systems of Plonsey and Heppner [16]. However, Bossetti et al. [2] have recently argued that in the case of short and fast impulses the

argument of [16] may be undermined. To test their ideas, Bossetti et al. relaxed the quasi-static assumption and simulated the resulting inhomogeneous Helmholtz equation. Their results show that whilst propagation and induction effects can be quite safely neglected, the system exhibits a higher sensitivity with respect to capacitive and conductive effects. In any case, numerical results obtained by the authors find a relative error of 5 to 13% for pulse widths within the range of typical neural events (25µs to 1 ms).

The unbranched sections of the dendritic tree of a neuron are typically much thinner than they are long, and so are amenable to a one-dimensional approximation that represents the potential within the dendrite as a function of the axial distance along a section. This assumption leads to a description in terms of electrical cable theory. A study of Lindsay et al. [12] has criticized this approach by observing that dendritic currents have an essential radial component that may not be neglected in the case of dendritic tapering. Lindsay et al. compared numerical simulations of the simplified 1-D model with ones incorporating higher order perturbations capturing the radial variation, and found that when synaptic input is weighted towards the distal end of the dendrite, the generated spike trains showed statistically significant differences between the models. On the other hand, if the weighting is towards the proximal end of the dendrite the spike trains are statistically indistinguishable, even though the actual firing patterns might differ.

Lastly, the third approximation concerns the distribution of ion channels and synapses. Ion channels are proteins that reside on the membrane of the cell, and are responsible for gating the flow of charged ions through the cell membrane. As such, ion channels are discrete phenomena, giving a spatially localised contribution to the transmembrane current by allowing single particles of charged ions to pass through their selectivity filters. As a consequence of the conductance based formalism these ion channels are not modelled individually. Rather, NEURON operates under the assumption that the contributions of these channels can be averaged across a section of dendrite where the density of their distribution can be considered uniform. Synapses, on the other hand, are modelled as discrete processes but are located on the dendrite only at nodes arising from the spatial discretization.

Once the appropriate model has been chosen for representing a neuron, the computational neuroscientist is still required to specify the value of some input parameters. Here we consider two sources of parameter error: morphology reconstruction and matching electrophysiological data.

Computer reconstructions are commonly used to provide cell-type specific morphologies as input to neural simulations. In this process, an expert controls a microscope to analyse a set of brain slices that have been stained in order to enhance visibility of cellular structures such as the soma, dendrites and axons. Unfortunately, during staining an error due to tissue shrinkage can be introduced; [18], for example, observed a shrinkage of approximately 10% in the X and Y directions and 25% in slice thickness. The work of Blackman et al. [1] investigated the effects of shrinkage and other reconstruction artefacts on the results of computer simulations conducted using the NEURON software. They consider two popular reconstruction techniques: biocytin histology and two-photon imaging. The authors identify some systematic defects introduced by both techniques. They note, however, that clustering of morphological types was successful with both techniques, and that electrophysiological parameters that could be affected by morphological properties of the dendrite, such as length constants, did not seem to be affected significantly by the uncertainty introduced by the two methods. Although recent techniques allow 3-D imaging of intact brains [4], many computer reconstruction techniques in use today require the reduction of the three dimensional brain to a set of two-dimensional slices. This process carries an intrinsic source of error: not only may some branches may be

severed, distorted or deformed during the slicing process; the human expert or image reconstruction software may also make mistakes while connecting multiple two dimensional slices into a three dimensional shape. It is important to consider these sources of error even if they are extremely difficult to quantify, as a neuron's morphology plays an important role in determining its electrophysiological properties [11].

Just as the different components of a computer chip give rise to meaningful emergent behaviour by exploiting differences in their electrical behaviours, it is thought that the ability of a neuron to *integrate* in a very specific way the synaptic input from other cells is at the core of the brain's ability to perform computations. As such, the main goal of developing a detailed neuron model is to be able to reproduce with high fidelity its experimentally measured electrophysiological behaviour, thus distinguishing it from other cells. Often this requires a two step process, which begins by obtaining valid experimental data. In this phase, errors may come from measurement sensitivity and noise. The comprehensive guide on patch clamp experiment setups [19] estimates that a realistic value for noise in patch clamp experiments is around 0.2 pA rms, which considering a typical cell resistance of 100 MΩ leads to a negligible noise amplitude of around 20 µV. Moreover, a typical value of the resolution of such experimental setups is in the order of 0.01 pA to 1 pA, which leads again to a resolution on the membrane potential of roughly 1 µV to 100 µV.

Once the experimental data has been obtained, computational neuroscientists often use an optimization algorithm to find the best set of model parameters to fit such data. Recently the community has shown interest in the use of evolutionary algorithms (see e.g. [5]). As fitting the raw experimental observations can be difficult or overwhelmingly costly in some cases, more and more focus is being placed on feature-based fitting, which consists of trying to replicate a set of secondary measures that can be computed from the experimental data (e.g. a neuron's firing rate under a certain stimulation protocol) rather than replicating the raw data directly (see [20] for a review). These optimization algorithms can suffer from intrinsic uncertainties and errors: they may get stuck in local minima, produce results that are not biologically plausible, incorrectly overfit the data or fail to replicate certain aspects of a neuron's electrophysiology due to a poor choice of the features. In addition, it is hard to quantify the degree of uncertainty in the raw experimental data that will affect the final uncertainty of the input parameters.

3 MATHEMATICAL MODEL

Before discussing numerical recipes for the simulation of neural circuits and error estimation, we need to clearly state the mathematical expression of the biological model. In this section we present the complete formulation of the equations typically solved by NEURON.

The behaviour of electrical signals in neurites is described by *cable theory*, which is derived from an electrical model with a constant axial resistivity and a leaky capacitive membrane. While a great variety of ion channel models are supported by NEURON, we restrict ourselves to the Hodgkin–Huxley model of voltage-gated sodium and potassium ion channels [10].

The model is defined over a domain Ω, comprising a disjoint collection of closed intervals of the real line; each interval corresponds to an unbranched section of dendrite, and the branching structure will be captured by additional boundary conditions applied at the interval end-points. Within the cable theory model, the membrane potential $v(x,t)$ is described by a variable depending only on time t and position x on the axis. In addition to these variables, we have to consider some quantities regulating the behaviour of the ion channels. The ion channel states in the Hodgin–Huxley model are described by dimensionless quantities $n(x,t)$, $m(x,t)$, and $h(x,t)$ with a voltage-dependent evolution in time. Finally, we consider the contribution of

discrete synaptic receptors on the surface of a section, which are governed by a dimensionless per-synapse state $y(t)$.

3.1 Differential equations

The governing equations for the systems are obtained through a balancing of trans-membrane currents and axial currents. The resulting system is the *cable equation* with additional terms capturing the ion channel currents,

$$\begin{aligned} 10^{-3} c_m \frac{\partial v}{\partial t} = {} & \frac{10^4}{2aR} \frac{\partial}{\partial x} \left(a^2 \frac{\partial}{\partial x} v \right) \\ & - g_{\mathrm{K}} n^4 (v - e_{\mathrm{K}}) - g_{\mathrm{Na}} m^3 h (v - e_{\mathrm{Na}}) - g_l (v - e_l) \\ & - \frac{10^2 g_{syn}(y, v)}{2\pi a} \delta_{syn}(x)(v - e_{syn}), \end{aligned} \tag{1}$$

where: the membrane potential v is expressed in mV; a is the dendrite radius [µm]; c_m is the membrane specific capacitance [µF·cm^{-2}]; R is the cytoplasmic resistivity [Ω·cm]; g_{K}, g_{Na}, g_l, and $g_{syn}(y, v)$ are specific conductances [S·cm^{-2}] (the latter depending on other variables); e_{K}, e_{Na}, e_l, and e_{syn} are reversal potentials [mV]; and $\delta_{syn}(x)$ is a Dirac delta centred on the synaptic receptor location x_{syn}.

The evolution of the ion channels states is described by nonlinear functions $\alpha_\times(v)$ and $\beta_\times(v)$ derived by Hodgin and Huxley [10],

$$\begin{aligned} \frac{\mathrm{d}n}{\mathrm{d}t} &= \alpha_n(v)(1 - n) - \beta_n(v) n, \\ \frac{\mathrm{d}m}{\mathrm{d}t} &= \alpha_m(v)(1 - m) - \beta_m(v) m, \\ \frac{\mathrm{d}h}{\mathrm{d}t} &= \alpha_h(v)(1 - h) - \beta_h(v) h. \end{aligned} \tag{2}$$

The evolution of the synaptic receptor state can be described by an ordinary differential equation (ODE) of the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \gamma(y) \tag{3}$$

where both the function $\gamma(y)$ and the behaviour of the specific conductance g_{syn} depend on the kind of synapse to be modelled. We consider here two different kinds of synaptic receptors:

- *Alpha-synapses* (see [3]) are simplified models, where the conductance $g_{syn}(y)$ is only dependent on the synapse state y.
- *Deterministic AMPA/NMDA-synapses* (described in [14]) are more realistic models, in which the conductance $g_{syn}(y, v)$ depends as well on the membrane potential to account for magnesium blocking of the NMDA receptor.

3.2 Initial and boundary conditions

The system of differential equations formed by (1), (2), and (3) are to be combined with initial and boundary conditions. For given initial conditions on the membrane potential

$$v|_{t=t^0} = v^0 \tag{4}$$

we recover the initial values for the ion channels states by considering the steady state solution of Equation (2),

$$n|_{t=t^0} = \frac{\alpha_n(v^0)}{\alpha_n(v^0)+\beta_n(v^0)}, \quad m|_{t=t^0} = \frac{\alpha_m(v^0)}{\alpha_m(v^0)+\beta_m(v^0)}, \quad h|_{t=t^0} = \frac{\alpha_h(v^0)}{\alpha_h(v^0)+\beta_h(v^0)}. \tag{5}$$

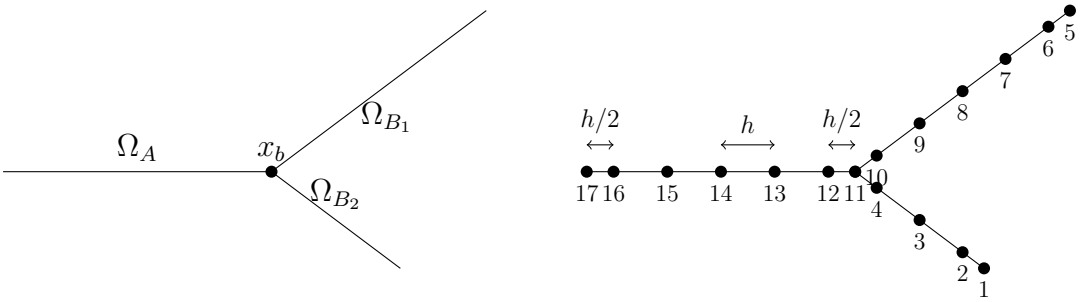
As for the boundary conditions for the membrane potential, we impose the balance of the axial current to obtain natural boundary conditions on the boundary points of each unbranched section Ω_i of the dendrite. We have to distinguish different cases depending on the position of the boundary point x_b within the dendritic tree, that is whether the boundary of the section is a terminal end of the dendrite or if it communicates with other sections through a branching (see [9]).

- On a *terminal point* x_b of the dendrite we have homogeneous conditions

$$\left.\frac{\partial v}{\partial x}\right|_{x=x_b} = 0. \tag{6}$$

- On a *branching point* x_b (see Figure 1a) the boundary conditions link together the gradients of the membrane potentials on the section Ω_A preceding the branching and on the sections Ω_{B_1} and Ω_{B_2} following it,

$$\left(a^2\frac{\partial v}{\partial x}\right)\Big|_{x=x_b}^{A} = \left(a^2\frac{\partial v}{\partial x}\right)\Big|_{x=x_b}^{B_1} + \left(a^2\frac{\partial v}{\partial x}\right)\Big|_{x=x_b}^{B_2}. \tag{7}$$



(a) Branched dendrite

(b) Space discretization and node ordering

Figure 1: Branched dendrite and its space discretization with an example of the node ordering used in NEURON to avoid fill-in during Gaussian elimination.

4 DISCRETIZATION IN SPACE

The strategy implemented in NEURON to solve (1) numerically is known as the *method of lines* (see [7]), which transforms the partial differential equation (PDE) into a system of ODEs through a suitable space discretization and then integrates the latter in time using some numerical scheme. In this section we discuss the discretization in space, while the integration in time is analysed in Section 5.

4.1 Numerical methods

The differential equations (1) constituting the mathematical model are discretized by finite differences to form a system of ODEs. On a given unbranched section, the nodes x_j of the mesh are chosen with an equal spacing h except for the boundary nodes where the spacing is $h/2$, as shown in Figure 1b.

It is important to mention that the numbering of the nodes of the dendrite has a critical role in the structure of the linear system to be solved after space discretization. NEURON adopts the node ordering scheme described in [9], such that branches that are further away from the soma have lower node indexes (see Figure 1b). This gives rise to a global stiffness matrix for the dendrite which is almost tridiagonal and which can be efficiently solved using an ad hoc version of Thomas algorithm with linear complexity, as the Gaussian elimination step requires the same number of arithmetic operation for a tree of N nodes as for a cable with N nodes.

Particular care must be paid to the second order term and to the Dirac delta when applying the finite differences discretization. Among the different finite difference approximations that are possible for the second order term, NEURON uses the discretization

$$\frac{\partial}{\partial x}\left(a^2(x)\frac{\partial v}{\partial x}(x)\right)\bigg|_{x=x_j} \approx \frac{1}{h}\left(a^2_{j+1/2}\frac{v_{j+1}-v_j}{h} + a^2_{j-1/2}\frac{v_{j+1}-v_j}{h}\right). \tag{8}$$

As for the Dirac delta term, NEURON uses the discretization

$$\delta_{syn}(x)v(x)\big|_{x=x_j} \approx \frac{\delta_{j,j_{syn}}}{h}v_j, \tag{9}$$

where $\delta_{j,j_{syn}}$ is the Kronecker delta and j_{syn} is the index of the node of the synaptic receptor.

Note that the finite difference scheme as described above coincides, up to higher order terms, with a finite element scheme with piecewise linear elements (see [17]). As a consequence, the order of convergence in the $L^2(\Omega)$ norm for the numerical solution v_h and its gradient ∇v_h can be expected to be quadratic and linear respectively. In addition, the equivalence of this finite difference scheme to the finite element method allows us to use error estimation techniques suited to finite element algorithms.

4.2 Error estimation

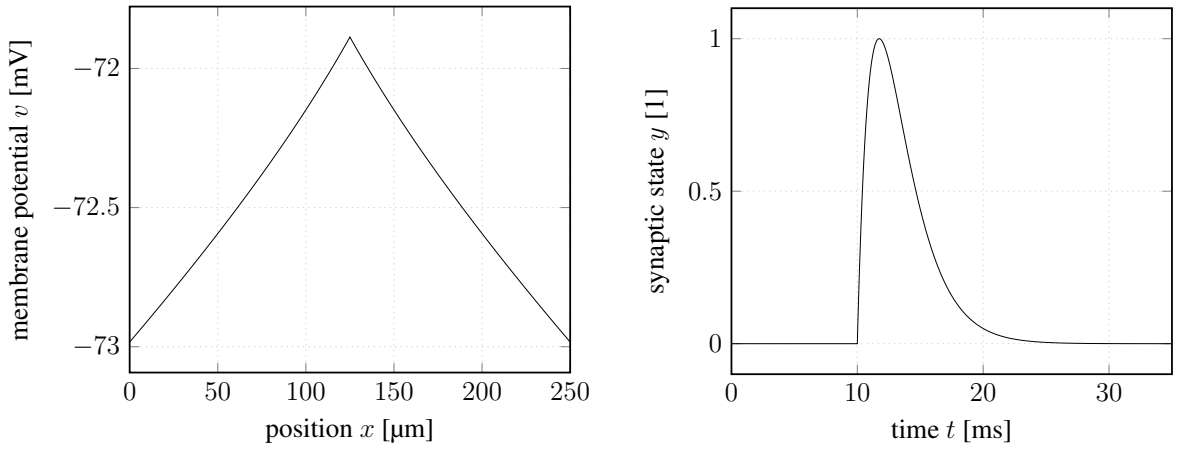
By interpreting the finite difference discretization used in NEURON as a variant of the finite element method, we can consider applying classical error estimation techniques for finite element methods. Residual-based estimators cannot be used, however, as the PDE (1) contains a non-integrable Dirac function. Instead, we use a Zienkiewicz–Zhu (ZZ) post-processing estimator based on a superconvergent patch recovery, as described in [21]. This method has been proved to be amongst the most robust error estimators available.

In our case, the superconvergent patch recovery technique is used to compute, from the numerical solution v_h of the finite difference scheme, a *better gradient* Gv_h converging quadratically to the exact gradient ∇v in the $L^2(\Omega)$ norm. The exact error in the gradient is thus approximated using

$$\|Gv_h - \nabla v_h\|_{L^2(\Omega)} = \|\nabla v - \nabla v_h\|_{L^2(\Omega)} + O(h^2). \tag{10}$$

The better gradient is a piecewise linear function computed locally using the values of the discrete solution v_h on triples of consecutive nodes x_{j-1}, x_j, and x_{j+1}. It is then clear that

this procedure provides worse estimates on boundary elements. Furthermore, we remark that this patch recovery fails where the exact solution contains a singularity (see Figure 2a). In order to adapt the method to our case, we split each section containing a synaptic receptor into subsections where the solution is smooth and then apply ZZ post-processing on each subsection.



(a) Singularity in space.

(b) Singularity in time

Figure 2: Singularities in the system: (a) membrane potential on a unbranched section of the dendrite with a singularity on the synaptic receptor ($x_{syn} = 125$ µm); (b) synaptic state over time with a singularity at spike arrival ($t_{rec} = 10$ ms).

5 INTEGRATION IN TIME

After the space discretization performed with the finite difference method described in Section 4, we obtain from (1) a system of ODEs that can be written together with (2) and (3) in the form

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = A(\boldsymbol{n}, \boldsymbol{v}, y)\boldsymbol{v} + a(\boldsymbol{n}, \boldsymbol{v}, y), \\ \dfrac{\mathrm{d}\boldsymbol{n}}{\mathrm{d}t} = B(\boldsymbol{v})\boldsymbol{n} + b(\boldsymbol{v}), \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = \gamma(y). \end{cases} \tag{11}$$

where $\boldsymbol{v}$ is the vector of the approximated membrane potential values v_j, and $\boldsymbol{n}$ is the vector of the gate states values n_j, m_j, and h_j at the mesh nodes x_j. We now discuss how to integrate (11) numerically and estimate the error in time. More specifically, we firstly present the numerical methods currently implemented in NEURON and we then suggest some techniques to improve their order of convergence and speed.

5.1 Numerical methods: current NEURON implementation

The structure of the system of ODEs (11) is particularly important for the choice of the numerical integration method. As mentioned in Section 4, the matrix $A(\boldsymbol{n}, \boldsymbol{v}, y)$ has a particularly convenient almost tridiagonal form, while it is easy to see that $B(\boldsymbol{v})$ is purely diagonal.

Another important consideration for the choice of the method is the stiffness of the system. The matrix $A(\boldsymbol{n}, \boldsymbol{v}, y)$ is obtained from the space discretization of the parabolic PDE (1) using the method of lines and thus it has a very large negative eigenvalue whose magnitude scales as

$1/h^2$ (see [7]), so that the ODE for $\boldsymbol{v}$ is highly stiff. For the particular form of the Hodgin–Huxley model's functions $\alpha_\times(v)$ and $\beta_\times(v)$ in (2) it can be shown that the matrix $B(\boldsymbol{v})$ has a spectral radius smaller than 10, so that the ODE for $\boldsymbol{n}$ non-stiff.

The system of ODEs (11) is highly nonlinear but is amenable to being split into two subsystems for $\boldsymbol{v}$ and for $\boldsymbol{n}$ and y respectively, to simplify its numerical integration. The idea of splitting methods (see [7]) is to approximate the exact flow φ_τ of an ODE $\mathrm{d}z/\mathrm{d}t = f^A(z) + f^B(z)$ using a composition of the flows of the partial ODEs as

$$\varphi_\tau \approx \varphi^A_{a_1\tau} \circ \varphi^B_{b_1\tau} \circ \cdots \circ \varphi^A_{a_J\tau} \circ \varphi^B_{b_J\tau}. \tag{12}$$

Goldman and Kaper [6] proved that for every splitting method of the form of (12) with order $p > 2$ there exist two indices $1 \leq j_1, j_2 \leq J$ such that $a_{j_1}, b_{j_2} < 0$ and thus the resulting method in unstable for stiff systems. Due to this order barrier, the second order *Strang splitting*

$$\varphi_\tau = \varphi^A_{\tau/2} \circ \varphi^B_\tau \circ \varphi^A_{\tau/2} + O(\tau^3) \tag{13}$$

is optimal. This is the splitting adopted in NEURON for the integration in time of the system of ODEs.

Let us now firstly assume that the synaptic receptor is modelled with a conductance $g_{syn}(y)$ independent from the membrane potential. Then we can drop the dependence on $\boldsymbol{v}$ of the terms on the right-hand side of the ODE for $\boldsymbol{v}$ and at each time step we have to compute $\boldsymbol{v}^{k+1} \approx \boldsymbol{v}(t^{k+1})$ and $\boldsymbol{n}^{k+1/2} \approx \boldsymbol{n}(t^{k+1/2})$, $y^{k+1/2} \approx y(t^{k+1/2})$ by solving

$$\begin{aligned}
\frac{\mathrm{d}\boldsymbol{n}}{\mathrm{d}t} &= B(\boldsymbol{v}^k)\boldsymbol{n} + b(\boldsymbol{v}^k), && t \in [t^{k-1/2}, t^{k+1/2}], \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \gamma(y), && t \in [t^{k-1/2}, t^{k+1/2}], \\
\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= A(\boldsymbol{n}^{k+1/2}, y^{k+1/2})\boldsymbol{v} + a(\boldsymbol{n}^{k+1/2}, y^{k+1/2}), && t \in [t^k, t^{k+1}].
\end{aligned} \tag{14}$$

The value of y^{k+1} can be often computed exactly, since $y(t)$ is typically a damped exponential or a similar function. The values of $\boldsymbol{v}^{k+1}$ and $\boldsymbol{n}^{k+1/2}$ could be computed by solving exactly (14), since after the splitting their ODEs are fully linearized. However the exact solution requires the computation of the exponential of A and B—the matrix A has a tridiagonal pattern, so its exponential is dense and its computation would be particularly expansive; on the other hand, the matrix B is diagonal so it is possible to compute its exponential with little effort. In NEURON, therefore, $\exp(\tau B)$ is computed exactly, whilst a Padé approximation (see [8]) is used to avoid the computation of $\exp(\tau A)$. Strang splitting introduces a second order error in time, therefore only Padé approximations of order $p \leq 2$ are convenient. Moreover, a stable approximation has to be chosen due to the stiffness of the ODE. In NEURON the exponential is approximated using the second order approximation equivalent to the Crank–Nicolson method

$$R_{CN}(\tau A) = \left(I - \tfrac{\tau}{2}A\right)^{-1}\left(I + \tfrac{\tau}{2}A\right) = \exp(\tau A) + O(\tau^3). \tag{15}$$

This second order approximation is obtained in NEURON by computing a half step of the first order Padé approximation $R_{BE}(\tau A) = (I - \tau A)^{-1}$ corresponding to backward Euler, and then recovers the Crank–Nicolson approximation through the relationship

$$R_{CN}(\tau A) = 2R_{BE}(\tfrac{\tau}{2}A) - I. \tag{16}$$

The resolution of an almost tridiagonal system (which can be solved using a modified Thomas algorithm with linear complexity) is thereby performed only once while maintaining a second order method.

Let us now consider the case of the more realistic AMPA/NMDA synapse model, where the conductance $g_{syn}(y, \boldsymbol{v})$ depends as well on the membrane potential. In this case the ODE for $\boldsymbol{v}$ contains a nonlinear term

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = A(\boldsymbol{n}^{k+1/2}, y^{k+1/2}, \boldsymbol{v})\boldsymbol{v} + a(\boldsymbol{n}^{k+1/2}, y^{k+1/2}, \boldsymbol{v}), \quad t \in [t^k, t^{k+1}]. \tag{17}$$

The current implementation in NEURON uses the value $\boldsymbol{v}^k$ (which is known) to simplify (17) and avoid nonlinearities, obtaining

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = A(\boldsymbol{n}^{k+1/2}, y^{k+1/2}, \boldsymbol{v}^k)\boldsymbol{v} + a(\boldsymbol{n}^{k+1/2}, y^{k+1/2}, \boldsymbol{v}^k), \quad t \in [t^k, t^{k+1}]. \tag{18}$$

In Section 6 we see that this technique unfortunately has the drawback of reducing the order of convergence of the numerical integration scheme from quadratic to linear.

5.2 Numerical methods: suggested improvements

We now want to present some numerical integration techniques to improve the results obtained using the aforementioned current implementation of NEURON. Our suggestions are aimed at enhancing both the speed of the integration in time and the accuracy of the results.

We mentioned above that NEURON, in its current implementation, integrates the ODE for $\boldsymbol{n}$ in (14), by computing the matrix exponential $\exp(\tau B)$ exactly. Such a high precision computation is however unnecessary, since the global accuracy of the method cannot exceed second order in time. As a consequence we can consider using a cheaper Padé approximation for the exponential of B. In this case the ODE is non-stiff, so we can choose an explicit approximation. The explicit second order Padé approximation of the exponential corresponds to the Heun method

$$R_H(\tau A) = I + \tau A + \tfrac{\tau}{2}A^2 = \exp(\tau A) + O(\tau^3). \tag{19}$$

This strategy does not reduce the order of convergence of the numerical method and at the same time avoids the computation of exponential functions performed in NEURON, resulting in a more efficient integration scheme.

Regarding the integration of realistic model of synapses with a conductance $g_{syn}(y, \boldsymbol{v})$ dependent on the potential, such as for the AMPA/NMDA model, we observed that the current implementation of NEURON simplifies the system of ODEs by eliminating the nonlinear terms with a technique that destroys the second-order accuracy of the numerical integration (see Section 6). We suggest here a different approach that allows us to preserve the quadratic order of convergence even in case of voltage dependent synaptic conductances. The idea is to use one Newton–Raphson iteration to solve the nonlinear equations corresponding to backward Euler with step size $\tau/2$; a second order method corresponding to Crank–Nicolson) is hence obtained by applying the formula (16) to the result. We point out that this strategy still requires only the solution of an almost tridiagonal system, so that no additional computational cost is required, while the accuracy of the method is significantly improved, as we show in Section 6.

Finally, a critical point is represented by the discontinuity in the ODE $\mathrm{d}y/\mathrm{d}t = \gamma(y)$. Indeed, $\gamma(y) = 0$ in the intervals between inter-spike intervals; when a spike is received at time $t = t_{rec}$ from a presynaptic neuron, $\gamma(y)$ presents then a discontinuity, generating a singularity in y (see

Figure 2b) and hence in the other variables of the system. It has been largely observed (see [7]) that crossing a discontinuity without some specific handling can introduce significant errors. In our simulations we observed (as we show in Section 6) that the current implementation of NEURON produces a global error which is $O(\tau)$ instead of $O(\tau^2)$ in every simulation that includes spike arrivals, as a result of this lack of adequate discontinuity handling. As shown by Hairer [7], the best way to compute a solution crossing a discontinuity consists in implementing an integrator that stops the computation at the point of discontinuity and restarts it with the new value of the right-hand side of the ODE. This technique is considerably more accurate than even a variable step size integrator with automatic error control. We therefore adopted this method for the handling of discontinuity, and the numerical results discussed in Section 6 demonstrate the recovery of a global $O(\tau^2)$ convergence.

5.3 Error estimation

Both local error (i.e. after one time step) and global error (i.e. at the end of the simulation time T) can be estimated using Richardson extrapolation. Given a step size τ, the global error in the numerical solution $\boldsymbol{v}_\tau(T)$ can be estimated from the numerical solution $\boldsymbol{v}_{\lambda\tau}(T)$ computed with a step size λ times larger (in our simulations we use $\lambda = 2$). These calculations, being independent, can be computed efficiently in parallel. A superconvergent approximation is then obtained using Richardson extrapolation

$$\boldsymbol{v}_{R(\tau)}(T) = \frac{\lambda^2 \boldsymbol{v}_\tau(T) - \boldsymbol{v}_{\lambda\tau}(T)}{m^2 - 1}, \tag{20}$$

and the exact global error is then approximated by

$$\|\boldsymbol{v}_{R(\tau)}(T) - \boldsymbol{v}_\tau(T)\|_2 = \|\boldsymbol{v}(T) - \boldsymbol{v}_\tau(T)\|_2 + O(\tau^3). \tag{21}$$

6 NUMERICAL RESULTS

To test the validity of the different numerical methods discussed in Section 4 and 5 we considered a simple model of a branched dendrite with uniform space discretization, such as the one shown in Figure 1b. The codes have been implemented in GNU Octave to replicate the results obtained in the current implementation of NEURON and to test the new ideas proposed here.

We first checked the order of convergence in space of the method, given the particular boundary condition at branching nodes (connecting the values of v on different sections) and the synapse point processes that produce singularities in the solution. The numerical results (see Figure 3) confirm that with respect to the L^2 norm, quadratic convergence in v_h and linear convergence in ∇v_h is preserved.

Regarding the order of convergence in time, we tested the different techniques presented in Section 5 first with alpha-synapses (where $g_{syn} = g_{syn}(y)$) and then with AMPA/NMDA synapses (where $g_{syn} = g_{syn}(y, v)$), and measured empirically the order of convergence of the global error in $\boldsymbol{v}_\tau(T)$.

In Figure 4a we show the results obtained for alpha-synapses. In particular, we see that without discontinuity handling, such as in the current NEURON implementation, the order of convergence is severely degraded. Our technique of discontinuity detection with stopping and restarting of the integration at the spike arrival time t_{rec} allows us to recover a full order 2 convergence.

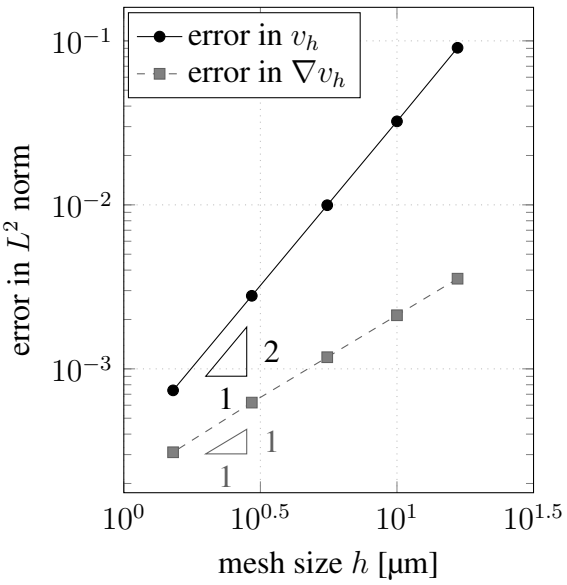
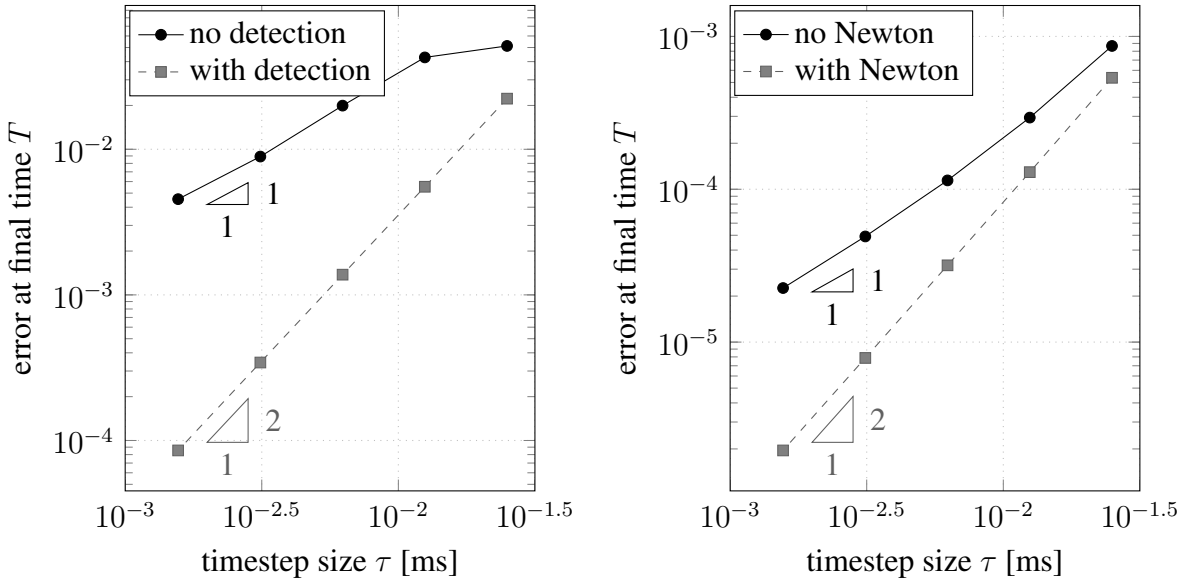


Figure 3: Convergence in space looking at the errors $\|v - v_h\|_{L^2}$ and $\|\nabla v - \nabla v_h\|_{L^2}$.



(a) Alpha-synapse

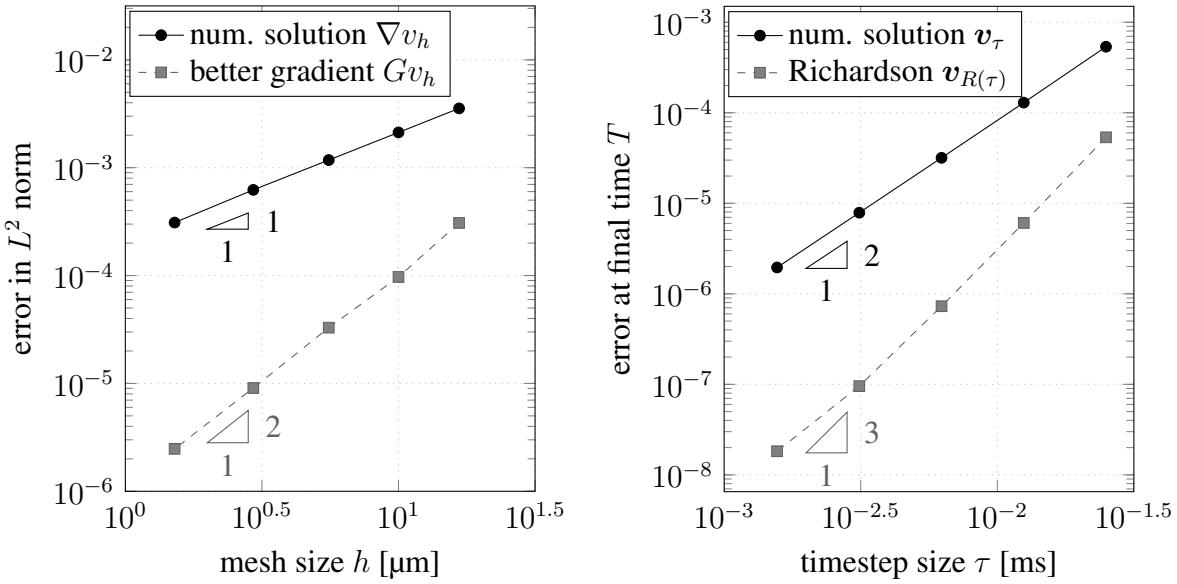
(b) AMPA/NMDA synapse

Figure 4: Convergence in time looking at the global error $\|\boldsymbol{v}(T) - \boldsymbol{v}_\tau(T)\|_2$ for different kinds of synapses and different numerical methods.

Figure 4b presents the results for the more realistic AMPA/NMDA synapse model. In this case we only considered methods incorporating the discontinuity detection technique. The simple technique adopted in NEURON (see Section 5.1) to resolve the nonlinearity still present after Strang splitting is shown to fail, reducing the method's convergence to order 1. However, the method we proposed based on performing only one step of Newton iteration to solve the nonlinear equation proves to be effective and restores quadratic convergence.

To summarise, the quadratic convergence in time is maintained through the use of the integration methods described in Section 5.2 that more effectively handle discontinuities and

nonlinearities in the system of ODEs, and is preserved when the computationally expensive exponential of τB is replaced by the much cheaper Heun method (19).



(a) ZZ patch recovery

(b) Richardson extrapolation

Figure 5: Superconvergence for error estimator in space (Zienkiewicz–Zhu patch recovered gradient) and in time (Richardson extrapolation).

Finally, we tested the validity of the post-processing methods for the estimation of the errors in space and time. In general, we define the *efficiency index* of an error estimator err^* for the true error err as the ratio

$$\theta = \frac{err^*}{err}. \tag{22}$$

It is desirable that the efficiency index $\theta \to 1$ at least linearly as the discretization is more and more refined. A sufficient condition to attain this goal is that the post-processed *better* solution is superconvergent with respect to the bare numerical solution itself. In Figure 5 we can can see that the better gradient obtained using Zienkiewicz–Zhu patch recovery and the better trajectory obtained using Richardson extrapolation are superconvergent with order $O(h^2)$ and $O(\tau^3)$, respectively, while the bare numerical solution converge with order $O(h)$ and $O(\tau^2)$, respectively. This validates our choice of error estimator.

7 CONCLUSIONS

NEURON is one of the most popular tools in the field of computational neuroscience for the simulation of morphologically detailed neural networks. We presented here an analysis of the diverse sources of error that can affect the results of NEURON simulations. Particular attention has been devoted to the numerical error arising from space and time discretization of the differential equations representing the mathematical model of the evolution of the membrane potential in a dendrite.

We examined the finite difference scheme and the numerical integration methods used in the current implementation of NEURON, and performed numerical experiments to investigate the orders of convergence of these schemes. In particular, for the time integration method we observed that the expected second order convergence was spoiled by the presence of singularities

in the solution and due to the simplifications introduced to eliminate the nonlinearities arising from the AMPA/NMDA synapse model.

We have presented here some techniques for handling the discontinuities in the system of ordinary differential equations and for addressing the nonlinearity in the AMPA/NMDA synapse model. These techniques allow the recovery of quadratic convergence without introducing extra computational cost. Moreover, we proposed the use of a second order Padé approximation for the computation of the ion channel gate states which reduces the computational cost without loss of precision.

Error estimators based on Zienkiewicz–Zhu patch recovery and Richardson extrapolation were presented, allowing a posteriori estimations of the errors introduced in space and time.

These methods have been validated numerically on a simple branched neuron model, demonstrating the gains in convergence order over the existing NEURON implementation. We plan to incorporate these methods into a future release of the NEURON simulator.

8 ACKNOWLEDGEMENTS

The work was supported by the EPFL Blue Brain Project Fund and the ETH Board funding to the Blue Brain Project.

REFERENCES

[1] Blackman, Arne V., et al. “A comparison of manual neuronal reconstruction from biocytin histology or 2-photon imaging: morphometry and computer modeling.” *Frontiers in neuroanatomy* 8 (2014).

[2] Bossetti, Chad A., Merrill J. Birdno, and Warren M. Grill. “Analysis of the quasi-static approximation for calculating potentials generated by neural stimulation.” *Journal of neural engineering* 5.1 (2007): 44.

[3] Carnevale, Nicholas T., and Michael L. Hines. *The NEURON book*. Cambridge University Press, 2006.

[4] Chung, Kwanghun et al. “Structural and molecular interrogation of intact biological systems.” *Nature* 497.7449 (2013): 332-337.

[5] Druckmann, Shaul, et al. “A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data.” Frontiers in neuroscience 1 (2007): 1.

[6] Goldman, Daniel, and Tasso J. Kaper. “N th-order operator splitting schemes and nonreversible systems.” *SIAM journal on numerical analysis* 33.1 (1996): 349-367.

[7] Hairer, Ernst, Syvert Paul Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations I: Nonstiff problems*. Springer Verlag, 2008.

[8] Hairer, Ernst, Syvert Paul Nørsett, and Gerhard Wanner. *Solving Ordinary Differential Equations II: Stiff and differential-algebraic problems*. Springer Verlag, 2010.

[9] Hines, Michael. “Efficient computation of branched nerve equations.” *International journal of bio-medical computing* 15.1 (1984): 69-76.

[10] Hodgkin, Alan L., and Andrew F. Huxley. "A quantitative description of membrane current and its application to conduction and excitation in nerve." *The Journal of physiology* 117.4 (1952): 500.

[11] Krichmar, Jeffrey L., et al. "Effects of dendritic morphology on CA3 pyramidal cell electrophysiology: a simulation study." Brain research 941.1 (2002): 11-28.

[12] Lindsay, K. A., J. R. Rosenberg, and G. Tucker. "From Maxwell's equations to the cable equation and beyond." *Progress in biophysics and molecular biology* 85.1 (2004): 71-116. w to denote units of dimensionless

[13] Larsson, Jonas. "Electromagnetics from a quasistatic perspective." *American Journal of Physics* 75.3 (2007): 230–239.

[14] Markram, Henry, et al. "Reconstruction and simulation of neocortical microcircuitry." *Cell* 163.2 (2015): 456-492.

[15] Ovcharenko, Aleksandr, et al. "Simulating Morphologically Detailed Neuronal Networks at Extreme Scale.", ParCo15 Conference, Edinburgh, Scotland, 2015.

[16] Plonsey, Robert, and Dennis B. Heppner. "Considerations of quasi-stationarity in electrophysiological systems." *The Bulletin of mathematical biophysics* 29.4 (1967): 657-664.

[17] Quarteroni, Alfio, and Alberto Valli. *Numerical approximation of partial differential equations*. Vol. 23. Springer Science & Business Media, 2008.

[18] Wang, Yun, et al. "Anatomical, physiological, molecular and circuit properties of nest basket cells in the developing somatosensory cortex." *Cerebral Cortex* 12.4 (2002): 395-410.

[19] Sherman-Gold, Rivka. The Axon Guide for Electrophysiology & Biophysics: Laboratory Techniques. Axon Instruments, 1993.

[20] Van Geit, Werner, Erik De Schutter, and Pablo Achard. "Automated neuron model optimization techniques: a review." Biological cybernetics 99.4-5 (2008): 241-251.

[21] Zienkiewicz, O. C., and J. Z. Zhu. "The superconvergent patch recovery (SPR) and adaptive finite element refinement." *Computer Methods in Applied Mechanics and Engineering* 101.1 (1992): 207-224.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EVALUATION OF SPARSE LINEAR ALGEBRA OPERATIONS IN TRILINOS

Mohammad Siahatgar[1,2], **Gabriele von Voigt**[1,2]

[1] Leibniz Universität IT Services
Leibniz Universität Hannover
Schloßwender Straße 5, 30159 Hannover, Germany
e-mail: {siahatgar, vonvoigt}@luis.uni-hannover.de

[2] L3S Research Center
Appelstrasse 9a, 30167 Hannover, Germany

Keywords: Performance Evaluation, Sparse Algebra, Energy Consumption, Trilinos, HPC

Abstract. *The performance of numerous scientific libraries and applications depends heavily on efficiency of sparse linear algebra operations. In this paper, we survey the performance of several parallel sparse vector and matrix kernels provided in the Trilinos framework on supercomputer systems Cray XC30/40 and IBM Blue Gene/Q. The linear algebra operations in Trilinos are handled by one of the two packages Epetra or Tpetra. While the former is the most-used, the latter is the target of future developments and supports larger scale problems as well as shared memory parallelism. We compare the results obtained from both packages together with the MPI only and hybrid solutions. The hybrid parallelism is managed by the package Kokkos, which aims for performance portability among different architectures. We report the efficiency of a single node of the system and demonstrate the scalability behavior of the benchmarks up to 38,400 cores of the HLRN-III systems. Furthermore, for the Intel processors used in the Cray system we present measurements of the energy consumption of the kernels and compare the Energy-to-Solution between different compilers and parallel programing paradigms. In addition, we discuss the effect on the performance and the energy consumption by linking the vendor provided libraries compared to the user-compiled versions. These extensive comparisons obtained on the top most performant supercomputer systems help users and developers as a starting point for determining an optimal development strategy.*

1 INTRODUCTION

Mathematical libraries contribute at the core level to any scientific application and compose a basic framework for scientific computing. Many complex but well-known and often-used algorithms have been already implemented, tested and are ready to be incorporated in user applications. With an increasing number and coverage, a key question is to determine the importance and effectiveness of any library. Developers often carefully consider the functionality of the libraries regarding the desired capability besides their performance on their target hardware architecture. In many cases, the latter point can be important and even is the decisive factor. For example, quantum chemistry and molecular dynamics packages benefit much from GPUs, since most of their calculations are in essence stream processing. Furthermore, great efforts are being made to optimize the essential and low-level libraries needed for general applications. Examples are the BLAS and LAPACK libraries, of which many HPC vendors (such as IBM, Intel, Cray, AMD, Nvidia and Intel) are providing highly tuned versions for their hardware. Moreover, the dependency as well as cooperation between related libraries is constantly increasing. With a growing number of options (in both software and hardware terms), evaluation of libraries is becoming more and more relevant in the lifetime of scientific applications. Such evaluations should help developers to make informed decisions about the value of a particular library in regard to their objectives.

Sparse linear algebra calculations play a significant role in many algorithms in large-scale computational codes. The sparse matrix-vector and matrix-matrix multiplications arise in many contexts such as iterative solvers, eigenvalue problems and graph algorithms. These two key kernels dominate the performance of many solvers and may easily correspond to the most of the total run time. Thus studying the behavior of the implementations on top of the hardware (such as many-core, many-node and hybrid systems) is highly influential concerning the performance of the application. Numerous efforts [1-6] have been made to implement these kernels with optimized algorithms and tuned to specific memory and processor architectures.

In this paper, we have evaluated several sparse vector and matrix operations provided from the Trilinos library. The Trilinos project [7,8] from Sandia National Laboratories aims to develop and implement robust algorithms and software packages within an object-oriented framework for the solution of large-scale, complex multi-physics engineering and scientific problems in parallel. It consists of many self-contained packages each designed for specific objectives. The project also helps leveraging the well-established libraries such as BLAS and LAPACK with emphasizing abstract interfaces for maximum flexibility of component interchanging. Furthermore, many applications are standardizing on the Trilinos APIs, resulting to access of all Trilinos solver components as well as underlying libraries without unnecessary interface modifications. Several supercomputing facilities provide an installed version of Trilinos for their users. For example, Cray supercomputers come with Trilinos installed as part of the *Cray Scientific and Math Libraries* [9]. This release includes the *Cray Adaptive Sparse Kernels* (CASK) which are tuned for the Cray environment, and provide improved performance over native installation under most circumstances especially for operations with very sparse matrices [10].

Trilinos relies on two of its packages which implement most of the linear algebra capabilities, namely Epetra and Tpetra. The package Epetra has been the core linear algebra package, but due to its limitations in design and lack of support for large systems and new architectures, the package Tpetra is the target of future developments. In summary, their features can be compared by the following criteria:

- **Data-type**: Epetra supports only double-precision floating-point operations (of C++ type *double*). Other types such as complex numbers or higher precision floating-points are exclusively in the scope of Tpetra.
- **Very large problems**: Initially, Epetra used the C++ *int* data-type for storing global and local indices (32 bits), therefore limiting the maximum number of degrees-of-freedom in the problem to less than 2 billion. Needless to say, this limit is undesirable for high performance applications and the current systems can go beyond this limit. A workaround has been introduced [11] for the support for 64-bit global indices (C++ data-type *long long*) in Epetra and other depending packages. Tpetra has also optional support for 64-bit local indices, which allows that many entities within a single MPI process.
- **Shared Memory Parallelism**: While supporting MPI parallelism, Epetra only supports OpenMP shared memory parallelism for a few kernels but usable up to a few threads per MPI rank. Tpetra has a more general "MPI+X" approach, where X stands for various shared memory parallel programming models. Tpetra implements this in a performance-portable way by using the Kokkos shared memory parallel programming package.

The standalone package Kokkos [12] is the Trilinos foundation for thread-scalable parallelization. It implements a C++ template library for writing performance-portable applications targeting major HPC platforms. For that purpose, Kokkos provides abstractions for both parallel execution of code and data management. The goal of the package is to implement a performance-portable shared memory parallel programming model and data containers, in order to let users to write an algorithm once, and then just change a template parameter to get the optimal data layout for a particular hardware.

2 BENCHMARKS

In this paper, we have surveyed parallel sparse linear algebra kernels provided by both packages Epetra and Tpetra. For the Epetra benchmark, we have used the Epetra Kernels Benchmark driver from the Mantevo project [13]. Mantevo proxy applications are open source software and the Epetra benchmarks is available for download under a GNU Lesser General Public License (LGPL). This driver executes key performance-impacting Epetra kernels for sparse matrix-vector and sparse matrix-multivector kernels. Additionally, we have refactored a version of this benchmark to use Tpetra. The operations performed are calculation of sparse vector 2-norm (NORM), dot product of two vectors (DOT), generalized sparse vector addition of the form $\alpha x + y$ (AXPY), and sparse matrix-vector (SpMV) as well as sparse matrix-matrix (SpMM, with 2, 4, and 8 columns) products (with same global dimensions for vectors and matrices). Among these, SpMV product is by far the most computationally expensive component of sparse iterative linear solvers [3]. SpMV is a memory bandwidth bound problem and its performance changes highly with respect to matrix characteristics. One should note that there is no "General Purpose" SpMV code, since the applicability of an optimization applies only to types of matrices obtained in different problem or set of problems. For our benchmarks, we have constructed a two dimensional 25-point finite difference stencil matrix, which is stored in the Compressed Row Storage (CSR) format.

2.1 Hardware resources

We have evaluated the performance of the two aforementioned benchmarks and compared the results in two supercomputing centers HLRN-III and JUQUEEN, ranked 83rd and 11th in the Top 500 list of November 2015, respectively [14]. The former system is operated by North-German Supercomputing Alliance (Norddeutscher Verbund zur Förderung des Hoch- und

Höchstleistungsrechnens – HLRN) and is a distributed supercomputer system hosted at two locations Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) and Leibniz Universität IT Services (LUIS, formerly RRZN) at University of Hannover. The HLRN-III system consists of two almost identical complexes based on Cray XC30/XC40 supercomputers and the total system comprises about 87,500 cores with a total peak performance of 2.7 PFLOP/s. The latter system is an IBM Blue Gene/Q system built for the Jülich Research Centre (Forschungszentrum Jülich – FZJ) with 458,752 cores and overall peak performance of 5.9 PFLOP/s.

The benchmarking on the Cray system is performed at the Hannover complex, which composes of two Cray systems with Intel processors and Cray Aries interconnect [15]. The nodes of the Cray XC30 system contain two Intel Xeon Ivy Bridge CPUs (E5-2695v2) operated at 2.6 GHz while in the Cray XC40 system the Intel Xeon Haswell CPUs (E5-2680v3) operated at 2.5 GHz are used. The IBM Blue Gene/Q system at Jülich uses IBM PowerPC A2 processors at 1.6 GHz with 16 cores per node [16]. The energy usage of each rack (1024 nodes, 16384 cores) without jobs running is about 30 kW and on load on average 60-70 kW [16]. The peak consumption of each A2 processor is about 55 W. This is more or less half the TDP value of the Intel Xeon CPUs installed at HLRN-III systems (with 12 cores per CPU). In comparison, the theoretical peak performance for the A2 processor is 204.8 GFLOP/s, and 230 GFLOP/s for the Intel Xeon Ivy Bridge CPUs at HLRN. For the comparison between two systems, we executed the benchmarks on subset of two systems with similar number of physical cores, remarking that the PowerPC A2 cores are capable of four-way multithreading.

2.2 Details of execution

For the Cray system, both benchmark compiled against a local installation of Trillions as well as the Cray provided library. The Trilinos library provided from Cray has the version 11.12 and supports only MPI parallelism. This is the latest version of Trilinos without the explicit requirement of C++11, which the Cray compiler (at the time of writing) does not support. For the local installation we used Trilinos version 12.2.1 linked against the Cray LibSci library [9] for BLAS and LAPACK functions, which includes the tuned CASK library. The codes have been compiled on Cray systems using Intel compiler version 15.0.1 and GNU compiler version 4.9.1. The architecture dependent instruction sets were enabled (i.e. AVX for Ivy Bridge and AVX2 and FMA for Haswell) and optimization flag –Ofast was used in all compilers. Additionally, 8 MB huge pages were also allocated. The effect of using such larger memory pages was found to be substantial (e.g. more than 60% increase in performance using 6,000 MPI processes).

For the Epetra Benchmark, we have compiled both serial (MPI only, without the OpenMP flag) and hybrid (MPI+OpenMP) versions. Cray MPI library uses MPICH2 distribution from Argonne, enhanced by tunings for the Aries interconnect. For the Tpetra benchmark, similar versions were produced as well as another hybrid option using POSIX threads (MPI+Pthreads). For the hybrid versions, we ran the benchmarks on a single node with varying number of threads from one to the number of cores (enabling hyper-threading did not lead to any improvement in performance). In case of hybrid parallelism, a single MPI process per node (with no core binding) has been assigned. For the MPI only case, one core has been reserved for one MPI rank. For comparison of the performance of different benchmarks with different programming model (MPI, OpenMP, Pthreads), the matrix size per node has been kept constant (i.e. solving precisely the same equation with same number of FLOPs in each node). The FLOP/s is calculated by $2nmN^2 / t$, in which n is the number of point in the stencil (in our case 25), m the number of trials (on average 100), N the size of the global matrix and t the time of the calculation of the kernel. The local matrix size per node kept constant and of the order 10^3.

We compiled Trilinos as well as both benchmarks on the JUQUEEN system using GNU compiler version 4.8.1 linking to the ESSL library for BLAS and LAPACK functions. Each JUQUEEN node (16 physical cores) is capable of running 64 threads at the same time (since each core is 4-way simultaneously multithreaded). For the hybrid MPI+OpenMP case, we have used all those 64 threads, but for the MPI only version, due to the resulting limited amount of memory per process, we have used only 16 MPI ranks per node.

2.3 Energy consumption

The Cray supercomputer operated by HLRN provides users and system administrators with extensive tools to comprehensively measure and monitor the energy consumed by the programs and the system [17]. Power consumption is sampled periodically and is monitored at the blade level with each blade reporting usage of the nodes and the network. Data is aggregated by the cabinet controllers and reported out-of-band to the system management workstation where it is logged for subsequent processing. Information on the allocation of nodes to jobs and the CPU time is logged at the same time, enabling energy consumption to be added to the data collected for each job. This enables the system administrators to measure and account for the energy efficiency of the system on production workload and the project partners to make power consumption measurements and optimizing for both performance and power consumption. With the help of power management enabled compute nodes, users are also able to easily access the per-node power and energy data in real-time in a set of virtual files. These counters are updated at 10 Hz frequency, however polling them causes a system interrupt that may affect the user code, so their use is not recommended for high-frequency performance monitoring. We have used the same approach for measuring the energy measurement of the sparse kernels as in Refs. [18,19], and recorded the total energy used per node before and after each calculation.

3 RESULTS

In the following we present the measurements for the performance and energy consumption of the sparse linear algebra kernels.

3.1 Single Node of Cray XC30/40

The results of the execution of the benchmarks for several sparse kernels on a single node of both Cray XC30 and XC40 systems using Epetra and Tpetra packages are shown in Figure 1. Performance from using all 24 cores of both MPI and OpenMP only versions compiled by different compilers have been shown in the plots. The GNU compiler surprisingly outperforms other compilers, with Cray compiler catching up in a few cases. Furthermore, the MPI only version performs the best. Tpetra (with help of Kokkos) incorporates better use of cache locality leading to remarkably better performances compared to Epetra. For the shared memory case, Kokkos is also able to manage POSIX threads directly, and achieving better performance in some cases compared to OpenMP.

Figure 2 corresponds to the same calculation but from the point of view of the energy consumed. MPI only version of Tpetra wins regarding the lowest energy consumption (mostly due to having less total run time). GNU and Cray compiled executables also consume the lowest power, signifying that the Cray compiler is heavily optimized for low-energy consumption for the CPUs used for the Cray systems.

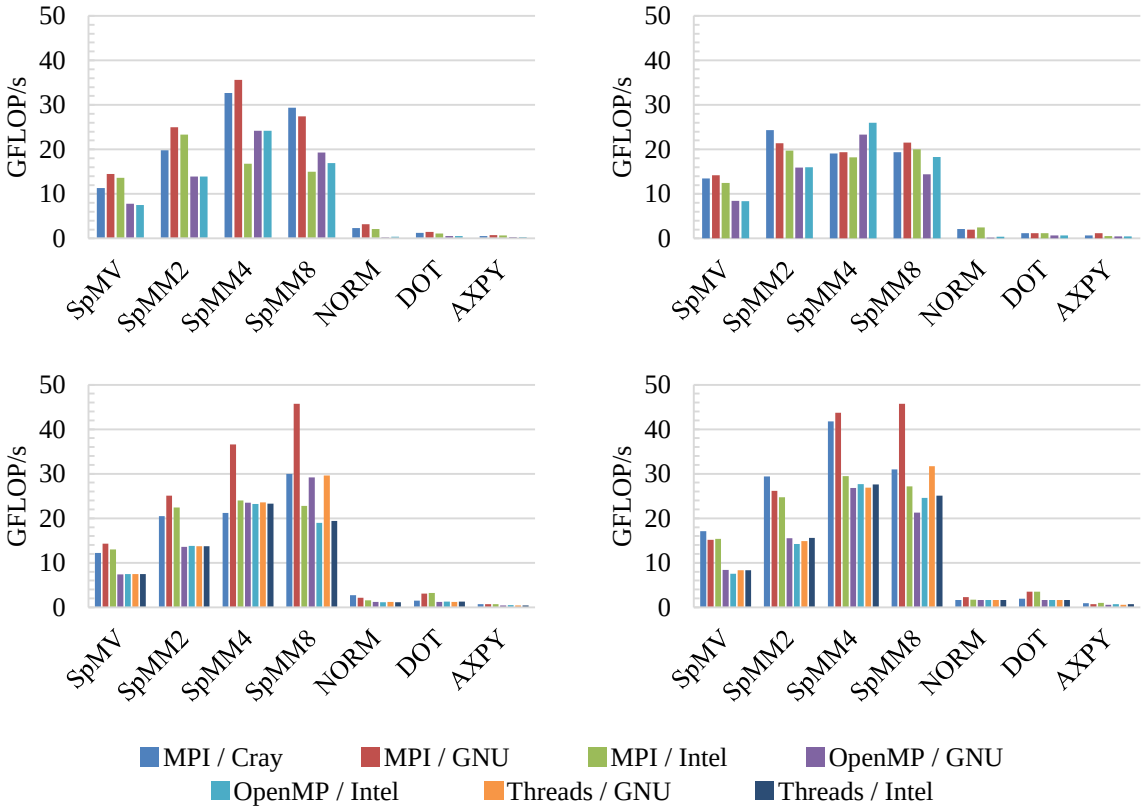


Figure 1. The performance of various sparse linear algebra kernels on one node of Cray XC30 (left) and Cray XC40 (right) using the Epetra (top) and Tpetra (bottom) packages.

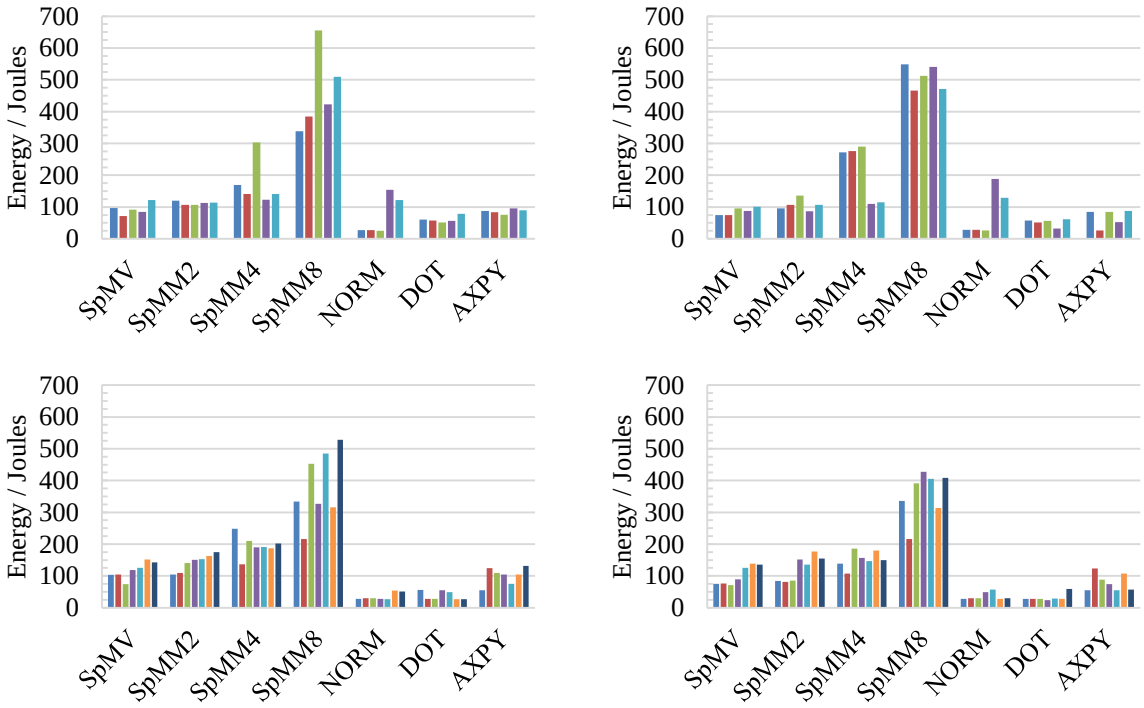


Figure 2. The energy consumption of kernels. The legend and the order of plots are the same as Figure 1.

To see the performance using the shared memory parallelism, we have also run the benchmarks with varying number of threads. Figure 3 demonstrates the scalability of the sparse matrix-multivector (of length 8) multiplication kernel on a single node of both Cray XC systems. The code uses threads from 1 to 24 (maximum per node).

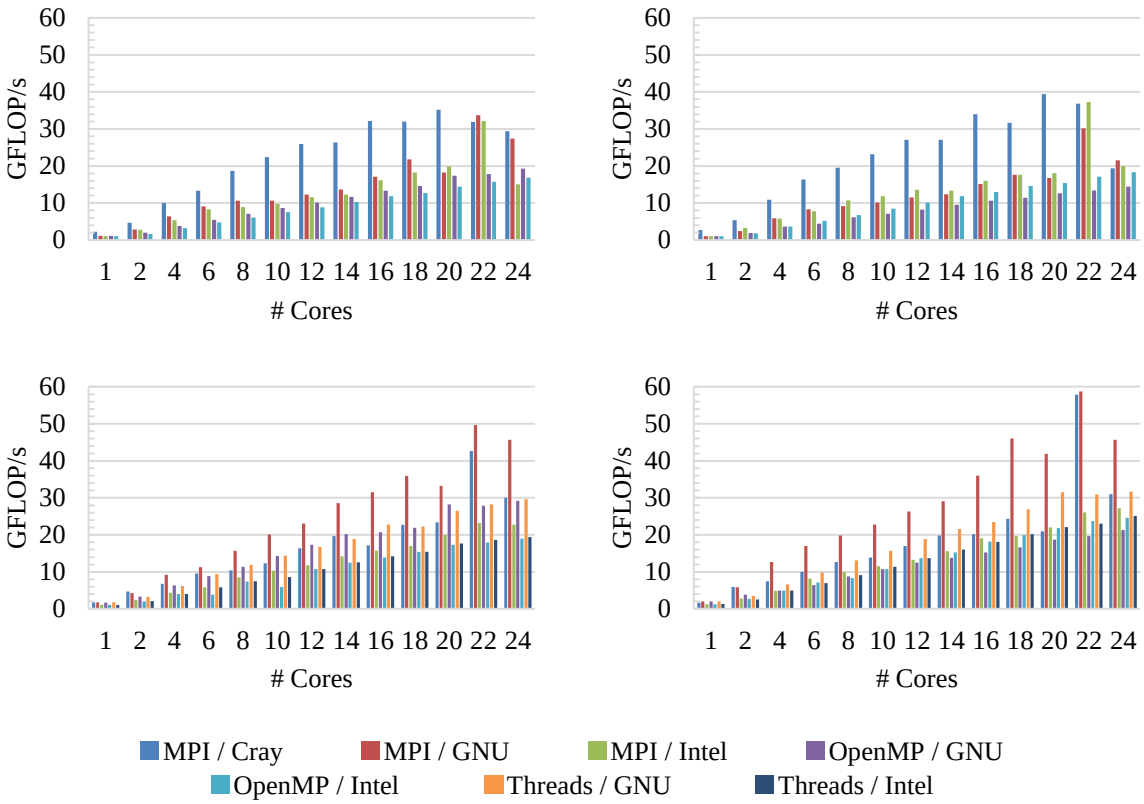


Figure 3. The performance of SPMM8. The order of plots is the same as Figure 1.

The Tpetra version of the benchmark compiled with GNU compiler using only MPI on the Haswell CPU is the winner in overall performance, with the Cray provided library being a close second. This is a hint to show that migrating from Epetra to Tpetra (even in case of not using shared memory parallelism) improves the performance. In the case of Epetra benchmark, the intrinsic support of CASK (included in Cray LibSci) leads to better performance.

It is also interesting to look at the energy consumption of the same operation in Figure 4. MPI only version of Tpetra again wins regarding the lowest energy consumption. Obviously allocating more cores decreases the energy consumption of the node, but with using more than half the cores, the total energy consumed by the node does not change considerably.

3.2 Many nodes

Here we present the performance results of running the benchmarks in parallel on the Cray XC30 system in Hannover. We have used the Cray provided library as well as our own installation, and using MPI only and MPI+OpenMP for both benchmarks. In the former case, all the cores within a (single) node have been assigned one MPI rank, and in the latter, one MPI rank is used per node with 24 OpenMP threads. Based on the results from single node performance, we omitted using the Intel compiler. The results of the weak scaling is presented in Figure 5. For Epetra, the Cray provided library performs the best, while the Tpetra delivers better overall performance using GNU compiler. The last point in the plot corresponds to using 1600 nodes

(38,400 cores), which is the largest partition in the Hannover system combining both Cray XC30 and XC40 systems. For that, the executable compiled for the Ivy Bridge CPUs has been used for all the nodes. As expected and also similar to the single node performance results, the MPI+OpenMP hybrid parallel benchmark is falling behind.

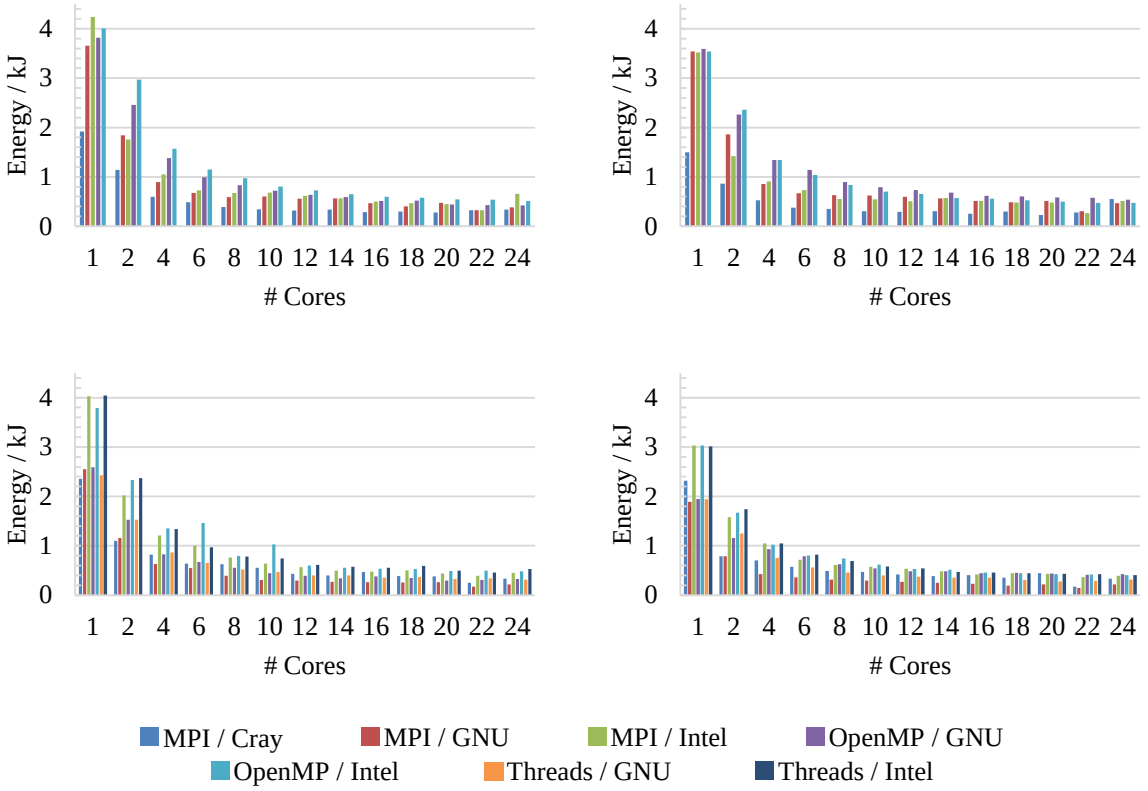


Figure 4. The energy consumption of SPMM8. The order of plots are the same as Figure 1.

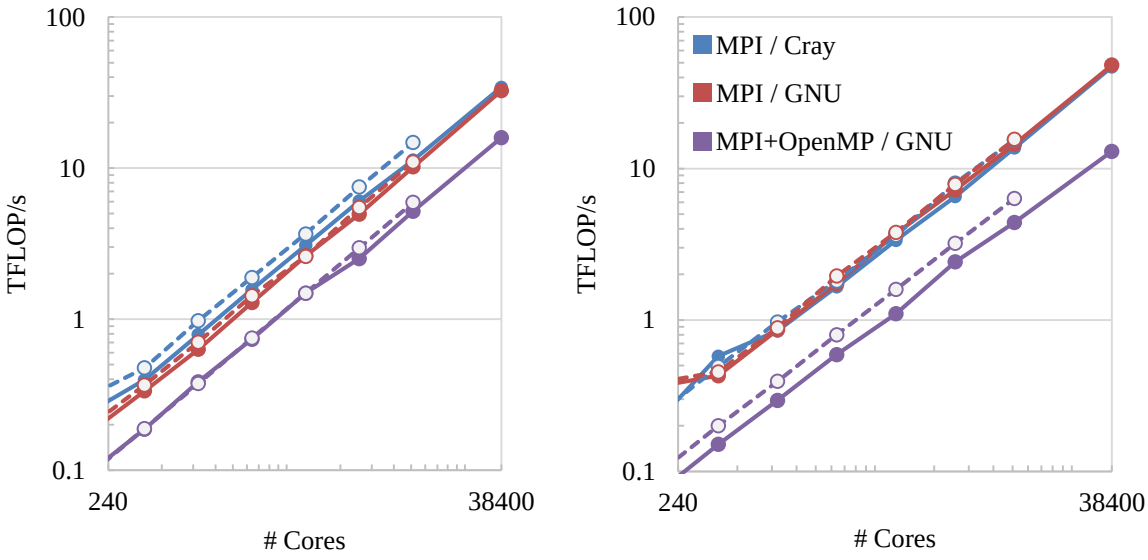


Figure 5. The weak scaling plot of SPMM8 using the Epetra (left) and Tpetra (right).
The solid (dashed) line belongs to the results from the Cray XC30 (XC40) system.

The MPI only Epetra version scales well and behaves in accordance with a single-node performance model. However, Tpetra performs remarkably better than Epetra, therefore, this can be seen as practically enough reason to suggest the migration from Epetra to Tpetra. With this move, support for shared memory parallelism also comes for free.

3.3 Comparison of Cray XC30 and IBM Blue Gene/Q

We have compared the performance of the Ivy Bridge CPUs with the PowerPC A2 processors for the exact same calculation on a similar number of physical cores on both systems. Table 1 contains the details of the execution of benchmarks.

Parallelism	# Nodes	# MPI Ranks per Node	# Threads per MPI Rank	# Cores	Epetra	Eff. %	Tpetra	Eff. %
IBM Blue Gene/Q								
MPI	64	16	1	256	72.2	0.6	193	1.5
MPI+OpenMP	64	16	4	1024	**216**	**1.6**	**460**	**3.5**
	64	8	8	1024	206	1.6	438	3.3
	64	4	16	1024	190	1.4	414	3.2
	64	1	64	1024	131	1.0	295	2.3
Cray XC30								
MPI	64	16	1	1024	**1,470**	**5.0**	**2,210**	**7.5**
MPI+OpenMP	48	6	4	1152	756	3.4	1,060	4.8
	48	3	8	1152	744	3.4	921	4.2
	64	1	16	1024	603	2.0	590	2.0
	64	1	24	1536	614	2.1	558	1.9

Table 1: Comparison of Cray XC30 and IBM Blue Gene/Q benchmarks. Performance is in GFLOP/s and the efficiency percentage is the ratio to the peak performance. The topmost performance is shown in boldface.

The results are from the same matrix problem on both systems using different parallelization techniques. The topmost FLOP/s in each system is denoted in bold. The same data has been shown in Figure 6, with the percentages displaying the ratio of the performance of the Blue Gene/Q to the Cray XC30 system (for a comparable total number of cores). We have observed that the Intel CPUs are more efficient and outperform PowerPC A2s for sparse matrix vector operations (despite consuming more power). Unlike the Cray system, hybrid parallelization leads to much better performance compared to the MPI only case. In practice, it is possible to fill up all four threads of each core of the A2 processor with MPI processors but the memory overhead from the MPI library limits the performance of the benchmark. On the other hand, the thread management overhead from using only OpenMP per node decreases the performance. Also, the Blue Gene/Q chip is able to use idle threads for asynchronous progress on MPI. We have found that one MPI rank per core and four OpenMP threads gives the best performance. Also the performance improvement of Tpetra over Epetra kernels is significantly larger compared to the Cray system.

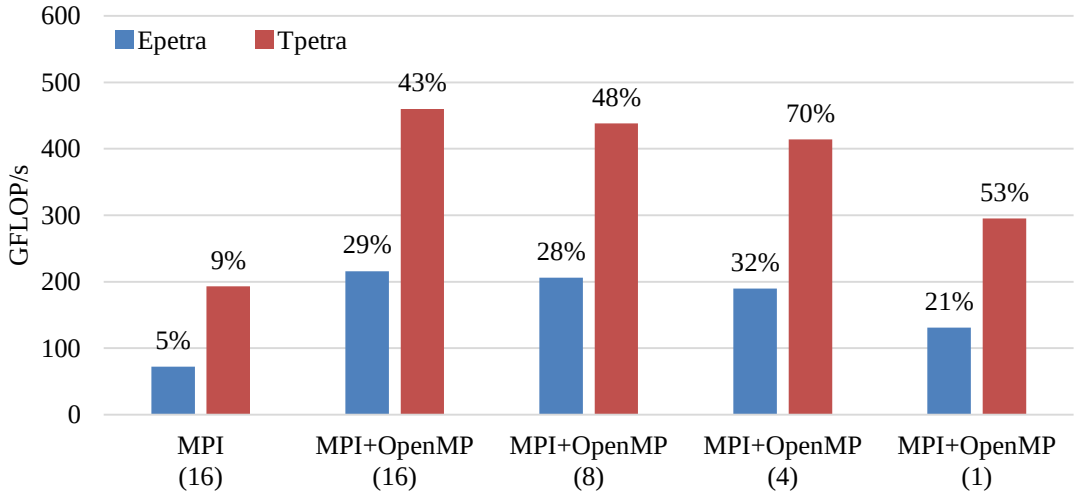


Figure 6. The comparison between the performance of SPMM8 kernel in IBM Blue Gene/Q and Cray XC30. The percentage show relative ratio of performance between two platforms. Number of MPI ranks per node is displayed in parenthesis.

4 CONCLUSIONS

In this article, we have surveyed the fundamental packages of Trillions for linear algebra operations, namely Epetra and Tpetra, regarding their real-life performance. We performed several sparse linear algebra operations using two packages on leading supercomputing systems, Cray XC30/40 and IBM Blue Gene/Q. The results from different compilers with tuned libraries have been compared suggesting that the MPI only Tpetra benchmark on the Cray XC system is getting better performance and efficiency. We have also reported results for actual energy consumed by the processors of the Cray system during the run time of the sparse algebra calculations.

Using different hardware architectures and software parallelism models, our results present examples in order to help decision making in incorporation of Trilinos in other scientific applications. The Tpetra package and its underlying package Kokkos show much improved performance and functionality over its counterpart standard linear algebra package, Epetra. We strongly advise the developers to plan accordingly in order to refactor their code base into Tpetra as well using other packages of Trilinos which are going through the same path. By migrating from Epetra to Tpetra and from internal OpenMP development (or similar programming models) to Kokkos, one can expect short term benefits (such as increase in performance) as well as opening the path for using extended set of architectures. Kokkos can also help in reaching more portability (along with performance portability) in shared memory parallelism.

5 ACKNOWLEDGEMENTS

This work has been funded by the European Research Council under the FP7 NUMEXAS project under grant agreement 611636. The authors gratefully acknowledge the Gauss Centre for Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS share of the supercomputer JUQUEEN [16] at Jülich Supercomputing Centre (JSC). GCS is the alliance of the three national supercomputing centers HLRS (Universität Stuttgart), JSC (Forschungszentrum Jülich), and LRZ (Bayerische Akademie der Wissenschaften), funded by the German Federal Ministry of Education and Research (BMBF) and the German State Ministries for Research of Baden-Württemberg (MWK), Bayern (StMWFK) and Nordrhein-Westfalen (MIWF). The authors would like to thank the anonymous referee for the comments.

REFERENCES

[1] S. Williams, L. Oliker, R. Vuduc, J. Shalf, K. Yelick, J. Demmel, Optimization of sparse matrix-vector multiplication on emerging multicore platforms. *Proceedings of the 2007 ACM/IEEE conference on Supercomputing (SC '07)*, Article 38, 2007.

[2] A. Buluc, J. R. Gilbert, Challenges and Advances in Parallel Sparse Matrix-Matrix Multiplication, *Parallel Processing, 2008. ICPP '08. 37th International Conference on*, 503-510, 2008.

[3] G. Schubert, H. Fehske, G. Hager, G. Wellein, Hybrid-parallel Sparse Matrix-vector Multiplication with Explicit Communication Overlap on Current Multicore-based Systems, *Parallel Process. Lett.* **21**, 339-358, 2011.

[4] J.D. Davis, E.S. Chung, SpMV: A Memory-Bound Application on the GPU Stuck Between a Rock and a Hard Place, *Microsoft Technical Report*, MSR-TR-2012-95, 2012.

[5] K. Akbudak, E. Kayaaslan, C. Aykanat, Hypergraph Partitioning Based Models and Methods for Exploiting Cache Locality in Sparse Matrix-Vector Multiplication, *SIAM Journal on Scientific Computing*, **35**(3), C237-C262, 2013.

[6] S. Dalton, L. Olson, N. Bell, Optimizing Sparse Matrix-Matrix Multiplication for the GPU, ACM Transactions on Mathematical Software (TOMS), **41**(4), Article 25, 2015.

[7] M.A. Heroux et al, An overview of the Trilinos project, *ACM Trans. Math. Softw.*, **31** (3), 397-423, 2005.

[8] http://www.trilinos.org/

[9] Cray Inc., Cray Programming Environment User's Guide, S-2529-116, 2014.

[10] Cray Inc., Cray Application Developer's Environment User's Guide, S-2396-610, 2012.

[11] C. Jhurani, T. M. Austin, M. A. Heroux, J. M. Willenbring, Supporting 64-bit global indices in Epetra and other Trilinos packages - Techniques used and lessons Learned, Sandia National Laboratories, SAND2013-5928, 2013.

[12] C. R. Trott, M. Hoemmen, S. D. Hammond, H. C. Edwards, Kokkos - The Programming Guide, Sandia National Laboratories, SAND2015-4178 O, 2015.

[13] M.A. Heroux et al, Improving Performance via Mini-applications, Sandia National Laboratories, SAND2009-5574, 2009.

[14] http://www.top500.org/

[15] https://www.hlrn.de/home/view/System3/CrayHardware

[16] Jülich Supercomputing Centre, JUQUEEN: IBM Blue Gene/Q Supercomputer System at the Jülich Supercomputing Centre, *Journal of large-scale research facilities*, **1**, A1, 2015.

[17] Cray Inc., Monitoring and Managing Power Consumption on the Cray XC System, S-0043-7204, 2015.

[18] G. Fourestey, B. Cumming, L. Gilly, T.C. Schulthess, First Experiences with Validating and Using the Cray Power Management Database Tool. *Proceedings of CUG2014*, Seattle, WA, USA, 2014.

[19] A. Hart, H. Richardson, J. Doleschal, T. Ilsche, M. Bielert, M. Kappel, User-level power monitoring and application performance on Cray XC30 supercomputers, *Proceedings of CUG2014*, Seattle, WA, USA, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

AUTOMATIC VISUALIZATION AND CONTROL OF ARBITRARY NUMERICAL SIMULATIONS

Jan P. Springer and **Helen Wright**

University of Hull

Keywords: Computational Steering, Markup Languages, Schema, Data Binding, Visualization

Abstract. *Visualization of numerical simulation data has become a cornerstone for many industries and research areas today. There exists a large amount of software support, which is usually tied to specific problem domains or simulation platforms. However, numerical simulations have commonalities in the building blocks of their descriptions (e. g., dimensionality, range constraints, sample frequency). Instead of encoding these descriptions and their meaning into software architecures we propose to base their interpretation and evaluation on a data-centric model. This approach draws much inspiration from work of the IEEE Simulation Interoperability Standards Group as currently applied in distributed (military) training and simulation scenarios and seeks to extend those ideas. By using an extensible self-describing protocol format, simulation users as well as simulation-code providers would be able to express the meaning of their data even if no access to the underlying source code was available or if new and unforseen use cases emerge.*

A protocol definition will allow simulation-domain experts to describe constraints that can be used for automatically creating appropriate visualizations of simulation data and control interfaces. Potentially, this will enable leveraging innovations on both the simulation and visualization side of the problem continuum.

We envision the design and development of algorithms and software tools for the automatic visualization of complex data from numerical simulations executed on a wide variety of platforms (e. g., remote HPC systems, local many-core or GPU-based systems). We also envisage using this automatically gathered information to control (or steer) the simulation while it is running, as well as providing the ability for fine-tuning representational aspects of the visualizations produced.

1 Introduction

Simulation and visualization of phenomena and processes has become an integral part in our approach to understand and manipulate the world around us. We increasingly rely on computer-based modelling and simulation in such diverse areas as engineering, fundamental research, economics, or even politics. For some time such simulations have been augmented by our ability to provide diverse visual representations based on simulation results improving comprehension of the underlying model, data, or processes. In some cases visualization is the crucial, enabling factor for comprehension and decision making. While for many problems simulation and just visualizing the results works fine, there are also problem domains where it is beneficial to change aspects of the simulation parameters while the simulation is running, likely based on an expert's analysis of visualized simulation results. In essence the unidirectional process of providing simulation results for visualization ideally should be complemented by the ability to control (or *steer*) the simulation, providing a bidirectional (or even multidirectional) flow of data.

The general approach to computational steering provides software infrastructure or an environment where simulations can be embedded to run experiments. Usually interfaces are provided that allow for transporting control or steering information to a simulation as well as returning results to the visualization component. The visualization component also contains encoded knowledge of how to represent steering controls. Past as well as current approaches for computation steering frameworks are based solely on software architectures and often tied to specific pieces of technology.

We are interested in how the connection between a simulation, its steering mechanism, and its visualization can be decoupled. We argue that software-interface approaches result in an unnecessary dependency between the intent of a simulation, its possible visualization(s), and the actual technology, i. e. the particular software API, used for realizing that coupling. We aim instead towards a flexible interface that is self-contained, platform-agnostic, and allows for longevity of simulation codes for future use. This can be seen as a generalized *glue* between the simulation and the visualization provider(s).

Our approach is motivated by the Distributed Interactive Simulation (DIS) protocol, developed since the late 1980s and standardized in IEEE:1278.1:2012 [14], which allows participation of independently developed software in a running simulation as well as the *replay* of captured events from a simulation session. Because DIS provides a well-defined protocol, including specification of extensions without breaking applications not enabled for specific extensions, it remains the standard of choice for large-scale training simulation. However, DIS does not allow to discover extensions without developing software for their appropriate evaluation. We argue that this difficulty can be circumvented by combining a data-based approach similar to DIS with a markup-language infrastructure. An approach based on XML Schema definitions [28, 29] provides for self-describing schemas and auto-generated data binding and in summary it offers all the advantages of a DIS-like approach with none of its fixed-format disadvantages.

The remainder of this paper is organized as follows. In section 2 we summarize the important role of software-oriented solutions in computational steering. Section 3 illustrates the idea behind a data-centric approach with an artificial example of a simplistic but generic computational-fluid dynamics application. In sections 4 and 5 we discuss the merits of our proposal, including comparing it with other recent work utilising XML, before summarising and offering suggestions for future work.

2 Related Work

Computational steering has a long and distinguished record in the wider computational science literature. Steering a computation interactively necessarily involves visualization, and the focus of early attempts at computational steering was on how to achieve the interworking of potentially diverse tools. For example, GRASPARC [4] linked computational and graphical elements via a data management system based on a tree structure (a *history tree*) that captured both the results of the calculation and the progress of the investigation. The user interface reflected this same structure: branches of the tree recorded snapshots of parameters and data that could be used to pause, modify, and restart the calculation without having to re-run the simulation from the beginning. Different visualization components could be connected and demonstrators of the system utilised both in-house developments and off-the-shelf commercial products in this regard. In CSE [23], a steering interface was built from 2d and 3d graphics objects whose properties were used both to control parameters of the simulation *and* furnish a visualization of results. Parameters were accessed in the simulation code using a data I/O library designed to be minimally invasive. As in GRASPARC, a data manager was central to the architecture, but in CSE it orchestrated steering by issuing event notifications to subscribed visualization, transformation, and calculation processes. An emerging theme of this and other steering research was the need to instrument the application source code which led in turn to the development of a number of steering libraries for this purpose. Notable contributions that particularly addressed steering in grid-computing environments were the gViz library [5] and the RealityGrid library [19].

While the aim of the majority of such steering frameworks and libraries was to enable interaction with existing computational codes, in contrast SCIRun [18] provided an environment to make combined compute-and-visualization ensembles from scratch. The architecture was based on the dataflow paradigm combined with visual programming, which had hitherto seen application in modular visualization environments (MVEs) such as AVS [8], IRIS Explorer [24], and DX (IBM Visualization Data Explorer, later OpenDX) [1]. Such systems allow the user to build a visualization application, dragging-and-dropping pre-packaged modules to accomplish the constituent steps of first reading in the data, then mapping to some graphical representation and finally rendering this to the screen. Whereas these latter supported steering by adding existing simulation codes into the system as new modules, the SCIRun module set [22] included algorithmic and mathematical elements as standard. As such it is as much a software workbench for building integrated computational and visualization applications, as it is a system for simply visualizing results.

The difficulty that besets all computational steering applications, whether based explicitly on the dataflow paradigm or not, is how to transmit control information upstream from the display to the simulation. The user and their interaction tools (e. g., panels hosting sliders, menu selections) are logically found at the output end of the pipeline, but the recipients of these tools' outputs (e. g., changes to parameters, pause and restart commands) are at the input end, all the while producing data. This becomes especially problematic when the steering-visualization pipeline is distributed [31]. The solution adopted depends on the architecture of the steering framework: GRASPARC [4] evolved a system of control structures, separate from its data pathways; CSE [23] adopted an event-based model; MVE-based systems generally abandoned strict adherence to dataflow principles in favour of architectures incorporating feedback mechanisms for implementing loop constructs, and sometimes caching data to preserve state (q. v. [1]). A novel approach described in [6], which also targeted an MVE, was to implement an input-oriented pipeline in parallel

with the more usual results-bearing pipeline. This allowed a description of the simulation's exposed parameters to be automatically searched and the correct scalar, vector, or positional interactor to be placed in the image of the output, with the correct position and appropriate value constraints already applied. Dubbed *multi-purpose image interaction*, the idea is closely allied to that described in the present paper, except the aim now is to capture the essence of both simulation parameters *and* results, independently of the software used. Using a self-describing, data-centric protocol, we assert that essential knowledge about a steering application can be preserved, for example in the absence of the original source code or, provided its outputs have been appropriately curated, potentially even after the binary can no longer be run.

3 Example

In the following we develop an artificial example for a simplistic but generic computational-fluid dynamics (CFD) simulation using XML Schema to define control and result types and derive possible source code based on those definitions.

3.1 Requirements for a Generic CFD

Computational steering of a generic CFD simulation can be separated into two elements: firstly, how to control the simulation and, secondly, what kind of results are returned. In the simplest case we assume that the simulation will be based on computing particles per simulation time step. Listing 1 shows pseudo code to encode this.

```
struct particle {
  unsigned id;
  float3   position;
  float3   velocity;
};

struct result {
  double          timestamp;
  list<particle> particles;
};
```

Listing 1: Result structure for a generic CFD simulation returning a simulation result for a particular time step and the computed particles placed in a list.

The `timestamp` attribute in the `result` structure contains the time in (local) simulation time for which the particle properties have been computed. The actual particle properties are returned in the `particles` attribute of the `result` structure. The `particle` structure contains the `position` and `velocity` as 3-component floating-point values and a unique `id` for each particle instance. This `id` maybe used to create traces for particles of interest. To control the simulation a `control` structure can be used as shown in listing 2.

```
struct control {
  enum etype { start, stop, update, };

  etype  type;
  float3 gravity;
  float3 force;
}
```

Listing 2: Control structure for a generic CFD simulation allowing for specification of starting, stopping, or updating the simulation.

The `control` structure allows for specifying if the simulation should `start`, `stop`, or if the `control` structure contains an `update` to the parameter set. This is encoded by the `type` attribute and expressed as an enumeration. In our example the simulation-control parameters are simply `gravity` and `force` expressed as 3-component floating-point values. These parameters can be set for the initial setup of the simulation, i. e. `type == start`, or when updating simulation conditions, i. e. `type == update`. In case the `type` attribute is set to `stop` the simulation should simply shutdown, though for more sophisticated simulations shutdown parameters might be used as well.

Separating the result description returned by a simulation from its possible control allows for using these structures for both communication between the simulation and visualization as well as within the visualization to represent their content. However, this is tightly coupled to a specific example. Any changes to accommodate different simulations or even just variants with slightly different control parameters will require some software-development effort for both the actual simulation as well as the visualization. Additionally, parameter sizes, intended meaning, or problem-domain-specific constraints are usually hard to express and to document for automated extraction.

3.2 XML Schema Definitions

An XML Schema definition (XSD) [28, 29] allows to define content structure to be instantiated in an actual XML document. While this could be any kind of document we will use the notion of an XML document here as a way of transporting data between interested parties, i. e. the simulation and visualization provider. XSDs use standard XML notation and can be evaluated using existing XML support infrastructure. An XSD can be interpreted as a class description from the view point of an object-oriented software approach. XML documents then play the role of objects, i. e. actual instances of a class at run time, although this comparison is only partially correct because XSDs only allow the definition of member attributes but not member functions.

XSDs support a variety of built-in data types, e. g., `string`, `int`, `float`, as well as creating compound structures, e. g., `simpleType`, `complexType`. Within a compound structure element definitions based on primitive or user-defined types may appear. These elements can be constrained to restrict their range or to represent a discontinuous range such as an enumeration. XSD also supports extending existing structures as shown in appendix A, listings 3–6. Listing 3 shows an XSD for n-component floating-point values which can be used to represent n-dimensional vectors. These n-component floating-point types are used in listing 4 to define a particle having position and velocity in 3d similar to the pseudo code in listing 1. XSD also allows including other XSDs, which enables sharing of definitions between separate XSD files. Listing 4 also provides an example for defining a list of particles. In addition aliases in the form of root elements to the actual compound definitions are provided, which are necessary to allow the automatic generation of serialization code. In listing 5 base types for the control and result structures are defined, which are independent of any specific simulation. The control structure contains an enumeration to express its three possible states, i. e. start, update, and stop, while the result structure contains a time stamp. Listing 6 shows the definition of the control and result structures for the generic CFD example in turn extending the general definition of control and result structures in listing 5, essentially expressing the same content as in listing 2.

XSDs provide a template for actual instances in the form of XML documents as shown in appendix B, listings 7–10. Listings 7–9 show sample control-structure instances for controlling start, update, and stop of the simulation, respectively. In the XSD for the `cfd_control` structure, the `gravity` and `force` properties are defined withinin an `xs:sequence`. This

necessitates that these properties appear in the XML document even if they are not required (e. g., in case of submitting a `cfd_control` instance of `type` stop). To make properties such as `gravity` optional the `minOccurs` attribute can be set to zero, which, for larger aggregations of properties, allows for reducing the actual size of the XML document. Listing 10 shows a sample `cfd_result` structure returned by the simulation for each simulation step. The `timestamp` property provides the local simulation time at which the result was produced while the `particle_list` property contains the list of particles and their state at that time. The `timestamp` property is defined as a `xs:time`, which is expressed in hours, minutes, and seconds format. Note that the seconds part of `xs:time` is based on a double-precision floating-point type and allows for arbitrary precise resolution in the sub-second range.

XSDs as well as XML in general are a complex topic and cannot be treated in full detail in the context of this paper. We refer the interested reader to one of the many books on this topic such as [12] or the standard specifications [26, 28, 29].

3.3 Code Generation

XML-based documents provide much flexibility. However, the actual processing of XML documents requires infrastructure for parsing and validation. Fortunately, there is a plethora of software available to accomplish this. Any programming language in use today provides XML processing capabilities either as part of the language (e. g., *Groovy* [11]) or as a library service. XML processing requires the use of a parser, which can be roughly categorized into stream-oriented (e. g., SAX [21]), tree-traversal or pull-parsing (e. g., DOM [30]), declarative transformation (e. g., XSLT [25, 27]), and data-binding parsing (e. g., for *C++* [32]). The data-binding parser category is of interest here because it allows to automatically create API bindings based on XSDs.

Similar to XML document processing, XML data binding using XSDs is also available for many programming languages and environments. However, we will concentrate on how to auto-generate *C++* source code from XSDs using [32]. A compiler is used to parse an XSD and produce language-dependent definitions. The structure of XSDs is easily transferable into object-oriented technology where XSD types, built-in or user-defined, are expressed as classes and XSD properties or attributes are expressed as class attributes. These auto-generated classes usually contain class-attribute access functions for getting or setting a specific attribute as well as for serialization of class instances. The XSD-to-*C++* compiler at [32] is also capable of extracting documentation and annotations as well as transforming this information for processing with documentation processing software such as `doxygen` [7]. Listing 11 in appendix C shows test driver code in *C++* for creating the XML documents in listings 7–9 and listing 10, respectively. Listing 12 in appendix C shows a test driver that reads XML documents created by listing 11 or manually constructed XML documents adhering to the XSDs in listings 3–6. It is worth noting that both test drivers are relatively small, support standard *C++* idioms for object creation and manipulation, and show that such auto-generated source code can be integrated into existing software with relatively little effort.

XML data binding also supports generation of parser mappings. The auto-generated source code using [32] is usually generated by setting the XSD-to-*C++* compiler into a tree mode. However, it also possible to run the XSD-to-*C++* compiler in a parser mode which allows for creating interfaces that connect to stream-oriented XML parsing infrastructure. The main difference here is that instead of using class abstractions to handle complex data types the stream-oriented interfaces allow for reflection on the data types usually by providing callback mechanisms to client code. This would allow for inspection of unknown entities and potentially

retrieving their definitions from URIs embedded in the XSDs itself. It could also be used to *interpret* the structure of an XSD to derive representational aspects such as building GUI elements on the fly.

4 Discussion

XSDs allow for describing complex data-structure types. These structure types are built from a rich set of built-in primitives as well as by introducing user-defined types by extending existing types. This is very similar to building abstract data types in the object-oriented paradigm. However, XSDs are independent from specific object-oriented programming languages, development environments, or execution platforms. Mature tools for many programming languages and environments exist to automatically generate source code for processing, serialization, etc. as well as documentation for a chosen target platform. This allows for using XML documents as instances of XSDs to exchange data in a truly platform-independent way. The relationship between an XSD and XML instances is similar to the class-and-its-objects relationship in the object-oriented paradigm with the restriction that XSDs do not support methods or messages between objects. However, this restriction is of little consequence because XSDs allow for encoding behavior using the command software-design pattern [9].

We acknowledge that our example in section 3, including the source code in appendices A–C, is rather simplistic. However, the example already captures the essence of computational steering at a high level, at least for particle-based simulations, and was developed in a very short amount of time. XSDs are already successfully used for defining structure in complex problem domains such as computer graphics (X3D [16, 17]) or geographic-information processing (GML [15]). In addition to automatic source-code generation XSDs also allow for interoperability between different sets of XSDs as well as run-time reflection of XML instances. Using the namespace feature of XSD it is possible to encapsulate and disambiguate which types are intended for use in a schema file. The namespace `xmlns:xs="http://www.w3.org/2001/XMLSchema"` introduces a short-hand prefix for the XSD namespace itself; in our example this is set to `xs` but could be any other unique string. A custom namespace can be built by providing a `targetNamespace` attribute. This allows reusing of already existing XSDs without the need for reinvention. Serialized XML instances from an XSD also provide information about their original definitions. By using the `schemaLocation` attribute, run-time parsing of some unknown element can be resolved by fetching the (missing) schema definition to be used for interpreting the XSD instance of the formerly unknown element.

We believe that our approach has the potential for resolving problems inherent to traditional approaches to computational steering, introducing greater flexibility for both the simulation provider as well as the visualization provider or improving the maintainability of infrastructure. We are aware that a large amount of computational-steering software is in existence and we do not propose to replace it by starting from scratch. However, there already seems to be awareness in the community that a more general and flexible approach to controlling (numerical) simulations in general is desirable. For example Biddiscombe et al. [2, 3] describe an approach where the virtual-file driver interface in HDF5 [13] was used to exchange information between a simulation and its desired visual representation. While rather application-specific in its solution Biddiscombe et al. [3] also propose a generic software API, embedded into the HDF5 customization, that would allow to transfer steering-control information from the visual interface to the simulation. Using our approach a simulation provider could simply add an XSD of the control elements and the structure of the simulation's result to existing software. Advertising the simulation-specific XSD would allow the development of independent visualization software

even in the event of not being able to recompile the simulation software or if no appropriate execution platform is available. The same is true for the visualization side. Experiments in finding new, useful, or problem-domain specific solutions could be carried out by concentrating on the data descriptions (and their meaning) while at the same time pseudo-simulators could generate randomized dummy data adhering to the simulation constraints to evaluate such visualization experiments under a variety of conditions. In addition, because XSD instances are communicated by serialized XML documents a large variety of simulation-visualization connection types are supported (e. g., within the same process and between processes on the same or separate machines).

Automatic source-code generation from XSDs provides a tight coupling between a simulation type and possible visualization and control interfaces. The alternative approach of XML data binding using parser mapping has the potential to automate this. Because the parser reports elements back to client code (via callbacks) not only unknown elements can be resolved (by fetching and evaluating their XSDs) but also automated transformations are possible based on the element types found. This is probably easiest to see in the case of automatically creating a GUI layout from some XSD. All built-in XSD types can be directly transferred to standard GUI elements for display or manipulation. Restricted elements such as enumerations can be easily expressed as selection elements in a GUI (e. g., pull-down list, list boxes). Compound structures may trigger grouping mechanisms usually also allowing collapsing or inflating their content. Assuming that some simulation-specific types are known a priori relevant visualization elements can be selected. The result structure in our example consists of a continuous value (`timestamp`) and a list of particles. This information could be used to provide a two-element layout containing a continuous slider for selecting a time step and a widget capable of processing and visualizing 3d data in a (cartesian) coordinate system. Similar approaches have been already explored in the gViz library [5] and the eViz framework [20] and may serve as a valuable starting point for experimentation as well as baselines for comparison. Intuitively, it seems that the more elements relevant to computational steering are known (and/or agreed upon) the more general such on-the-fly visualization-interface building might become. However, it would be naïve to suggest we can provide a full solution to this problem without deeper insights into the problem domain and its requirements. The only definite advantage our approach provides here is the potential for flexible run-time inspection of complex data types in contrast to traditional software APIs.

5 Conclusions and Future Work

We have presented a data-centric approach based on extensible self-describing protocols for communication between simulation and visualization providers in computational steering. Using XML Schema definitions we developed a protocol for the control and result elements of a simple generic CFD simulation. Such XSDs allow for automatic source-code generation in many programming languages and environments as well as the extraction of documentation. This in turn provides simulation providers with the ability to communicate how a specific simulation can be controlled as well as how its results should be interpreted. Visualization providers also benefit by being able to choose the most appropriate visualization and interaction techniques. XSD also enables decoupling of support infrastructure because only the XSD needs to be transformed into code for use in software while the XSD itself is completely independent from specific software APIs, technologies, or execution environments.

Our approach has the potential to describe and use computational steering in a much more flexible way than hitherto. This flexibility reaches into topics as varied as long-term maintain-

ability, connectivity, execution-platform improvements, or innovative approaches to simulation and visualization itself.

Our simple example is exactly that: simplistic. While we believe that the high-level separation into a control and a result structure is valid for a large number of (numerical) simulations, we are unable to provide a formalization here. In part this stems from the fact that computational steering does not have a formalized definition or standards document. We are aware of the efforts of the working group for a simple API for grid applications (SAGA) to define a core API [10] for general grid computing. However, while valuable in providing problem-domain analysis and terminology, SAGA only advocates a software-API approach. We believe that our approach is more suitable to standardizing communication in computational steering and suggest that the development of such a consensus includes discussion and experimentation within the whole community.

The increased use of GPU architectures provides a challenge for our approach. GPU-based simulations do have a great appeal in that they provide faster evaluation per time step below certain problem sizes as well as that the simulation results are already available for further graphics processing on the GPU. Essentially, the simulation provider and visualization component are the same and the use of a generic data-centric communication approach will require further in-depth contemplation.

Manual deduction of appropriate visualization techniques from numerical simulations requires a high degree of expert knowledge in both the specific problem domain of a simulation as well as in the domain of visualization techniques. The ideal solution would be to be able to automatically deduce the inherent structure of a simulation, for both the set of control parameters and the result(s), and to select appropriate visualization techniques. Our data-centric protocol already provides a general approach to the structure-deduction problem for simulation control and its results and we are interested in further investigating this problem domain with respect to automatic deduction of appropriate visualizations techniques.

We believe that our data-centric approach using a structured and self-describing format for communication in computational steering provides a good starting point for further development that enables true platform independence as well as future-proof reusability.

A Example XML Schemas

In this section we provide the XSD sources for n-component floating-point types ($n \in \{1, 2, 3, 4\}$) using simple inheritance (listing 3), a generic particle type providing position and velocity in 3d as well as a unique id and a type for a list of such particles (listing 4), base types for control and result structures suitable in the context of computational steering (listing 5), and specializations of the control and result types for a generic CFD simulation (listing 6), which extend the previously defined base types for control and result structures.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:complexType name="float1">
    <xs:sequence>
      <xs:element name="x" type="xs:float"     />
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="float2">
    <xs:complexContent>
```

```
      <xs:extension base="float1">
        <xs:sequence>
          <xs:element name="y" type="xs:float" />
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>

  <xs:complexType name="float3">
    <xs:complexContent>
      <xs:extension base="float2">
        <xs:sequence>
          <xs:element name="z" type="xs:float" />
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>

  <xs:complexType name="float4">
    <xs:complexContent>
      <xs:extension base="float3">
        <xs:sequence>
          <xs:element name="w" type="xs:float" />
        </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>

</xs:schema>
```

Listing 3: An XML schema definition for one-, two-, three-, and four-component floating-point types, which is stored in `floats.xsd`.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:include schemaLocation="floats.xsd" />

  <xs:complexType name="particle_t">
    <xs:sequence>
      <xs:element name="position" type="float3"         />
      <xs:element name="velocity" type="float3"         />
    </xs:sequence>
    <xs:attribute name="id"       type="xs:unsignedInt"
                                  use="required"        />
  </xs:complexType>

  <xs:element name="particle" type="particle_t" />

  <xs:complexType name="particle_list_t">
    <xs:sequence>
      <xs:element name="particle" type="particle_t"
                                  minOccurs="0"
                                  maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>

  <xs:element name="particle_list" type="particle_list_t" />

</xs:schema>
```

Listing 4: An XML schema definition for a generic particle and a list of particles, which is stored in `particle.xsd`.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
  <xs:complexType name="control_t">
    <xs:sequence>
      <xs:element name="type">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="start"  />
            <xs:enumeration value="stop"   />
            <xs:enumeration value="update" />
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>

  <xs:element name="control" type="control_t"/>

  <xs:complexType name="result_t">
    <xs:sequence>
      <xs:element name="time" type="xs:time" />
    </xs:sequence>
  </xs:complexType>

  <xs:element name="result" type="result_t" />

</xs:schema>
```

Listing 5: An XML schema definition for steering control and result base types, which is stored in `steering.xsd`. The base control type always contains an enumeration for starting, stopping, or updating the simulation. The base result type always includes a time stamp with respect to the simulation time.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">

  <xs:include schemaLocation="particle.xsd" />
  <xs:include schemaLocation="steering.xsd" />

  <xs:complexType name="cfd_control_t">
    <xs:complexContent>
      <xs:extension base="control_t">
         <xs:sequence>
           <xs:element name="gravity" type="float3" />
           <xs:element name="force"   type="float3" />
         </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>

  <xs:element name="cfd_control" type="cfd_control_t"/>

  <xs:complexType name="cfd_result_t">
    <xs:complexContent>
      <xs:extension base="result_t">
         <xs:sequence>
           <xs:element name="particle_list"
                       type="particle_list_t" />
         </xs:sequence>
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>

  <xs:element name="cfd_result" type="cfd_result_t"/>

</xs:schema>
```

Listing 6: An XML schema definition for steering a generic CFD, which is stored in `cfd.xsd`.

B Example XML Documents

In this section we provide the XML sources for XSD instances of control structures of type start (listing 7), update (listing 8), and stop (listing 9), and a result structure (listing 10).

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<cfd_control xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="cfd.xsd">
  <type>start</type>
  <gravity>
    <x>0</x>
    <y>0</y>
    <z>-9.81</z>
  </gravity>
  <force>
    <x>0</x>
    <y>0</y>
    <z>0</z>
  </force>
</cfd_control>
```

Listing 7: An XML document representing a control structure of type start as defined by the XSD in listing 6 and produced by the program in listing 11.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<cfd_control xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="cfd.xsd">
  <type>update</type>
  <gravity>
    <x>0</x>
    <y>0</y>
    <z>-9.81</z>
  </gravity>
  <force>
    <x>0</x>
    <y>1</y>
    <z>0</z>
  </force>
</cfd_control>
```

Listing 8: An XML document representing a control structure of type update as defined by the XSD in listing 6 and produced by the program in listing 11.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<cfd_control xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="cfd.xsd">
  <type>stop</type>
  <gravity>
    <x>0</x>
    <y>0</y>
    <z>0</z>
  </gravity>
  <force>
    <x>0</x>
    <y>0</y>
    <z>0</z>
  </force>
</cfd_control>
```

Listing 9: An XML document representing a control structure of type stop as defined by the XSD in listing 6 and produced by the program in listing 11. Note that the gravity and force attributes may or may not be used in the shutdown process but the definition of the control structure in listing 6 requires their appearance in the XML document.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<cfd_result xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:noNamespaceSchemaLocation="cfd.xsd">
  <time>00:00:00.0005</time>
  <particle_list>
    <particle id="0">
      <position>
        <x>0</x>
        <y>0</y>
        <z>0</z>
      </position>
      <velocity>
        <x>0</x>
        <y>0</y>
        <z>0</z>
      </velocity>
    </particle>

        ⋮

    <particle id="9">
      <position>
        <x>0</x>
        <y>0</y>
        <z>0</z>
      </position>
      <velocity>
        <x>0</x>
        <y>0</y>
        <z>0</z>
      </velocity>
    </particle>
  </particle_list>
</cfd_result>
```

Listing 10: An XML document representing a result structure as defined by the XSD in listing 6 and produced by the program in listing 11.

C Example Test Drivers

In this section we provide the *C++* source code of simple test drivers (listings 11 and 12) for creating the XSD instances serialized to XML documents in listings 7–10 and processing XML documents based on the XSD of generic CFD simulation in listing 6.

```
#include <cstdlib>
#include <fstream>
#include <iostream>

#include <cfd.hpp>

int main(int, char*[])
{
  int result(EXIT_SUCCESS);

  try {
    // required for de-serialization
    xml_schema::namespace_infomap map;

    map[""].name   = "";
    map[""].schema = "cfd.xsd";

    float3 const g (0,0,-9.81);
    float3 const f0(0,0, 0);
    float3 const fy(0,1, 0);

    {
```

```
      cfd_control_t const ctrl(type::start, g, f0);
      std::ofstream       ofs ("../cfd_control_start.xml");

      cfd_control(ofs, ctrl, map);
    }

    {
      cfd_control_t const ctrl(type::update, g, fy);
      std::ofstream       ofs ("../cfd_control_update.xml");

      cfd_control(ofs, ctrl, map);
    }

    {
      cfd_control_t const ctrl(type::stop, f0, f0);
      std::ofstream       ofs ("../cfd_control_stop.xml");

      cfd_control(ofs, ctrl, map);
    }

    {
      cfd_result_t::particle_list_type particles;

      for (unsigned i(0); i < 10; ++i) {
        particles.particle().push_back(particle_t(f0, f0, i));
      }

      using time_type = cfd_result_t::time_type;

      cfd_result_t const result(time_type(0,0,0.0005), particles);
      std::ofstream      ofs ("../cfd_result.xml");

      cfd_result(ofs, result, map);
    }
  }

  catch (xml_schema::exception const& e) {
    std::cerr << e << std::endl;

    result = EXIT_FAILURE;
  }

  return result;
}
```

Listing 11: Test driver using auto-generated source code, provided by the include statement for `cfd.hpp`, to produce the XML documents in listings 7–10.

```
#include <cstdlib>
#include <iostream>

#include <cfd.hpp>

int main(int argc, char* argv[])
{
  int result(EXIT_SUCCESS);

  try {
    if (2 <= argc) {
      bool processed(false);

      try {
        if (!processed) {
          cfd_control_t const ctrl(*(cfd_control(argv[1]).get()));

          std::cout << argv[0] << ": " << ctrl << std::endl;

          processed = true;
        }
```

```
      }

      catch (xml_schema::unexpected_element const& e) {
        std::cerr << e << std::endl;
      }

      try {
        if (!processed) {
          cfd_result_t const result(*(cfd_result(argv[1]).get()));

          std::cout << argv[0] << ": " << result << std::endl;

          processed = true;
        }
      }

      catch (xml_schema::unexpected_element const& e) {
        std::cerr << e << std::endl;

        result = EXIT_FAILURE;
      }

      if (!processed) {
        std::cout << argv[0] << ": " << "unable to process '" << argv[1] << "'" << std::endl;
      }
    } else {
      std::cout << argv[0] << ": " << "no file name supplied" << std::endl;
    }
  }

  catch (xml_schema::exception const& e) {
    std::cerr << e << std::endl;

    result = EXIT_FAILURE;
  }

  return result;
}
```

Listing 12: Test driver using auto-generated source code, provided by the include statement for `cfd.hpp`, for reading and displaying the content of XML documents containing control or result structures as defined in listing 6.

References

[1] G. Abram and L. Treinish. An Extended Data-flow Architecture for Data Analysis and Visualization. In *Proceedings of ACM SIGGRAPH 95*, Computer Graphics Proceedings, Annual Conference Series, pages 17–21. ACM, 1995. DOI: 10.1145/204362.204366.

[2] J. Biddiscombe, J. Soumagne, G. Oger, D. Guibert, and J.-G. Piccinali. Parallel Computational Steering and Analysis for HPC Applications using a ParaView Interface and the HDF5 DSM Virtual File Driver. In *Eurographics Symposium on Parallel Graphics and Visualization*, pages 91–100. Eurographics, 2011. DOI: 10.2312/EGPGV/EGPGV11/091-100.

[3] J. Biddiscombe, J. Soumagne, G. Oger, D. Guibert, and J.-G. Piccinali. Parallel Computational Steering for HPC Applications using HDF5 Files in Distributed Shared Memory. *IEEE Trans. Vis. Comput. Graphics*, 18(6):852–864, 2012. DOI: 10.1109/TVCG.2012.63.

[4] K. Brodlie, A. Poon, H. Wright, L. Brankin, G. Banecki, and A. Gay. GRASPARC – A Problem Solving Environment Integrating Computation and Visualization. In *Proceedings IEEE Visualization*, pages 102–109. IEEE, 1993. DOI: 10.1109/VISUAL.1993.398857.

[5] K. Brodlie, D. Duce, J. Gallop, Sagar M., J. Walton, and J. Wood. Visualization in Grid Computing Environments. In *Proceedings IEEE Visualization*, pages 155–162. IEEE, 2004. DOI: 10.1109/VISUAL.2004.112.

[6] F. Chatzinikos and H. Wright. Enabling Multipurpose Image Interaction in Modular Visualization Environments. In *Visualization and Data Analysis 2003*, pages 203–214. SPIE, 2003. DOI: 10.1117/12.474026.

[7] Doxygen – Generate documentation from source code. URL http://www.stack.nl/~dimitri/doxygen/.

[8] J. M. Favre and M. Valle. AVS and AVS/Express. In C. Johnson and C. Hansen, editors, *Visualization Handbook*, pages 655–672. Academic Press, 2004.

[9] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1995.

[10] T. Goodale, S. Jha, H. Kaiser, T. Kielmann, P. Kleijer, A. Merzky, J. Shalf, and C. Smith. *A Simple API for Grid Applications (SAGA)*. Open Grid Forum, 2013. URL http://www.ogf.org/documents/GFD.90.pdf.

[11] Groovy Programming Language. URL http://groovy-lang.org/.

[12] E. R. Harold. *XML 1.1 Bible*. John Wiley & Sons, 3rd edition, 2004.

[13] HDF Group. URL http://www.hdfgroup.org/.

[14] IEEE:1278.1:2012. *Standard for Distributed Interactive Simulation: Application Protocols*. IEEE, 2012. IEEE Standard 1278.1-2012.

[15] ISO:19136:2007. *Geographic Information – Geography Markup Language (GML)*. ISO/IEC, 2013. International Standard ISO/IEC 19136.

[16] ISO:19775-1:2013. *Information Technology – Computer Graphics, Image Processing and environmental data representation – Extensible 3D (X3D) – Part 1: Architecture and Base Components*. ISO/IEC, 2013. International Standard ISO/IEC 19775-1.

[17] ISO:19776-1:2015. *Information Technology – Computer Graphics, Image Processing and Environmental Data Representation – Extensible 3D (X3D) Encodings – Part 1: Extensible Markup Language (XML) Encoding*. ISO/IEC, 2015. International Standard ISO/IEC 19776-1.

[18] S. G. Parker, D. M. Weinstein, and C. R. Johnson. The SCIRun Computational Steering Software System. In *Modern Software Tools for Scientific Computing*, pages 5–44. Springer Verlag, 1997. DOI: 10.1007/978-1-4612-1986-6˙1.

[19] S. M. Pickles, R. Haines, R. L. Pinning, and A. R. Porter. A Practical Toolkit for Computational Steering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1833):1843–1853, 2005. DOI: 10.1098/rsta.2005.1611.

[20] M. Riding, J. Wood, K. W. Brodlie, J. M. Brooke, M. Chen, D. Chisnall, C. Hughes, N. W. John, D. M. Jones, and N. Roard. eViz: Towards an Integrated Framework for High Performance Visualization. In *Proceedings of UK e-Science All Hands Meeting*, pages 344–351, 2005. http://www.allhands.org.uk/2005/proceedings/papers/344.pdf.

[21] SAX Project. URL http://www.saxproject.org/.

[22] SCIRun. URL http://www.scirun.org/.

[23] R. van Liere, J. D. Mulder, and J. J. van Wijk. Computational Steering. *Future Generation Computer Systems*, 12(5):441–450, 1997. DOI: 10.1016/S0167-739X(96)00029-5.

[24] J. P. R. B. Walton. NAG's IRIS Explorer. In C. Johnson and C. Hansen, editors, *Visualization Handbook*, pages 633–654. Academic Press, 2004.

[25] W3C:XSLT1.0:1999. *XSL Transformations (XSLT) Version 1.0 – W3C Recommendation 16 November 1999*. World Wide Web Consortium (W3C), 1999. URL http://www.w3.org/TR/1999/REC-xslt-19991116.

[26] W3C:XML1.1:2006. *Extensible Markup Language (XML) 1.1 (Second Edition) – W3C Recommendation 16 August 2006*. World Wide Web Consortium (W3C), 2006. URL http://www.w3.org/TR/2006/REC-xml11-20060816/.

[27] W3C:XSLT2.0:2007. *XSL Transformations (XSLT) Version 2.0 – W3C Recommendation 23 January 2007*. World Wide Web Consortium (W3C), 2007. URL http://www.w3.org/TR/2007/REC-xslt20-20070123/.

[28] W3C:XSD1.1P1:2012. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures – W3C Recommendation 5 April 2012*. World Wide Web Consortium (W3C), 2012. URL http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/.

[29] W3C:XSD1.1P2:2012. *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes – W3C Recommendation 5 April 2012*. World Wide Web Consortium (W3C), 2012. URL http://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/.

[30] W3C:DOM4:2015. *W3C DOM4 – W3C Recommendation 19 November 2015*. World Wide Web Consortium (W3C), 2015. URL http://www.w3.org/TR/2015/REC-dom-20151119/.

[31] H. Wright, R. H. Crompton, S. Kharche, and P. Wenisch. Steering and Visualization: Enabling Technologies for Computational Science. *Future Generation Computer Systems*, 26(3):506–513, 2010. DOI: 10.1016/j.future.2008.06.015.

[32] XSD - XML Schema to C++ Data Binding Compiler - Project Page. URL http://www.codesynthesis.com/projects/xsd/.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DIRECT TIME-DOMAIN INTEGRATION APPROACH FOR VISCOELASTIC SYSTEMS INVOLVING VARIOUS DAMPING MODELS

Zhe Ding, Li Li, and Yujin Hu

State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

e-mail: dingzhe@hust.edu.cn, lili_em@hust.edu.cn, yjhu@hust.edu.cn

Keywords: Nonviscous damping, various damping models, Time-domain, State-space method, Dynamic response.

Abstract. *Mechanical engineering systems with two or more parts of significantly different energy dissipation levels are encountered frequently in dynamical designs nowadays. Although dynamic analyses of only one exponential-like damping model in independent structures are studied by authors for decades, their counterpart for nonviscous systems involving various damping models seems not well-developed. In this paper, a time-domain analysis of various damping models subjected to arbitrary initial conditions is presented. The proposed approach is built on an extended state-space representation of the equations of motion. The method can be applied in both conditions when the damping matrices are of full rank or rank deficient in a uniform expression. The dynamic responses of the various damping models are calculated only by the original system matrices without solving the eigensolutions of the systems. This advantage particularly makes this method numerically efficient. Finally, a numerical example is offered to illustrate the accuracy of the derived method.*

1 INTRODUCTION

Damping is longtime regarded as a significant factor in the dynamic behavior of vibration systems and is a characteristic parameter describing the energy dissipation. The viscous damping model which was first proposed by Lord Rayleigh [1] has been extensively used in the past centuries for its theoretical simplicity. However, with the development of many industrial applications (such as composite structural materials, active control in spacecraft, ships, rockets and railway tracks, sandwich structures), the classical viscous damping model may not properly reveal the dissipative forces. As a result, there has been an increasing demand in recent years on nonviscous damping models with the purpose of describing dissipative forces in a more general and accurate manner compared to the limited scope offered by the viscous damping model. Majority of the nonviscous damping models, like Biot model [2, 3], Golla-Hugher-McTavish (GHM) model [4, 5], Anelastic Displacement Field (ADF) model [6, 7], are presented considering the entire velocity history via convolution integrals over some kernel functions.

Classic dynamic analysis methods for linear viscous damping cannot be directly applied in the nonviscous damping model systems. Any modifications of nonviscous model may lead to some mathematical difficulties and changes for solving the corresponding dynamic equations. In recent years, researchers have considered the dynamic responses of nonviscous damping models. Some authors developed exact state-space approaches [4-12] based on internal variables. Others proposed approximated approaches [13-20] and model reduction methods [21-26] to solve the eigenproblem for the exponential-like damping models. The previously mentioned methods are devoted to efficiently solve the nonviscous eigenproblem under different damping models. However, the solution of this eigenproblem is time-consuming. Then, a direct time-domain integration approach for exponentially damped systems was proposed by Adhikari and Wager [27]. The method used only original system matrices to compute the dynamic response which improved the efficiency significantly. Later, Cortés et al. [28] proposed a direct integration formulation for exponentially damped systems, which did not employ internal variables. Liu [29] proposed analysis method for non-viscously damped structures based on Newmark's assumption of acceleration. Pan and Wang [30] developed a frequency-domain method for exponentially damped systems using the Discrete Fourier Transform method in combination with Fast Fourier Transform which can deal with non-zero initial conditions.

The mentioned previous studies are restricted to the case of the systems with only one damping model. Nowadays, in practice, structure systems with two or more parts with remarkably different levels of energy dissipation are encountered frequently in mechanical engineering. Therefore, these damping systems often involve multiple damping models.

Although the exponential damping model, Biot model, GHM model and ADF model are physically different, they can be mathematically represented by a friction formula of rational polynomials [31]. The uniform way to express damping models can simplify the theoretical analysis of dynamic responses and be applied conveniently in multiple damping models. Recently, Li and Hu [32] have proposed a state-space approach for the analysis of linear systems with multiple damping models. The dynamic response of the system was calculated by mode-superposition method using the state-space eigensolutions which was time-consuming.

Based on the problems mentioned above, a time-domain approach for various damping models is proposed. The presented time-domain approach can be applied in both conditions when the damping matrices are of full rank or rank deficient. The advantage of the proposed method over the traditional mode superposition method is that the final recurrence formulas to be solved contain nothing but a linear combination of the system matrices. It will be demonstrated by a numerical example that the proposed method shares the same accuracy with exact values but is of greater efficiency.

The paper is organized as follows: in Section 2, we will quickly review some preliminary concepts and definitions of various damping models and its corresponding state-space formalism. A direct time-domain approach for various damping models is proposed to calculate the dynamic response of the systems in Section 3. We briefly summarize the method in Section 4 to make it convenient to be coded up. The results of a numerical example are presented in Section 5 in order to assess the performance of the proposed method. The paper ends with some conclusions in section 6.

2 REVIEW OF THE STATE-SPACE FORMALISM

The equations of motion of an *N*-degree-of-freedom linear system with various damping models can be expressed by

$$\mathbf{M}\ddot{\mathbf{u}}(t)+\mathbf{C_0}\dot{\mathbf{u}}(t)+\sum_{k=1}^{n}\mathbf{C}_k\int_{-\infty}^{t}g_k(t-\tau)\frac{\partial\mathbf{u}(\tau)}{\partial\tau}d\tau+\mathbf{K}\mathbf{u}(t)=\mathbf{f}(t) \tag{1}$$

together with the initial conditions

$$\mathbf{u}(t=0)=\mathbf{u_0}\in\mathbb{R}^N \quad \text{and} \quad \dot{\mathbf{u}}(t=0)=\dot{\mathbf{u}}_\mathbf{0}\in\mathbb{R}^N \tag{2}$$

Here $\mathbf{u}(t)\in\mathbb{R}^{N\times N}$ is the displacement vector, $\mathbf{M}\in\mathbb{R}^{N\times N}$ is the mass matrix, $\mathbf{K}\in\mathbb{R}^{N\times N}$ is the stiffness matrix, $\mathbf{C_0}\in\mathbb{R}^{N\times N}$ is the frequency-independent damping matrix, $\mathbf{C}_k\in\mathbb{R}^{N\times N}$ are the coefficient matrices of frequency-dependent damping, $g_k(t)$ is the *k*th damping kernel function, $\mathbf{f}(t)$ is the forcing vector.

In fact, any causal model which can make the energy dissipation function nonnegative, may be considered as a candidate for a damping model. So far, several physically realistic mathematical forms of damping models are available in the literature. The Biot model [2, 3] can be given by

$$g_k(t)=a_0\delta(t)+\sum_{k=1}^{m}a_k\exp(-b_kt) \tag{3}$$

The GHM model [4, 5] and ADF model [6, 7]can be respectively expressed as

$$g_k(t)=\sum_{k=1}^{m}\alpha_k\frac{\hat{b}_{2k}e^{-\hat{b}_{1k}t}-\hat{b}_{1k}e^{-\hat{b}_{2k}t}}{\hat{b}_{2k}-\hat{b}_{1k}} \text{ and } g_k(t)=\sum_{k=1}^{m}\Delta_k\exp(-\Omega_kt) \tag{4}$$

where

$$\hat{b}_{1k} = \hat{\omega}_k \hat{\xi}_k - \hat{\omega}_k \sqrt{\hat{\xi}_k^2 - 1}, \hat{b}_{2k} = \hat{\omega}_k \hat{\xi}_k + \hat{\omega}_k \sqrt{\hat{\xi}_k^2 - 1}$$

The uniform way of expressing damping models including the exponential damping model, Biot model, GHM model and ADF model can be represented by a fraction formula of rational polynomials [32]

$$G_k(s) = \frac{c_{p_k} s^{p_k} + c_{p_{k-1}} s^{p_{k-1}} + \cdots + c_0}{d_{q_k} s^{q_k} + d_{q_{k-1}} s^{q_{k-1}} + \cdots + d_0} \tag{5}$$

The previous equation can be rewritten as the following form based on the theory of minimal realization

$$G_k(s) = \frac{c_{p_k} s^{p_k} + c_{p_{k-1}} s^{p_{k-1}} + \cdots + c_0}{d_{q_k} s^{q_k} + d_{q_{k-1}} s^{q_{k-1}} + \cdots + d_0} = \mathbf{P}_k^T (\mathbf{E}_k - s\mathbf{W}_k)^{-1} \mathbf{Q}_k \tag{6}$$

where $\mathbf{E}_k, \mathbf{W}_k \in \mathbb{R}^{q_k \times q_k}$ and vectors $\mathbf{P}_k, \mathbf{Q}_k \in \mathbb{R}^{q_k \times q_k}$.

It should be noted that the coefficient of the highest order denominator can be normalized to 1, which has

$$d_{q_k} = 1 \tag{7}$$

The other matrices and vectors can be expressed as

$$\mathbf{P}_k = \begin{Bmatrix} c_{q_k-1} - d_{q_k-1} c_{q_k} \\ c_{q_k-2} - d_{q_k-2} c_{q_k} \\ \vdots \\ c_0 - d_0 c_{q_k} \end{Bmatrix} \left(\forall j > p_k, c_j = 0\right), \mathbf{E}_k = \begin{bmatrix} -d_{q_k-1} & \cdots & -d_1 & -d_0 \\ 1 & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ 0 & \mathbf{0} & 1 & 0 \end{bmatrix}, \mathbf{W}_k = \mathbf{I}_{q_k}, \mathbf{Q}_k = \begin{Bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$$

In the actual mechanical structure, the damping materials are only a small part of the whole system. As a result, the coefficient matrix of frequency-dependent damping is rank deficient. We assume that in general

$$r_k = rank(\mathbf{C}_k) \le N, \forall k = 1, \cdots n \tag{8}$$

Thus, the coefficient matrix $\mathbf{C}_k$ has the rank-revealing decomposition

$$\mathbf{C}_k = \mathbf{L}_k \mathbf{R}_k^T \tag{9}$$

where $\mathbf{L}_{k,} \mathbf{R}_k \in \mathbb{R}^{N \times r_k}$ are of full column rank r_k .

Eq.(1) can be represented in the first-order form as

$$\mathbf{B}\dot{\mathbf{z}}(t)=\mathbf{A}\mathbf{z}(t)+\mathbf{r}(t) \tag{10}$$

where

$$\mathbf{B}=\begin{bmatrix}\mathbf{C}_0+\mathbf{L}\mathbf{E}^{-1}\mathbf{R}^T & \mathbf{M} & \mathbf{L}\mathbf{E}^{-1}\mathbf{W}\\ \mathbf{M} & 0 & 0\\ \mathbf{E}^{-1}\mathbf{R}^T & 0 & \mathbf{E}^{-1}\mathbf{W}\end{bmatrix},\mathbf{A}=\begin{bmatrix}-\mathbf{K} & 0 & 0\\ 0 & \mathbf{M} & 0\\ 0 & 0 & \mathbf{I}_y\end{bmatrix},\mathbf{z}(t)=\begin{bmatrix}\mathbf{u}(t)\\ \mathbf{v}(t)\\ \mathbf{y}(t)\end{bmatrix},\mathbf{r}(t)=\begin{bmatrix}\mathbf{f}(t)\\ 0\\ 0\end{bmatrix}$$

$$\mathbf{v}(t)=\dot{\mathbf{u}}(t),\ \mathbf{y}(t)=\ell^{-1}\left[s(\mathbf{E}-s\mathbf{W})^{-1}\mathbf{R}^T\right]\mathbf{u}(t)$$

$$\mathbf{E}=diag(\mathbf{I}_{r_1}\otimes\mathbf{E}_1,\mathbf{I}_{r_2}\otimes\mathbf{E}_2,\cdots,\mathbf{I}_{r_n}\otimes\mathbf{E}_n),\mathbf{W}=diag(\mathbf{I}_{r_1}\otimes\mathbf{W}_1,\mathbf{I}_{r_2}\otimes\mathbf{W}_2,\cdots,\mathbf{I}_{r_n}\otimes\mathbf{W}_n)$$

$$\mathbf{L}=\left[\mathbf{L}_1(\mathbf{I}_{r_1}\otimes\mathbf{Q}_1)^T,\mathbf{L}_2(\mathbf{I}_{r_2}\otimes\mathbf{Q}_2)^T,\cdots,\mathbf{L}_n(\mathbf{I}_{r_n}\otimes\mathbf{Q}_n)^T\right]$$

$$\mathbf{R}=\left[\mathbf{R}_1(\mathbf{I}_{r_1}\otimes\mathbf{P}_1)^T,\mathbf{R}_2(\mathbf{I}_{r_2}\otimes\mathbf{P}_2)^T,\cdots,\mathbf{R}_n(\mathbf{I}_{r_n}\otimes\mathbf{P}_n)^T\right]$$

The symbol $\otimes$ in the previous expression denotes the Kronecker product.

3 DIRECT TIME-DOMAIN APPROACH

First, assume the displacement and the velocities vectors at random time *t* can be expressed by the linear approximations

$$\mathbf{u}(t)=\mathbf{u}_j(1-\frac{t}{h})+\mathbf{u}_{j+1}\frac{t}{h} \tag{11}$$

and

$$\mathbf{v}(t)=\mathbf{v}_j(1-\frac{t}{h})+\mathbf{v}_{j+1}\frac{t}{h} \tag{12}$$

The internal variables $\mathbf{y}(t)$ can also be written in a linear manner

$$\mathbf{y}(t)=\mathbf{y}_j(1-\frac{t}{h})+\mathbf{y}_{j+1}\frac{t}{h} \tag{13}$$

The step size $h=T/\mathrm{N}_d$, where T equals the time which the response calculation is required and N_d is the number of divisions in the time axis. In order to simplify the calculation, in this paper, we only consider the constant time-steps. To get the totally discretized one-step formulation which allows $\mathbf{z}_{j+1}$ to be calculated by $\mathbf{z}_j$, the last and the next step considers the integration of the Eq.(10) with respect to the time interval with $t/h\in\left[j,j+1\right]$. The trapezoidal rule of numerical integration is used.

$$\left[\mathbf{B}-\frac{h}{2}\mathbf{A}\right]\mathbf{z}_{j+1}=\left[\mathbf{B}+\frac{h}{2}\mathbf{A}\right]\mathbf{z}_j+\mathbf{i}_r,\mathbf{i}_r=\int_{jh}^{(j+1)h}\mathbf{r}(t)\,\mathrm{d}t \tag{14}$$

where

$$\mathbf{z}_0^T=\left[\mathbf{u}_0^T,\mathbf{v}_0^T,\mathbf{0}^T,\cdots,\mathbf{0}^T\right] \tag{15}$$

Eq.(14) can be used to compute $\mathbf{z}(t)$ at discrete time.

Combining Eq.(10) and the second row of Eq.(14), we can get the following expression

$$\mathbf{v}_{j+1}=\frac{2}{h}[\mathbf{u}_{j+1}-\mathbf{u}_j]-\mathbf{v}_j \tag{16}$$

The third row in Eq.(14) is used to express the internal variables inters of the displacements as

$$\mathbf{y}_{j+1}=(\mathbf{E}^{-1}\mathbf{W}-\frac{h}{2}\mathbf{I}_y)^{-1}[-\mathbf{E}^{-1}\mathbf{R}^T(\mathbf{u}_{j+1}-\mathbf{u}_j)+(\mathbf{E}^{-1}\mathbf{W}+\frac{h}{2}\mathbf{I}_y)\mathbf{y}_j] \tag{17}$$

If we denote $\mathbf{a}_1=(\mathbf{E}^{-1}\mathbf{W}-\frac{h}{2}\mathbf{I}_y)^{-1},\mathbf{a}_2=(\mathbf{E}^{-1}\mathbf{W}+\frac{h}{2}\mathbf{I}_y)$, Eq.(17) can be expressed as

$$\mathbf{y}_{j+1}=\mathbf{a}_1[-\mathbf{E}^{-1}\mathbf{R}^T(\mathbf{u}_{j+1}-\mathbf{u}_j)+\mathbf{a}_2\mathbf{y}_j] \tag{18}$$

Substituting Eq.(16) into the first row of Eq.(14), we can receive a totally discretized one-step formulation of the displacements $\mathbf{u}_{j+1}$

$$\mathbf{a}_3\mathbf{u}_{j+1}=(\mathbf{a}_3-h\mathbf{K})\mathbf{u}_j+2\mathbf{M}\mathbf{v}_j+\mathbf{a}_4\mathbf{y}_j-\mathbf{a}_4\mathbf{y}_{j+1} \tag{19}$$

where

$$\mathbf{a}_3=(\mathbf{C}_0+\mathbf{L}\mathbf{E}^{-1}\mathbf{R}^T+\frac{2}{h}\mathbf{M}+\frac{h}{2}\mathbf{K}) \tag{20}$$

$$\mathbf{a}_4=\mathbf{L}\mathbf{E}^{-1}\mathbf{W} \tag{21}$$

Substituting Eq.(18) into Eq.(19), we have

$$\begin{aligned}\left[\mathbf{a}_3-\mathbf{a}_4\mathbf{a}_1\mathbf{E}^{-1}\mathbf{R}^T\right]\mathbf{u}_{j+1}=&\left[\mathbf{a}_3-\mathbf{a}_4\mathbf{a}_1\mathbf{E}^{-1}\mathbf{R}^T-h\mathbf{K}\right]\mathbf{u}_j+2\mathbf{M}\mathbf{v}_j\\&+\left[\mathbf{a}_4-\mathbf{a}_4\mathbf{a}_1\mathbf{a}_2\right]\mathbf{y}_j+\int_{jh}^{(j+1)h}\mathbf{f}(t)\,\mathrm{d}t\end{aligned} \tag{22}$$

Eq.(22) together with Eqs.(16), (18) allow $\mathbf{u}_{j+1}$ to be calculated by means of $\mathbf{u}_j$. The above process can be applied in both cases when $\mathbf{C}_k$ matrices are rank deficient or of full rank

4 SUMMARY OF THE METHOD

Based on the procedure mentioned above, the time-domain response of a various damping models can be obtained followed by the next steps:

(1) Transform each various damping model into the formalism of Eq.(5) and obtain the corresponding computing matrices $\mathbf{P}_k, \mathbf{Q}_k, \mathbf{E}_k, \mathbf{W}_k, \mathbf{L}_k, \mathbf{R}_k$

(2) Compute matrices $\mathbf{E}, \mathbf{W}, \mathbf{L}, \mathbf{R}$

(3) Select a sufficiently small time step size h and construct the computing matrices

$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ according to Eqs.(18)-(21)

(4) Calculate the matrices as

$$\mathbf{S}_1 = \left[\mathbf{a}_3 - \mathbf{a}_4 \mathbf{a}_1 \mathbf{E}^{-1} \mathbf{R}^T \right] \tag{23}$$

$$\mathbf{S}_2 = \left[\mathbf{a}_3 - \mathbf{a}_4 \mathbf{a}_1 \mathbf{E}^{-1} \mathbf{R}^T - h\mathbf{K} \right] = \mathbf{S}_1 - h\mathbf{K} \tag{24}$$

$$\mathbf{S}_3 = \left[\mathbf{a}_4 - \mathbf{a}_4 \mathbf{a}_1 \mathbf{a}_2 \right] \tag{25}$$

(5) Solve for displacements

$$\mathbf{S}_1 \mathbf{u}_{j+1} = \mathbf{S}_2 \mathbf{u}_j + 2\mathbf{M}\mathbf{v}_j + \mathbf{S}_3 \mathbf{y}_j + \int_{jh}^{(j+1)h} f(t)\,\mathrm{d}t \tag{26}$$

Velocities

$$\mathbf{v}_{j+1} = \frac{2}{h}\left[\mathbf{u}_{j+1} - \mathbf{u}_j \right] - \mathbf{v}_j \tag{27}$$

Auxiliary variable

$$\mathbf{y}_{j+1} = \mathbf{a}_1 \left[-\mathbf{E}^{-1} \mathbf{R}^T (\mathbf{u}_{j+1} - \mathbf{u}_j) + \mathbf{a}_2 \mathbf{y}_j \right] \tag{28}$$

5 NUMERICAL EXAMPLE

A four DOF system with two different damping models is considered as shown in Fig.1.

The equations of motion of this model system can be expressed by Eq.(1). The system matrices are given by

$$\mathbf{M}=\begin{bmatrix} m & & & \\ & m & & \\ & & m & \\ & & & m \end{bmatrix},\mathbf{K}=\begin{bmatrix} 2k_e & -k_e & & \\ -k_e & 2k_e & -k_e & \\ & -k_e & 2k_e & -k_e \\ & & -k_e & 2k_e \end{bmatrix},G_1(s)=a_0+\sum_{k=1}^{2}\frac{a_k}{s+b_k},$$

$$\mathbf{C}_1=\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},G_2(s)=\sum_{j=1}^{2}\frac{\mu_j}{s+\mu_j},\mathbf{C}_2=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Here $m=1\text{kg},k=100\,\text{N/m},a_0=1.497\times10^{-2},a_1=1.0132\times10^{4},a_2=1.2022\times10^{4}$, $b_1=5.5893,b_2=102.59,\mu_1=5$ and $\mu_2=100$. It should be noted that $G_1(s)$ is a Biot damping model and $G_2(s)$ is an exponential damping model both with two series. Obviously, both the damping coefficient matrices are rank deficiency with $r_1=rank(\mathbf{C}_1)=1\le4$ and $r_2=rank(\mathbf{C}_2)=1\le4$. The order of the system matrices expressed by Eq.(10) in the state-space can be obtained as $m=2\times4+(2\times1+2\times1)=12$.

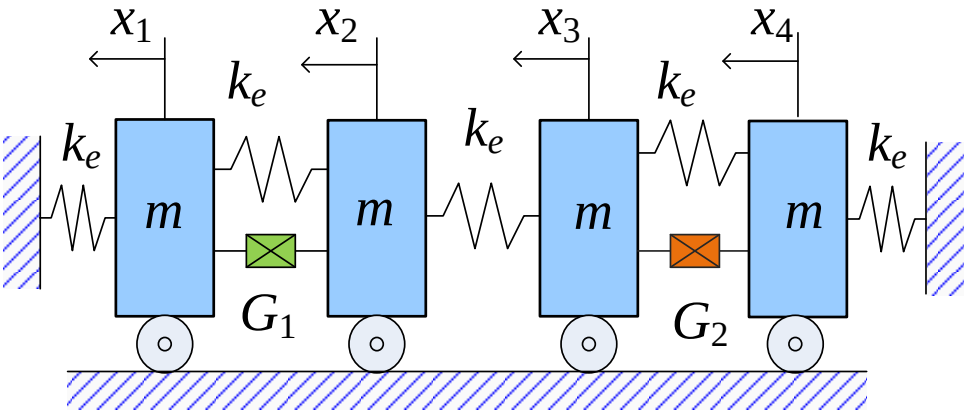


Figure 1: A four DOF system with two different damping models

We consider the dynamic response of the system subjected to a unit impulse excitation at the first DOF. The unit impulse excitation can be equally transformed into an unit initial displacement according to the theorem of impulse, that is, $\mathbf{f}(i\omega)=\mathbf{0},\mathbf{u}_0=\mathbf{0},\mathbf{v}_0=\{1,0,0,0\}^{\mathrm{T}}$. The response of the four masses are obtained using the above proposed direct time-domain approach with time step $h=0.01s$.

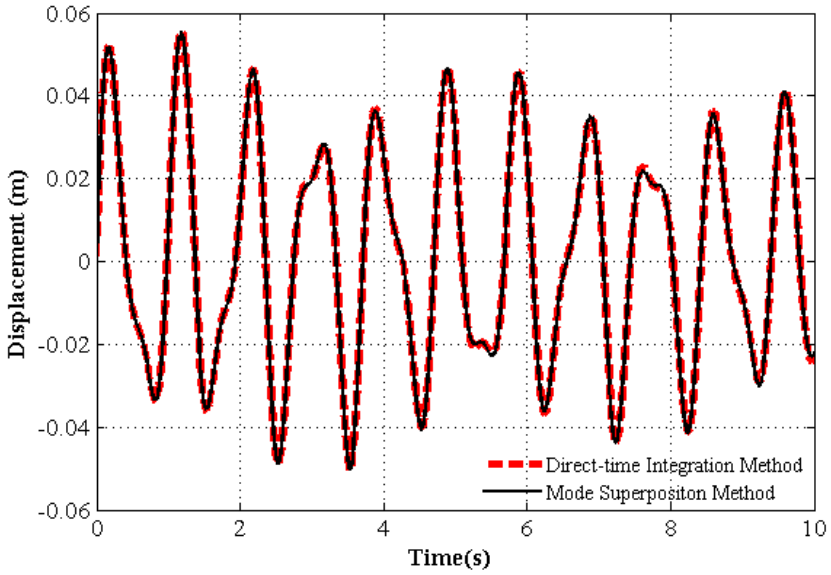


Figure 2： Displacement at the first DOF.

Fig.2 shows the response of the first DOF as a function of time. The results of the system on the same initial condition are obtained using the exact state-space mode superposition method and shown in the same figure. The eigenvalues and eigenvectors of the four DOF system are list in paper [32]. However, for large-scaled problems, it is often unnecessary and impossible to obtain all the eigenpairs of the system, the high-order mode truncation problem is frequently encountered and leads to errors [33-36]. But for the proposed direct-time domain method for various damping models, we need not compute the eigensolutions which will be a significant advantage over the traditional superposition method.

It is shown that the results obtained using both approaches match with excellent accuracy. The time-domain method proposed in this paper has a higher efficiency because the computation of state-space eigensolutions is avoided.

6 CONCLUSIONS

In this paper, a direct time-domain formulation of linear viscoelastic systems with various damping models has been presented. The various damping models described by a fraction formula of rational polynomials are attempting to express different damping models in a uniform way. The proposed method is on the base of an extended state-space formulation of various viscoelastic damping models. The method can be utilized in both conditions when the damping matrices $\mathbf{C}_k$ are of full rank or rank deficient. The advantage of the proposed method over the traditional mode superposition method is that the final recurrence formulas to be solved contain noting but a linear combination of the system matrices $\mathbf{M},\mathbf{C}_k,\mathbf{K}$ and time-stepping scheme *h*. We do not need to compute the eigenvalues and handle the non-viscous modes of the viscoelastic models. Thus, although the problem is originally organized in the extended state-space, the final solution is in effect implemented in the original reduced physical space. Finally, a four DOF system with two different damping models is used to show the validity of the proposed method.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 51375184) and the National Science and Technology Major Project of China (Grant No. 2012ZX04003-021).

REFERENCES

[1] J.W.S.B. Rayleigh, *The theory of sound*, Macmillan, 1896.

[2] M. Biot, Theory of Stress‐Strain Relations in Anisotropic Viscoelasticity and Relaxation Phenomena, *Journal of Applied Physics*, **25**, 1385-1391, 1954.

[3] M. Biot, Variational principles in irreversible thermodynamics with application to viscoelasticity, *Physical Review*, **97**, 1463, 1955.

[4] D.F. Golla, P. Hughes, Dynamics of viscoelastic structures—a time-domain, finite element formulation, *ASME J. Appl. Mech*, **52**, 897-906, 1985.

[5] D. McTavish, P. Hughes, Modeling of linear viscoelastic space structures, *Journal of Vibration and Acoustics*, **115**, 103-110, 1993.

[6] G.A. Lesieutre, D.L. Mingori, Finite element modeling of frequency-dependent material damping using augmenting thermodynamic fields, *Journal of Guidance, Control, and Dynamics*, **13**, 1040-1050, 1990.

[7] G.A. Lesieutre, Finite elements for dynamic modeling of uniaxial rods with frequency-dependent material properties, *International Journal of Solids and Structures*, **29**, 1567-1579, 1992.

[8] J. Zhang, G. Zheng, The Biot model and its application in viscoelastic composite structures, *Journal of Vibration and Acoustics*, **129**, 533-540, 2007.

[9] C. Vasques, R. Moreira, J.D. Rodrigues, Viscoelastic Damping Technologies–Part I: Modeling and Finite Element Implementation⋆, *Mechanical Engineering*, **1**, 96-110, 2010.

[10] A. Palmeri, S. Adhikari, A Galerkin-type state-space approach for transverse vibrations of slender double-beam systems with viscoelastic inner layer, *Journal of Sound and Vibration*, **330**, 6372-6386, 2011.

[11] N. Wagner, S. Adhikari, Symmetric state-space method for a class of nonviscously damped systems, *AIAA journal*, **41**, 951-956, 2003.

[12] S. Adhikari, N. Wagner, Analysis of asymmetric nonviscously damped linear dynamic systems, *Journal of Applied Mechanics*, **70**, 885-893, 2003.

[13] J. Woodhouse, Linear damping models for structural vibration, *Journal of Sound and Vibration*, **215**, 547-569, 1998.

[14] S. Adhikari, Damping models for structural vibration. University of Cambridge, 2001.

[15] F. Cortés, M.J. Elejabarrieta, Computational methods for complex eigenproblems in finite element analysis of structural systems with viscoelastic damping treatments, *Computer methods in applied mechanics and engineering*, **195**, 6448-6462, 2006.

[16] S. Adhikari, B. Pascual, Eigenvalues of linear viscoelastic systems, *Journal of Sound and Vibration*, **325**, 1000-1011, 2009.

[17] S. Adhikari, Dynamics of nonviscously damped linear systems, *Journal of engineering mechanics*, **128**, 328-339, 2002.

[18] Q. Chen, Y. Chan, Integral finite element method for dynamical analysis of elastic–viscoelastic composite structures, *Computers & Structures*, **74**, 51-64, 2000.

[19] S. Adhikari, B. Pascual, Iterative methods for eigenvalues of viscoelastic systems, *Journal of Vibration and Acoustics*, **133**, 021002, 2011.

[20] Y. Lei, M. Friswell, S. Adhikari, A Galerkin method for distributed systems with non-local damping, *International Journal of Solids and Structures*, **43**, 3381-3400, 2006.

[21] S. Zghal, M.L. Bouazizi, N. Bouhaddi, R. Nasri, Model reduction methods for viscoelastic sandwich structures in frequency and time domains, *Finite Elements in Analysis and Design*, **93**, 12-29, 2015.

[22] F. Boumediene, J.-M. Cadou, L. Duigou, A reduction model for eigensolutions of damped viscoelastic sandwich structures, *Mechanics Research Communications*, **57**, 74-81, 2014.

[23] A. De Lima, D. Rade, F.L. Neto, An efficient modeling methodology of structural systems containing viscoelastic dampers based on frequency response function substructuring, *Mechanical Systems and Signal Processing*, **23**, 1272-1281, 2009.

[24] A. De Lima, A. Da Silva, D. Rade, N. Bouhaddi, Component mode synthesis combining robust enriched Ritz approach for viscoelastically damped structures, *Engineering Structures*, **32**, 1479-1488, 2010.

[25] M.A. Trindade, Reduced-order finite element models of viscoelastically damped beams through internal variables projection, *Journal of Vibration and Acoustics*, **128**, 501-508, 2006.

[26] M. Trindade, A. Benjeddou, R. Ohayon, Piezoelectric active vibration control of damped sandwich beams, *Journal of Sound and Vibration*, **246**, 653-677, 2001.

[27] S. Adhikari, N. Wagner, Direct time-domain integration method for exponentially damped linear systems, *Computers & structures*, **82**, 2453-2461, 2004.

[28] F. Cortés, M. Mateos, M.J. Elejabarrieta, A direct integration formulation for exponentially damped structural systems, *Computers & Structures*, **87**, 391-394, 2009.

[29] Q. Liu, Computational Method of the Dynamic Response for Nonviscously Damped Structure Systems, *Journal of Engineering Mechanics*, **140**, 04014085, 2014.

[30] Y. Pan, Y. Wang, Frequency-domain analysis of exponentially damped linear systems, *Journal of Sound and Vibration*, **332**, 1754-1765, 2013.

[31] L. Li, Y. Hu, W. Deng, L. Lü, Z. Ding, Dynamics of structural systems with various frequency-dependent damping models, *Frontiers of Mechanical Engineering*, **10**, 48-63, 2015.

[32] L. Li, Y. Hu, State-Space Method for Viscoelastic Systems Involving General Damping Models, *AIAA Journal*, In Press.

[33] L. Li, Y. Hu, Generalized mode acceleration and modal truncation augmentation methods for the harmonic response analysis of nonviscously damped systems, *Mechanical Systems and Signal Processing*, **52**, 46-59, 2015.

[34] L. Li, Y. Hu, X. Wang, Improved approximate methods for calculating frequency response function matrix and response of MDOF systems with viscoelastic hereditary terms, *Journal of Sound and Vibration*, **332**, 3945-3956, 2013.

[35] L. Li, Y. Hu, X. Wang, Eliminating the modal truncation problem encountered in frequency responses of viscoelastic systems, *Journal of Sound and Vibration*, **333**, 1182-1192, 2014.

[36] L. Li, Y. Hu, X. Wang, L. Ling, Computation of eigensolution derivatives for nonviscously damped systems using the algebraic method, *AIAA journal*, 50, 2282-2284, 2012.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON THE PERFORMANCE OF A COMPUTATIONAL COST REDUCTION TECHNIQUE WHEN APPLIED TO COOLING TOWERS TRANSIENT ANALYSIS

Aram Soroushian[1], Amin Jahani Mehrnoosh[2], Yalda Zarabimanesh[1], Mohammad Hadi Ghondaghsaz[2], Ali Baiani[3], Ali Zakizade[1]

[1]International Institute of Earthquake Engineering and Seismology (IIEES)
S. Lavasani (Farmaniyeh), North Dibaji, West Arghavan, No. 21, Tehran 19537, Iran
e-mail: {a.soroushian, y.zarabimanesh, zakizade.ali}@iiees.ac.ir

[2]West Tehran Branch Islamic Azad University
Ashrafi Esfehani, Bagh-E-Azari, Azari Street, Tehran 14556, Iran
e-mail: {jahani.amin1,hadi.ghondaghsaz}@gmail.com

[3]University of Pooyandegan Danesh
Mazandaran, Chalus, Felestin 46131, Iran
e-mail: Ariancivil@yahoo.com

Keywords: Transient Analysis, Time Integration, Digitized Excitation, Computational Cost, Accuracy, Cooling Tower

Abstract. *Time integration is a versatile tool in transient analysis of structural systems. Considerable computational cost is a weak point for time integration. To reduce the computational cost of analyses against digitized excitations, a technique is proposed in 2008. Experiences on the implementation of the technique in seismic analysis of frames, multistory buildings, silos, bridges, space structures, reservoirs, etc. implies the adequate performance of the technique. The objective in this paper is to study the performance in implementation in seismic analysis of cooling towers. The study is in progress; still the achieved results display the good performance of the technique, when applied to a cooling tower analysis against earthquake records.*

1 INTRODUCTION

Time integration is the most versatile tool to analyze the transient behaviors of semi-discretized structural models. Nevertheless, the computational costs are considerable, and the responses are approximations. In a review on the time integration process, after discretization of the mathematical models in space, by methods such as finite elements, finite volumes, and boundary elements, we arrive at the semi-discretized models, typically stated below:

$$\begin{aligned} &\mathbf{M}\ddot{\mathbf{u}}(t)+\mathbf{f}_{\text{int}}(t)=\mathbf{f}(t), \qquad 0<t\le t_{end} \\ &\text{Initial conditions:} \quad \left|\begin{array}{l}\mathbf{u}(t=0)=\mathbf{u}_0\\ \dot{\mathbf{u}}(t=0)=\dot{\mathbf{u}}_0\\ \mathbf{f}_{\text{int}}(t=0)=\mathbf{f}_{\text{int}_0}\end{array}\right. \\ &\text{Additional constraints: } \mathbf{Q} \end{aligned} \tag{1}$$

where, t and t_{end} imply the time and the duration of the dynamic behavior, $\mathbf{M}$ is the mass matrix, $\mathbf{f}_{\text{int}}(t)$ and $\mathbf{f}(t)$ stand for the vectors of internal force and excitation, $\mathbf{u}$(t), $\dot{\mathbf{u}}(t)$, and $\ddot{\mathbf{u}}(t)$ denote the vectors of displacement, velocity, and acceleration; $\mathbf{u}_0$, $\dot{\mathbf{u}}_0$, and $\mathbf{f}_{\text{int}_0}$ define the initial status of the model (regarding the essentiality of considering $\mathbf{f}_{\text{int}_0}$ in Eqs. (1), see [1, 2]), and $\mathbf{Q}$ represents the nonlinearity restrictions, e.g. see [3, 4]. Time integration determines the status at discrete time instants $\Delta t, 2\Delta t, 3\Delta t$ …. (Δt stands for the integration step size), according to approximate formulations, introducing the integration method [5-8]. Because of the approximation, the integration step size cannot be large; and, because of the step-by-step nature, and the resulting high computational cost, the step size cannot be desirably small. Considering this, and the unconditional stability of conventional time integration methods (see [8, 9]), the integration step size is basically being obtained from:

$$\Delta t=\begin{cases}\dfrac{T}{10} & \text{linear problems}\\ \dfrac{T}{100} & \text{nonlinear problems}\\ \dfrac{T}{1000} & \text{nonlinear problems involved in impact}\end{cases} \tag{2}$$

where, T is the smallest period with considerable contribution in the response [5, 7, 9-11]. In some practical cases, e.g. time integration analysis against earthquake records, where,

$$\mathbf{f}(t)=-\mathbf{M}\,\Gamma\,\ddot{\mathbf{u}}_g(t) \tag{3}$$

and $\ddot{\mathbf{u}}_g(t)$ denotes the vector of ground motion at some different degrees of freedom (unless in multi-support excitations, $\ddot{\mathbf{u}}_g$ reduces to a function of time) and Γ is a matrix implying the effect of ground motion at different degree of freedom [12], $\mathbf{f}(t)$ is available as a vector of digitized records, and Eq. (2) is being replaced with

$$\Delta t=\min\left({}_f\Delta t,\frac{T}{\chi}\right) \tag{4}$$

where, $_f\Delta t$ implies the step size, by which, $\mathbf{f}(t)$ is digitized, and in view of Eq. (2),

$$\chi = \begin{cases} 10 & \text{linear problem} \\ 100 & \text{nonlinear problem} \\ 1000 & \text{nonlinear problem involved in impact} \end{cases} \tag{5}$$

When, $_f\Delta t$ governs Eq. (4), some additional computational cost, e.g. about

$$\frac{T - \chi_f \Delta t}{\chi_f \Delta t} > 0 \tag{6}$$

for linear problems is being dictated to the analysis, specially, in order to take into account the total information of $\mathbf{f}(t)$; see also [13].

In 2008, a technique is proposed for reducing the above-mentioned additional considerable computational cost, while taking into account the total earthquake record, and not sacrificing considerable accuracy (compared to analyses with steps obtained from Eq. (4)) [13-16]. The past experiences on implementation of the technique were successful; see Table 1. In continuation of the different case studies reported in Table 1 and with attention to the importance

System(s) analyzed	The cost reduction (%)	Source
SDOF system	75	[14]
2-DOF nonlinear system	49.27	[14]
Eight storey shear frame	80	[17]
Thirty-storey building	50	[18]
3-component earthquakes	66.7	[19]
Silo	77.65	[20, 21]
Water tank	66.7	[16, 22]
Building in pounding	12.7	[23]
Bridge with linear and nonlinear behaviors	45-80	[24-26]
Power stations	> 50	[27]
Regular residential buildings	50-87	[28, 29]
Bridges with pre-stressed elements, subjected to multi-support excitation and nonlinearities	30-70	[30]
Residential building with irregularities in height	50-80	[31]
Space Structures	>50	[32]
A telecommunication tower	>50	[33]

Table 1: Experiences on the computational cost reduction technique proposed in [14].

of cooling towers in industry as well as the considerably large and complicated structure systems (with high number of degrees of freedom) of cooling towers, in this paper, the authors intention is to answer to the following questions:

1. Whether for cooing tower, $_f\Delta t$ can be the governing term in Eq. (4).

2. Whether, if the response to the first question is positive, the performance of the technique proposed in [14] can be adequate, when implemented in transient analysis of cooling towers, against earthquake records.

A real cooling tower, two real ground motions and an integration method, are taken into account as the basis of the study. The models, and the structural behaviors, are briefly addressed in Section 2; the consequence of implementation of the computational cost reduction technique is reported in Section 3; the observations are briefly discussed in Section 4; and finally, the paper is closed in Section 5 with a brief set of the conclusions.

2 THE MODELS

The cooling tower, displayed in Fig.1, is constructed in Gazvin, Iran, with the construction end at about 2011, and the main structural properties as addressed in Table 2. (There is no special reason for the selection of this cooling tower, unless the availability of the computer finite foment model.) With attention to the existing recommendations regarding analysis methods in studies of tall structural system [35], the ground motions are considered, as displayed in Fig. 2, where $\ddot{u}_g$ stands for the ground acceleration, and g implies the gravity constant. The almost exact responses obtained from time integration with very small integration steps are as displayed in Fig. 3 and linear in view of the superposition principle [36].

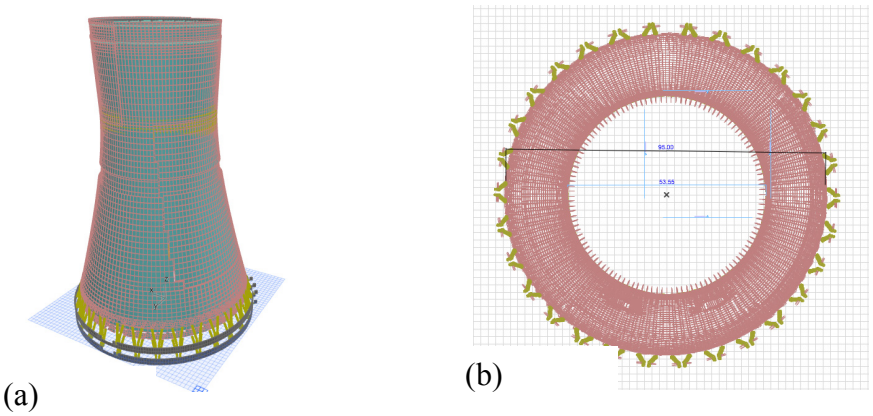


Figure 1: The cooling tower under study: (a) 3 D appearance, (b) planer cross section from top.

Height (m)	External Diameter (m)			Wall thickness (m)	Occupancy	Soil type	Structural system	Finite element
	Base	Top	Mid height					
146	95	59.5	49.5	3	Power Plant (steam)	III [34]	Concrete Shell with Steel Cover	Shell

Table 2: Main structural properties of the cooling tower in Fig. 1.

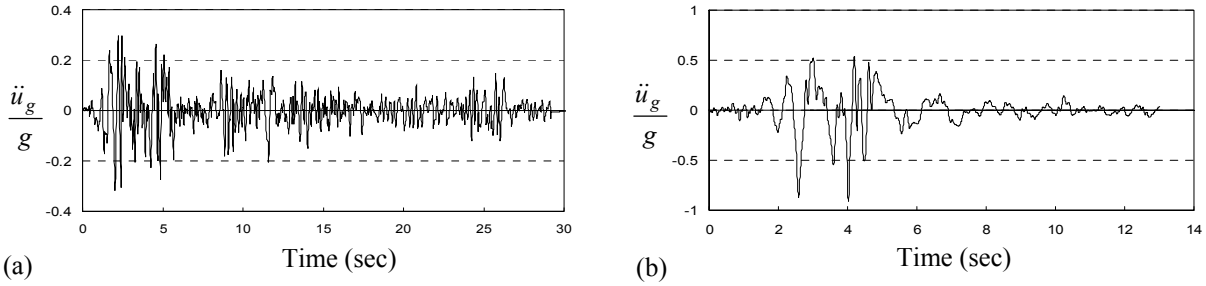


Figure 2: The ground strong motions taken into account: (a) an E1 Centro record, (b) a Kobe record.

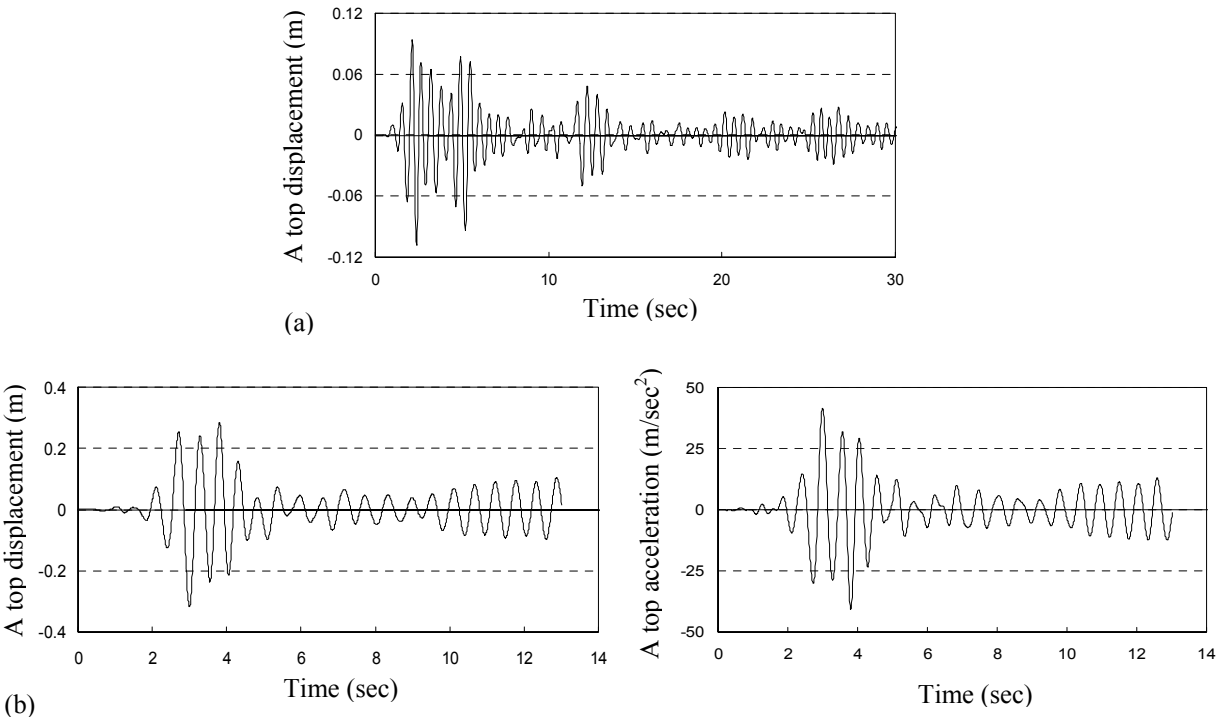


Figure 3: Almost exact responses against: (a) the ground motion in Fig. 2(a), (b) the ground motion in Fig. 2(b).

And finally for time integration analyses (numerical model), the average acceleration method of Newmark [37] with the formulation below [5, l2, 37]:

$$
\begin{aligned}
&\mathbf{M}_i\ddot{\mathbf{u}}_i + \mathbf{C}_i\dot{\mathbf{u}}_i + \mathbf{K}_i\mathbf{u}_i = \mathbf{f}_i \\
&\mathbf{u}_i = \mathbf{u}_{i-1} + \Delta t\,\dot{\mathbf{u}}_{i-1} + \frac{\Delta t^2}{2}\left[(1-2\beta)\ddot{\mathbf{u}}_{i-1} + 2\beta\ddot{\mathbf{u}}_i\right] \\
&\dot{\mathbf{u}}_i = \dot{\mathbf{u}}_{i-1} + \Delta t\,\dot{\mathbf{u}}_{i-1} + \Delta t\left[(1-\gamma)\ddot{\mathbf{u}}_{i-1} + \gamma\ddot{\mathbf{u}}_i\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1,2,3\ldots \\
&\gamma = 2\beta = 0.5 \\
&\mathbf{u}_0 = \mathbf{u}(t=0) \\
&\dot{\mathbf{u}}_0 = \dot{\mathbf{u}}(t=0) \\
&\ddot{\mathbf{u}}_0 = \mathbf{M}_1^{-1}\left(\mathbf{f}_0 - \mathbf{C}_1\dot{\mathbf{u}}_0 - \mathbf{K}_1\mathbf{u}_0\right) \\
&\mathbf{f}_0 = \mathbf{f}(t=0)
\end{aligned}
\tag{7}
$$

($\mathbf{M}_i$, $\mathbf{C}_i$, and $\mathbf{K}_i$ respectively represent the mass damping and stiffness matrices at the i th step, γ and β stand for the parameters of the Newmark method, and each top dot implies once differentiation with respect to time) is considered as the analysis tool.

3 RESULTS OF THE NUMERICAL STUDY

As mentioned in the ending lines of Section 2, the behaviors of the cooling tower against both records are linear. Accordingly, in Eq. (4),

$$\chi = 10 \tag{8}$$

With attention to Fig. 3, the values of T are addressed in Table 3. Considering Table 3 and ${}_f\Delta t = 0.02, 0.01$ respectively for the records in Figs. 2(a) and 2(b), obviously ${}_f\Delta t$ is smaller than $\frac{T}{\chi}$ in all cases. Therefore, the response to the first question in Section 1 is positive and we are to proceed to the second question.

Response	Fig. 2(a) as Earthquake record	Fig. 2(b) as Earthquake record
A top displacement	0.25	0.5
A top acceleration	-	0.5

Table 3: Values of T in different cases.

In view of the above mentioned values of $T, \chi, {}_f\Delta t$ and the recommendations on the maximum values to be assigned to the enlargement values [24], n

$$n = \frac{\Delta t}{{}_f\Delta t} \tag{9}$$

the selections below:

$$n = 2,3,4,5 \tag{10}$$

are considered in the numerical study. The finite element model of the cooling tower is analyzed by the average acceleration method of Newmark [5, 12, 37], with integration steps equal to

$$\Delta t = n \ {}_f\Delta t \tag{11}$$

while, for changing the records digitized at steps sized Δt to records digitized at $n_f\Delta t$ the technique proposed in [14] is implemented. The resulting responses are as displayed in Figs. 4-6, respectively for top displacements obtained under the effect of the El Centro (Fig. 2(a)) and the Kobe (Fig. 2(b)) record, and a top acceleration because of the Kobe record (Fig. 2(b)). The corresponding reductions of computational costs are addressed in Table 4. Obviously, the performance is adequate. Even, the inaccuracy in the case $n = 5$ and partly $n = 4$, more considerable for analysis against the record in Fig 2 (b) is acceptable, with attention to the fact that with a closer look to Figs. 3,

$$\frac{T}{\chi \ {}_f\Delta t} \begin{cases} < 2 & \text{for the excitation in Fig. 2(a)} \\ \cong 3.2 & \text{for the excitation in Fig. 2(b)} \end{cases} \tag{12}$$

Consequently in response to the second question in Section 1, we can claim that the performance of the computational cost reduction technique can be adequate in transient analysis of cooling towers' structural system against ground motions.

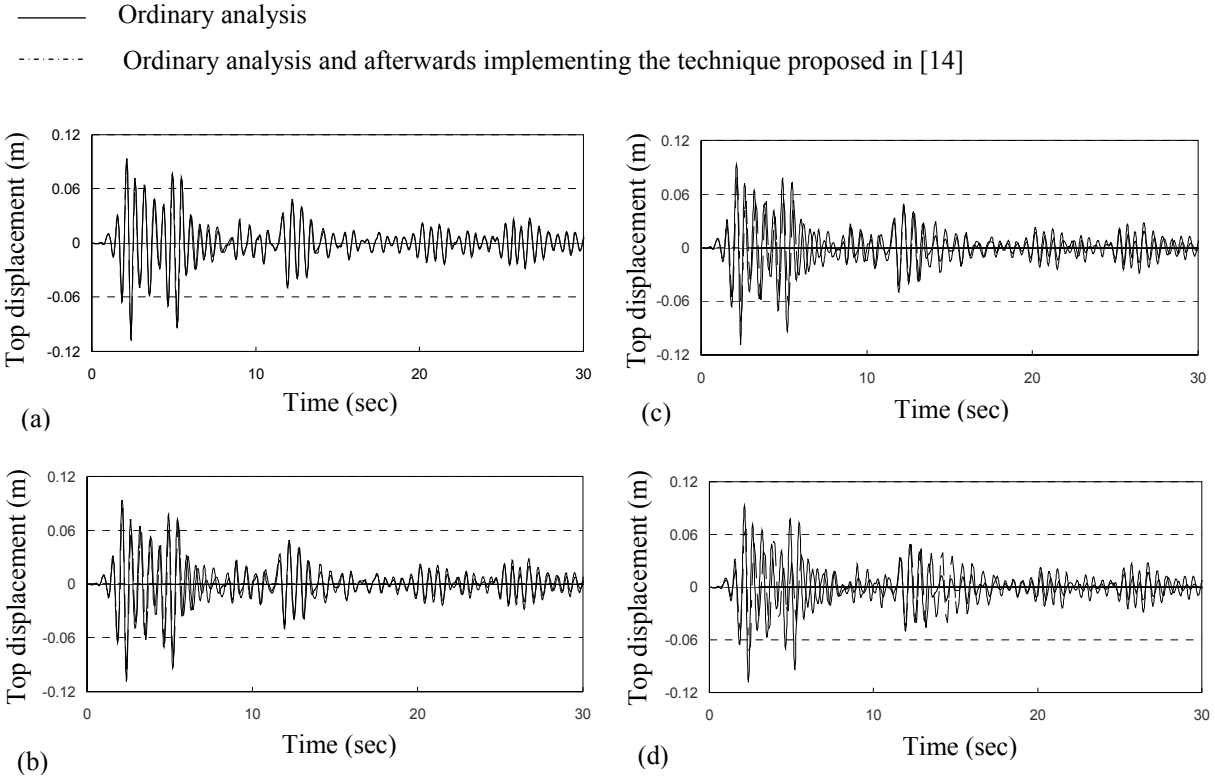


Figure 4: A top displacement obtained from average acceleration analysis of the model introduced in Fig. 1 and Table 2 against the record in Fig. 2(a): (a) $n=2$, (b) $n=3$, (c) $n=4$, (d) $n=5$.

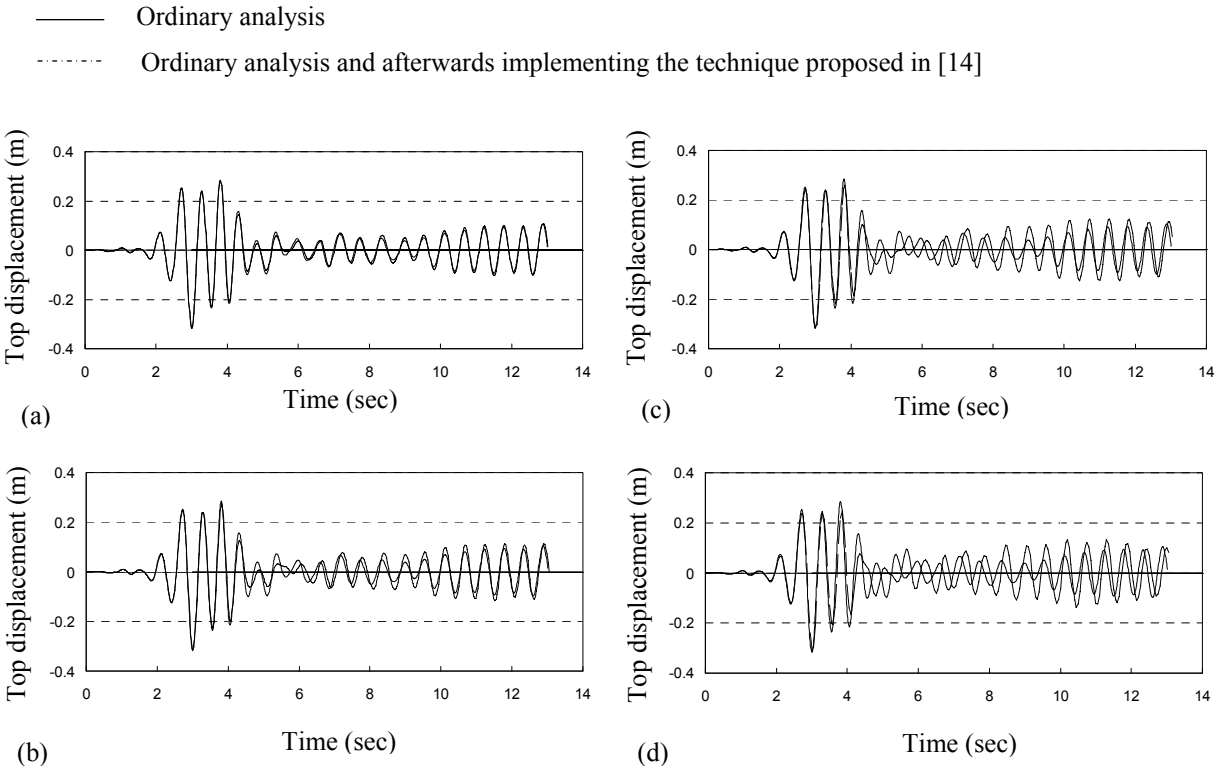


Figure 5: A top displacement obtained from average acceleration analysis of the model introduced in Fig. 1 and Table 2 against the record in Fig. 2(b): (a) $n=2$, (b) $n=3$, (c) $n=4$, (d) $n=5$.

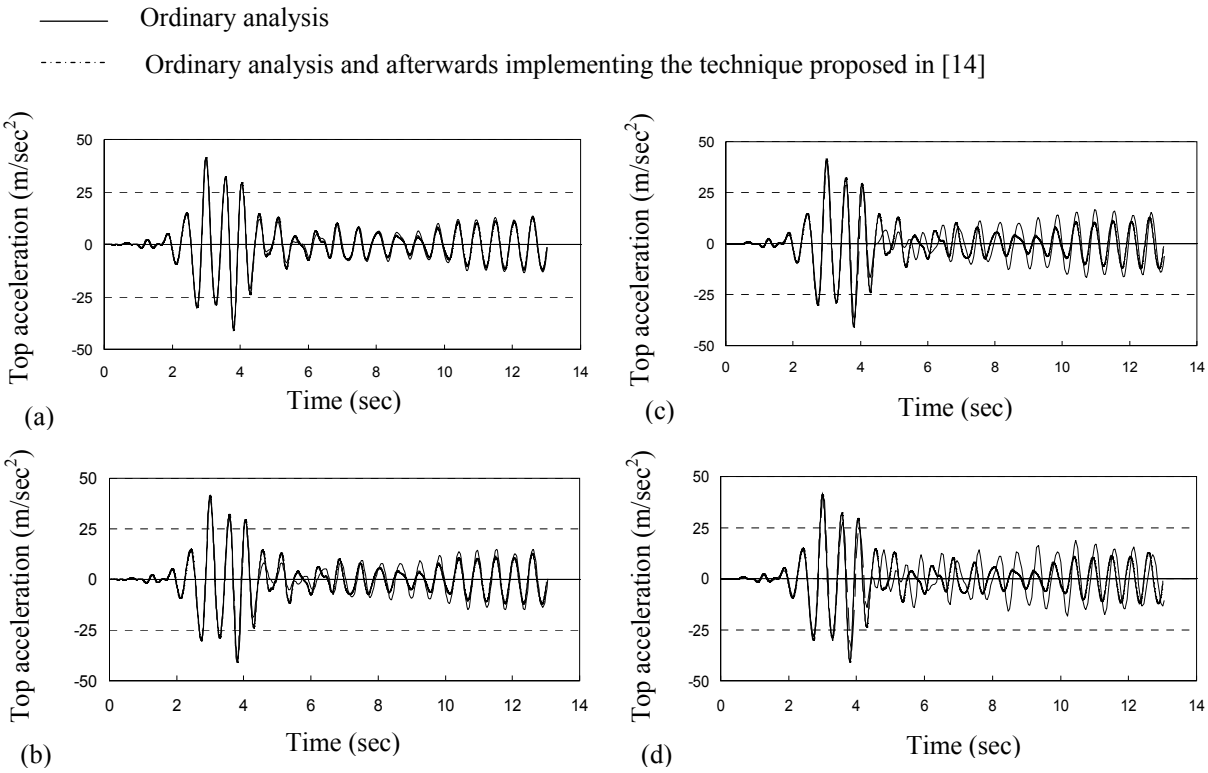


Figure 6: A top acceleration obtained from average acceleration analysis of the model introduced in Fig. 1 and Table 2 against the record in Fig. 2(b): (a) $n=2$, (b) $n=3$, (c) $n=4$, (d) $n=5$.

n	1	2	3	4	5
Reduction (%)	0	50	67	75	80

Table 4: Reduction of computational costs in Figs. 4-6.

4 BRIEF DISCUSSION

The observations in Section 3 are evidences for adequate performance of the technique proposed in [14]. However, with attention to the vagueness and complexity in the notion and computation of T, the fact that the value of ${}_f\Delta t$ might become smaller in future (see [38, 39]), and finally the less accuracy for the cases $n=4$ and $n=5$ when subjected to the excitation in Fig. 2(b), it is reasonable to continue investigations towards advanced techniques for integration step size enlargement, such that the reduction of computational cost can be provided in the price of even less sacrifice of accuracy. For further extensions, it is also essential to study other cases of cooling towers, earthquake records and integration methods.

5 CLOSURE

• In analysis of cooling towers against ground motion records, the step size needed for the accuracy of integration might be larger than the record step size.

• Implementation of the technique proposed in [14] for reducing the cost of transient analysis of cooling towers against earthquake records might considerably reduce the computational cost in the price of negligible loss of accuracy.

• Further investigations both to enhance the performance of the technique proposed in [14] and also to extend the numerical study is essential and recommended.

REFRENCE

[1] H. Gavin, *Structural dynamics*. Class Notes CE 283, Duke University, 2001.

[2] A. Soroushian, *New methods to maintain responses' convergence and control responses' errors in the analysis of nonlinear dynamic models of structural systems*. PhD Dissertation, University of Tehran, 2003. (in Persian)

[3] P. Wriggers, *Computational contact mechanics*. John Wiley & Sons, 2002.

[4] T.J.R. Hughes, K.S. Pister, R.L. Taylor, Implicit-explicit finite elements in nonlinear transient analysis. *Computer Methods in Applied Mechanics and Engineering,* **17/18**, 159-182, 1979.

[5] R.W. Clough, J. Penzien, *Dynamics of structures*. McGraw-Hill, 1993.

[6] J. Argyris, J.P. Mlejnek, *Dynamics of structures*. Elsevier, 1991.

[7] K.J. Bathe, *Finite element procedures*. Prentice-Hall, 1996.

[8] T. Belytschko, T.J.R. Hughes, *Computational methods for transient analysis*. Elsevier, 1983.

[9] T.J.R. Hughes, *The finite element method: linear static and dynamic finite element analysis*. Prentice-Hall, 1987.

[10] J.F. McNamara, Solution schemes for problems of nonlinear structural dynamics. *Journal of Pressure Vessels (ASME)*, **96**, 149-155, 1974.

[11] NZS, *Structural design actions, Part 5: Earthquake actions-New Zealand*. New Zealand Standard, 2004.

[12] A.K. Chopra, *Dynamics of structures: theory and application to earthquake engineering*. Prentice-Hall, 1995.

[13] A. Soroushian, Integration step size and its adequate selection in seismic dynamic analysis of structural systems. *In* Computational methods in earthquake engineering, vol. 3. *Edited by* M. Papadrakakis and N. Lagaros and V. Plevris. Springer. (accepted for publication)

[14] A. Soroushian, A technique for time integration with steps larger than the excitation steps. *Communications in Numerical Methods in Engineering*, **24**, 2087-2111, 2008.

[15] A. Soroushian, *Direct time integration with step larger than the steps by which the excitations are digitized*, *Report 7510*. International Institute of Earthquake Engineering and Seismology (IIEES), 2011. (In Persian)

[16] A. Soroushian, P. Farshadmanesh, S. Azad, On the essentiality of techniques to enlarge analysis steps in time integration against digitized excitations. *International Journal for Earthquake Engineering and Seismology (JSEE)*, **17**, 2015. (accepted for publication)

[17] A. Soroushian, On the performance of a conventional accuracy controlling method applied to linear and nonlinear structural dynamics. Proc., *17th International Congress on Sound & Vibration (ICSV17)*, Cairo, Egypt, 2010.

[18] A. Soroushian, A. Aziminejad, A more efficient seismic analysis of tall buildings by implementing a recently proposed technique. Proc., *6th International Conference Seismology and Earthquake Engineering (SEE6)*, Tehran, Iran, 2011.

[19] O. Bahar, S. Ramezani, Faster time integration analysis for building structures subjected to 3-component earthquakes. Proc., *3rd III ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMDYN 2011)*, Corfu, Greece, 2011.

[20] F. Nateghi, M. Yakhchalian, On less computational costs for analysis of silos seismic behaviors by time integration. Proc., *3rd ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2011)*, Corfu, Greece, 2011.

[21] F. Nateghi, M. Yakhchalian, (2011) An investigation into the effectiveness of a technique proposed for reducing computational cost of time integration in the analysis of silos seismic behaviors. Proc., *11th US National Congress on Computational Mechanics (USNCCM 2011)*, Minneapolis, USA, 2011.

[22] A. Soroushian, A. Saaed, M. Arghavani, M. Rajabi, M.M. Sharifpour, Less computational costs in the analysis of reservoirs seismic behaviours by time integration. Proc., *10th Biennial Conference on Vibration Problems (ICoVP-2011)*, Prague, Czech Republic, 2011.

[23] A. Soroushian, On the performance of a recent technique for more efficient time integration in severe seismic conditions. Proc., *1st International Conference on Advances in Structural Engineering and Mechanics (ASEM'11)*, Seoul, South Korea, 2011.

[24] S. Azad, *A study on accelerating time history analysis of bridges*, M.Sc. Dissertation, International Institute of Earthquake Engineering and Seismology (IIEES), Iran. (In Persian), 2015.

[25] A. Soroushian, M. Hosseini, A. Vasseghi, On the performance of a technique for more efficient time integration when applied to bridge structures seismic analysis Proceedings, *3rd ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering*, Corfu, Greece, May, 2011.

[26] A. Soroushian, A. Vasseghi, M. Hosseini, On practical performance of a technique for more efficient dynamic analysis in view of real seismic analysis of a bridge structural system. *In* Computational methods in earthquake engineering, *Edited by* M. Papadrakakis and M. Fragiadakis and V. Plevris. Springer, The Netherlands, 2013.

[27] M. Bastami, A technique for more efficient time integration applied to seismic analysis of power substation equipment. Proc., *11th World Congress on Computational Mechanics (WCCM XI)*, Barcelona, Spain, 2014.

[28] A. Sabzei, *On the performance of a recent technique for seismic analyses computational cost reduction when applied to buildings structural systems*, M.Sc. Dissertation,

International Institute of Earthquake Engineering and Seismology (IIEES), Iran. (In Persian), 2013.

[29] A.Y. Reziakolaei, A.Sabzei, A. Soroushian, On the performance of a structural analysis cost reduction technique when applied to residential buildings. Proc., *4th ECCOMAS Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2013)*, Kos Island, Greece, 2013.

[30] S. Azad, A. Soroushian, S. Malek, On the performance of a computational cost reduction technique when applied to bridges multi-support excitations, SEE7, May (in Persian), 2015.

[31] A.A. Hadad, *Reducing computational costs in time integration analyses of buildings with irregularities in height because of mass*, M.Sc. Dissertation, International Institute of Earthquake Engineering and Seismology (IIEES), Iran. (In Persian), 2015.

[32] A. Garakaninezhad, R.K. Moghadas, On the performance of a technique to accelerate time integration when applied to space structures analyses. Proc., *5th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2013)*, Kos, Greece, June, 2015.

[33] A. Soroushian, S.R. Mirghaderi, Y. Zarabimanesh, A.J. Mehrnoush, A. Z. Zadeh, M.H. Ghondaghsaz, On reduction of computational costs in transient analysis of telecommunication towers. Proc., *12th World Congress on Computational Mechanics (WCCM XII*), Seoul, July, 2016.

[34] BHRC, *Iranian Code of Practice for Seismic Resistant Design: Standard No. 2800–*. Iran, 2015. (in Persian)

[35] S.E. Azam, S. Mariani, Investigation of computational and accuracy issues in POD-based reduced order modeling of dynamic structural systems. *Engineering Structures*, **54**, 150-167, 2013.

[36] L. Meirovitch, *Fundamentals of vibrations*. McGraw Hill, 2001.

[37] N.M. Newmark, A method of computation for structural dynamics, *Journal of Engineering Mechanics (ASCE)* **85**, 67-94, 1959.

[38] J. Havskov, G. Alguacil, *Instrumentation in Earthquake Seismology (Modern Approaches in Geophysics)*. Springer, 2005.

[39] Earthquake Planning and Protection Organisation, *Code GS (2000) Earthquake resistant design of structures*. Greece, 2000.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A DIFFERENT LOOK AT THE RICHARDSON EXTRAPOLATION LEADING TO A NEW PROPOSITION

A. Soroushian

Structural Engineering Research Centre, International Institute of Earthquake Engineering and Seismology (IIEES)
S. Lavasani (Farmaniyeh), North Dibajee, West Arghavan, No. 21, Tehran 19537, Iran

Keywords: Approximate Computation, Richardson Extrapolation, Convergence, Accuracy, Order of Accuracy, Taylor Series.

Abstract. *Richardson extrapolation is proposed about a century ago and broadly accepted in the scientific community as a mean to accelerate the convergence of approximate computations. Many implementations, generalizations, and applications, of Richardson extrapolation, are reported in the literature. In this paper, considering the Taylor series expansion of approximate computations with respect to the algorithmic parameters, Richardson extrapolation is introduced as a tool towards higher accuracies, and the higher orders of accuracy are achieved as by-products. On this basis, a new proposition on Richardson extrapolation is stated, explained, and demonstrated, in an example.*

1 INTRODUCTION

Richardson extrapolation [1, 2], first introduced in the starting decades of the twentieth century, is a technique to accelerate sequential computations approach towards exact solutions, i.e. increase the order of accuracy and rate of convergence [3, 4]. This is a significant achievement addressed in the literature as the most usefulness of Richardson extrapolation for practical computations or turning lead to gold [5, 6]. Some of the applications of Richardson extrapolation are different advanced integration methods, e.g. the Romberg method, and a recent error estimation approach [4, 6]. Order of accuracy and its increase would lose its meaning without convergence. However, convergence is an essentiality in approximate computations [4, 7, 8] (and this highlights the applicability and versatility of Richardson extrapolation. Convergence implies the potential of arbitrary closeness to the exact solutions, i.e.

$$\lim_{\lambda \to 0} \mathbf{U}^a = \mathbf{U} \tag{1}$$

where, $\mathbf{U}^a$, and $\mathbf{U}$ respectively stand for the approximately computed and exact solutions, and λ is the algorithmic parameter [4, 9, 10]. Pictorial representation of Eq. (1) is as noted in the convergence plot displayed in Fig. 1, where, E stands for the error [11], i.e.

$$E = \left\| \mathbf{U}^a - \mathbf{U} \right\| \tag{2}$$

q denotes the rate of convergence, such that

$$q \le q_0 \tag{3}$$

q_0 represents the order of accuracy and the difference between q_0 and q is well studied and under control; e.g. see [10, 12, 13], the mid-section of the convergence plot sloped $\cong q$ in Fig. 1 is the region of proper convergence [14], with a length that depending on the problem, computational method, and round-off, might even reduce to zero and vanish. The effect of Richardson extrapolation on convergence is as typically displayed in Fig. 2; see also [4]. As apparent in Fig. 2, Richardson extrapolation seems to have the potential to add the accuracy besides the order of accuracy. Based on this consideration, the attempt in this paper is to re-derive the Richardson extrapolation as a technique to provide more accuracy, and to display the resulting higher order of accuracy, as a by-product, and finally, with attention to the essentialities of a specific application, to arrive at a new proposition on Richardson extrapolation, present an example, and discuss the practical consequences.

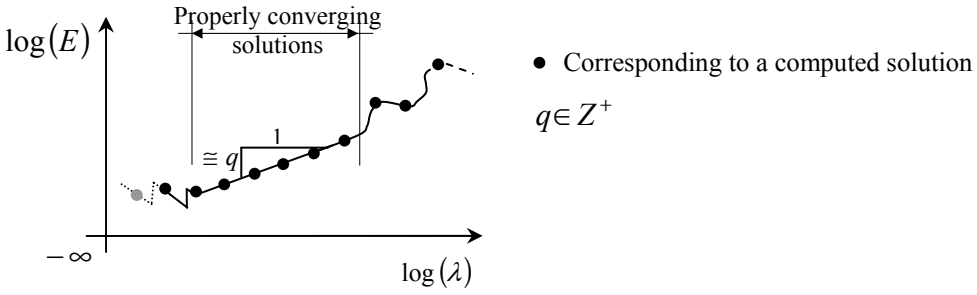


Figure 1: Typical changes of solutions errors with respect to the algorithmic parameters.

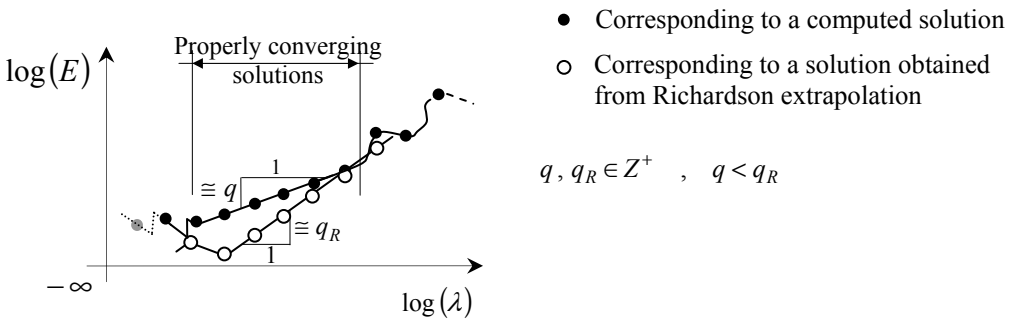


Figure 2: Typical effect of Richardson extrapolation on solutions convergence.

2 AN ACCURACY BASIS FOR RICHARDSON EXTRAPOLATION

Almost any approximate computation can be stated as noted below:

$$\mathbf{U}^a = \mathbf{F}\left(\lambda, p_1, p_2, \ldots p_n, \gamma_{i=1,2,\ldots n'}\right) \tag{4}$$

where, $p_{i=1,2,\ldots n}$ introduce the actual parameters defining the mathematical model, $\gamma_{i=1,2,\ldots n'}$ define the details of the numerical method (model) and λ stands for the algorithmic parameter, defined, via Eq. (1). By extending the definition of λ to a continuously changing parameter and considering the fact that converging approximate computations are generally well behaved at close neighborhood of the exact solutions, we can express $\mathbf{U}^a$ in terms of the Taylor series expansion at $\mathbf{U}^a\left(\lambda \to 0^+\right)$, to arrive at

$$\begin{aligned} \mathbf{U}^a &= \mathbf{U} + \sum_{n=1}^{\infty} \frac{s_n}{n!} \lambda^n \\ \mathbf{s}_n &= \left. \frac{\partial^n \mathbf{F}}{\partial \lambda^n} \right|_{\lambda \to 0^+} \\ & \qquad\qquad n = 1,2,3,\ldots \end{aligned} \tag{5}$$

also equivalent to:

$$\begin{aligned} &\mathbf{U}^a - \mathbf{U} = \mathbf{C}_0 \lambda^q + \mathbf{C}_1 \lambda^{q+1} + \ldots \qquad q \in Z^+ \\ &\mathbf{C}_i = \frac{\mathbf{s}_{q+i}}{(q+i)!} \qquad i = 0,1,2,3,\ldots \end{aligned} \tag{6}$$

where,

$$\begin{aligned} &\mathbf{s}_n = \overline{\mathbf{O}} \qquad , \qquad n = 1,2,3,\ldots q-1 \\ &\mathbf{s}_q \neq \overline{\mathbf{O}} \end{aligned} \tag{7}$$

and even in a more detailed expression

$$\begin{aligned} &\mathbf{U}^a - \mathbf{U} = \mathbf{C}'_0 \lambda^{q_0} + \mathbf{C}'_1 \lambda^{q_1} + \mathbf{C}'_2 \lambda^{q_2} + \ldots \\ &\mathbf{C}'_i = \frac{\mathbf{s}_{q_i}}{q_i!} \quad , \quad i = 1,2,\ldots \end{aligned} \tag{8}$$

where,

$$\mathbf{s}_n = \overline{\mathbf{O}} \quad , \quad n \neq q_i$$
$$\mathbf{s}_n = \left.\frac{\partial^{q_i}\mathbf{F}}{\partial\lambda^{q_i}}\right|_{\lambda\to 0^+} \quad , \quad n = q_i \qquad i = 1,2,\ldots \tag{9}$$

and

$$q_0 < q_1 < q_2 < \ldots \tag{10}$$

In view of Eqs. (6) or (8), for sufficiently small values of λ, the error as introduced in Eq. (2), mainly depend on $\mathbf{C}'_0$, not on $\mathbf{C}'_1$, $\mathbf{C}'_2$, . . . In more detail, by gradually decreasing the value assigned to λ, the role of $\mathbf{C}'_i$ decreases starting from the higher values of i (with higher exponents of λ), and in some stage, it is only $\mathbf{C}'_0$ that contributes E (in continuation, the role of $\mathbf{C}'_0$ will also vanish at convergence; see. Eq. (1)). This is in complete consistence with Fig. 1 and specifically the mid-section implying proper convergence. The role of $\mathbf{C}'_{i=1,2,..}$ is never zero; however, in view of Eq. (10), is negligible, depending on the problem at hand and the difference between q_0 and $q_{i=1,2,...}$. Accordingly, if in defining a new computational method we can eliminate the term $\mathbf{C}'_0\lambda^{q_0}$ from the right hand side of Eq. (8), and at the same time, guarantee a reasonably moderate change in $\mathbf{C}'_1$, we would be able to expect more accuracy from the resultant of the elimination.

With this idea, consider two computations for one problem by one computational method,

$$\begin{aligned}\mathbf{U}_1^a &= \mathbf{U}^a(\lambda_1) = \mathbf{U} + \mathbf{C}'_0\lambda_1{}^{q_0} + \mathbf{C}'_1\lambda_1{}^{q_1} + \ldots \\ \mathbf{U}_2^a &= \mathbf{U}^a(\lambda_2) = \mathbf{U} + \mathbf{C}'_0\lambda_2{}^{q_0} + \mathbf{C}'_1\lambda_2{}^{q_1} + \ldots\end{aligned} \tag{11}$$

and define a new computation, ${}^R\mathbf{U}_1^a$, for that problem, as stated below:

$${}^R\mathbf{U}_1^a = \alpha\mathbf{U}_1^a + \beta\mathbf{U}_2^a \tag{12}$$

where, α and β are constants to be determined. In view of Eqs. (5) and (11), in order to preserve convergence to the exact solution $\mathbf{U}$, it is essential to be restricted to

$$\alpha + \beta = 1 \tag{13}$$

and since, in Eq. (8), $\mathbf{C}'_0 \neq \overline{\mathbf{O}}$ (see Eq. (1)), in order to omit the terms with exponent q_0, it is essential to satisfy

$$\alpha\lambda_1{}^{q_0} + \beta\lambda_2{}^{q_0} = 0 \tag{14}$$

Consequently, from Eqs. (13) and (14),

$$\alpha = \frac{\lambda_2{}^{q_0}}{\lambda_2{}^{q_0} - \lambda_1{}^{q_0}} \quad , \quad \beta = \frac{\lambda_1{}^{q_0}}{\lambda_1{}^{q_0} - \lambda_2{}^{q_0}} \tag{15}$$

and hence from Eq. (12),

$${}^R\mathbf{U}_1^a = \frac{\lambda_2{}^{q_0}\mathbf{U}_1^a - \lambda_1{}^{q_0}\mathbf{U}_2^a}{\lambda_2{}^{q_0} - \lambda_1{}^{q_0}} = \frac{\lambda_1{}^{q_0}\mathbf{U}_2^a - \lambda_2{}^{q_0}\mathbf{U}_1^a}{\lambda_1{}^{q_0} - \lambda_2{}^{q_0}} \tag{16}$$

is a candidate for more accurate solution. The resulting change in the term $\mathbf{C}_1'\lambda^{q_1}$ is as noted below (${}^{R}\mathbf{C}_1'$ is the coefficient of second converging term in the Taylor series expansion of ${}^{R}\mathbf{U}_1^a$ with respect to λ; see also Eqs. (6), (8) or (11)) :

$$\frac{\left\|{}^{R}\mathbf{C}_1'\right\|}{\left\|\mathbf{C}_1'\right\|}=\frac{\left|\lambda_1{}^{q_0}\lambda_2{}^{q_1}-\lambda_2{}^{q_0}\lambda_1{}^{q_1}\right|}{\left|\lambda_1{}^{q_0}-\lambda_2{}^{q_0}\right|}\frac{1}{\left|\lambda_1{}^{q_1}\right|} \tag{17}$$

which considering

$$\lambda_1=r\lambda_2 \qquad , \qquad r>1 \tag{18}$$

leads to

$$\frac{\left|{}^{R}\mathbf{C}_1'\right|}{\left|\mathbf{C}_1'\right|}=\frac{\left|r^{q_0}-r^{q_1}\right|}{\left|r^{q_0}-1\right|}\frac{1}{\left|r^{q_1}\right|}<1 \tag{19}$$

that taking into account that λ is small (at convergence) and r and q_0 are finite, implies the decrease of $\mathbf{C}_1'$, specifically in the mid-section in Fig. 1. Accordingly, for sufficiently small values of λ, specifically in the mid-section in Fig. 1, the ${}^{R}\mathbf{U}_1^a$ defined in Eq. (16) will be more accurate than $\mathbf{U}_1^a$. For ${}^{R}\mathbf{U}_1^a$ the Taylor series expansion can be stated as noted below:

$${}^{R}\mathbf{U}_1^a=\mathbf{U}+\mathbf{C}_1''\lambda^{q_1}+\mathbf{C}_2''\lambda^{q_2}+\mathbf{C}_3''\lambda^{q_3}+\ldots \tag{20}$$

where, $\mathbf{C}''$ can be simply computed in a process such that implemented in arriving at Eq. (19), and it is worth noting that, in consistence with Eqs. (17)-(19),

$$\mathbf{C}_i''={}^{R}\mathbf{C}_i' \tag{21}$$

In view of Eq. (10), Eq. (20) displays the higher order of accuracy for the computation:

$${}^{R}\mathbf{U}_1^a={}^{R}\mathbf{F}(\lambda,p_i,\ldots) \tag{22}$$

Therefore, it is correct to claim that, ${}^{R}\mathbf{U}_1^a$, being defined based on increasing accuracy (compared to $\mathbf{U}_1^a$), also leads to higher order of accuracy. With attention to Eq. (16), Eq. (21) coincides the Richardson extrapolation addressed in the literature [1-4], herein derived using the Taylor series expansion of a computation (in terms of the algorithmic parameter) on the basis of accuracy. A more rigorous explanation, especially for higher orders of Richardson extrapolation is the fact that as implied in Eqs. (12)-(16), for Richardson extrapolation of order J, J terms are to be omitted from the Taylor series by solving a linear $(J+1)\times(J+1)$ algebraic system. However, with attention to the role and restrictions on λ and λ^{q_i}, there is only one set of $\alpha_1, \alpha_2, \ldots \alpha_{J+1}$ to appropriately combine the solutions $\mathbf{U}^a(\lambda_1)$, $\mathbf{U}^a(\lambda_2)$, $\ldots$ $\mathbf{U}^a(\lambda_{J+1})$, and hence, any other appropriate combination of $\mathbf{U}^a(\lambda_1)$, $\mathbf{U}^a(\lambda_2)$, $\ldots$ $\mathbf{U}^a(\lambda_{J+1})$ should not exist, and consequently, the proposed approach and Richardson extrapolation lead to identical formulation regardless of the order of Richardson extrapolation.

Finally, it is worth noting that the presented derivation has also a geometrical explanation. In view of Fig. 1, for enhancement of accuracy in the proper convergence region of the convergence plot (more valid for smaller values of λ), there is no way but to make the slope of

convergence trend more steep; the opposite is also correct. Albeit Richardson extrapolation is merely one alternative to enhance the accuracy and order of convergence; see also [15].

3 THE RESULTING PROPOSITION

In Section 2, Richardson extrapolation is introduced as the process of elimination of non-zero converging terms from the Taylor series expansion of the computation with respect to the algorithmic parameter, by linear combination of the computations at different values of the algorithmic parameters, while the elimination is started from the smallest exponents. This is summarized in the relations below:

$$\begin{gathered}\mathbf{U}^a = \mathbf{F}\left(\lambda, \mathbf{C}_0\left(p_1, p_2, , \ldots, \gamma_{i=1,2,\ldots n'}\right), \mathbf{C}_1\left(p_1, p_2, , \ldots, \gamma_{i=1,2,\ldots n'}\right), \ldots\right) \\ \Downarrow \quad \text{Richardson Extrapolation of order } J \\ {}^{R_J}\mathbf{U}^a = {}^{R_J}\mathbf{F}\left(\lambda, \mathbf{C}'_J\left(p_1, p_2, , \ldots, \gamma_{i=1,2,\ldots n'}\right), \mathbf{C}'_{J+1}\left(p_1, p_2, , \ldots, \gamma_{i=1,2,\ldots n'}\right), \ldots\right)\end{gathered} \tag{23}$$

where, J is a positive integer, denoting the order of Richardson extrapolation, and meanwhile, the original computation can be considered as the Richardson extrapolation of order zero, i.e.

$${}^{R_0}\mathbf{U}^a = \mathbf{U}^a \tag{24}$$

all in consistence with the formulation presented for Richardson extrapolation [1-4]. In other words, for arriving at Richardson extrapolation of order J, the ordinary computation need to be considered at J+1 values of λ and then the computations should be combined linearly, such that to preserve convergence and eliminate the J terms $\mathbf{C}'_i\lambda^{q_i}, i = 0,1,2,\ldots J-1$ from the combination. From this perspective, it seems obvious that the J terms can be eliminated in many ways, e.g. all together with one computation, one by one in J separate stages, etc., all resulting in identical results [6, 16]. Consequently, the J th order Richardson extrapolation is identical to the 1st order Richardson extrapolation of the J-1 th order Richardson extrapolation.

The above idea, or to say better claim, is true, because: (1) linear combination of linearly combined computations is a linear combination of the starting computations, (2) linear combination of $J+1$ terms such that to preserve convergence and omit $\mathbf{C}'_0\lambda^{q_0}, \mathbf{C}'_1\lambda^{q_1}, \ldots \mathbf{C}'_{J-1}\lambda^{q_{J-1}}$, is unique, and hence, there is no difference, whether to arrive at Richardson extrapolation by directly omitting $\mathbf{C}'_0\lambda^{q_0}, \mathbf{C}'_1\lambda^{q_1}, \ldots \mathbf{C}'_{J-1}\lambda^{q_{J-1}}$ terms using an appropriate linear combination, or first computing a Richardson extrapolation of order J-m

$$0 < m < J \tag{25}$$

and omit J-m terms and then compute the Richardson extrapolation of order J, by computing the Richardson extrapolation of order m of the Richardson extrapolation of order J-m, i.e.

$${}^{R_J}\mathbf{U}^a = {}^{R_m}\left({}^{R_{J-m}}\mathbf{U}^a\right) \tag{26}$$

A direct extension of Eq. (26) is as stated below:

$${}^{R_{m_1+m_2+m_3+\ldots m_N}}\mathbf{U}^a = {}^{R_{m_1}}\left({}^{R_{m_2}}\left({}^{R_{m_3}}\left(\ddots\left({}^{R_{m_N}}\mathbf{U}^a\right)\right)\right)\right) \tag{27}$$

Equation (27) is essential in formation of a recent convergence purification method [17, 18].

4 NUMERICAL STUDY

Direct time integration is a versatile approach in analysis of structural systems against earthquake-induced ground motions [19, 20], and average acceleration [21], central difference [22], Wilson-θ [23-26], HHT [27], and C-H [28] methods, are of the broadly accepted time integration methods [20, 29-33]. In general, the responses obtained from time integration are approximations, and hence, the concepts of convergence, Richardson extrapolation, and that stated in Sections 2 and 3, are to be applicable; see also [3, 4, 34, 35]. Considering these, the structural system, introduced in Fig. 3(a) and Table 1, is analyzed against the ground motion introduced in Fig. 3(b), once by the average acceleration method and then again by the C-H ($\rho = 0.8$) method. Since, in time integration analyses, the integration step size implies the algorithmic parameter [10], i.e.

$$\lambda = \Delta t \tag{28}$$

the results of Richardson extrapolation of second order is first computed directly, in view of the relation below [4]:

$${}^{R_2}\mathbf{U}^a = \frac{r^{q_0+q_1}\mathbf{U}_3^a - \left(r^{q_0}+r^{q_1}\right)\mathbf{U}_2^a + \mathbf{U}_1^a}{r^{q_0+q_1} - \left(r^{q_0}+r^{q_1}\right)+1} \tag{29}$$

where,

$$\begin{gathered} \frac{\lambda_1}{\lambda_2} = \frac{\lambda_2}{\lambda_3} = r = \text{const.} = 2 > 1 \\ \mathbf{U}_3^a = \mathbf{U}^a\left(\lambda = \lambda_3\right) \quad , \quad \mathbf{U}_2^a = \mathbf{U}^a\left(\lambda = \lambda_2\right) \quad , \quad \mathbf{U}_1^a = \mathbf{U}^a\left(\lambda = \lambda_1\right) \end{gathered} \tag{30}$$

and the values of q_0 and q_1 for the two integration methods are as noted below [36]:

$$\begin{aligned} &\text{Average acceleration}: && q_0 = 2 \quad , \quad q_1 = 4 \\ &\text{CH}\,(\rho = 0.8): && q_0 = 2 \quad , \quad q_1 = 3 \end{aligned} \tag{31}$$

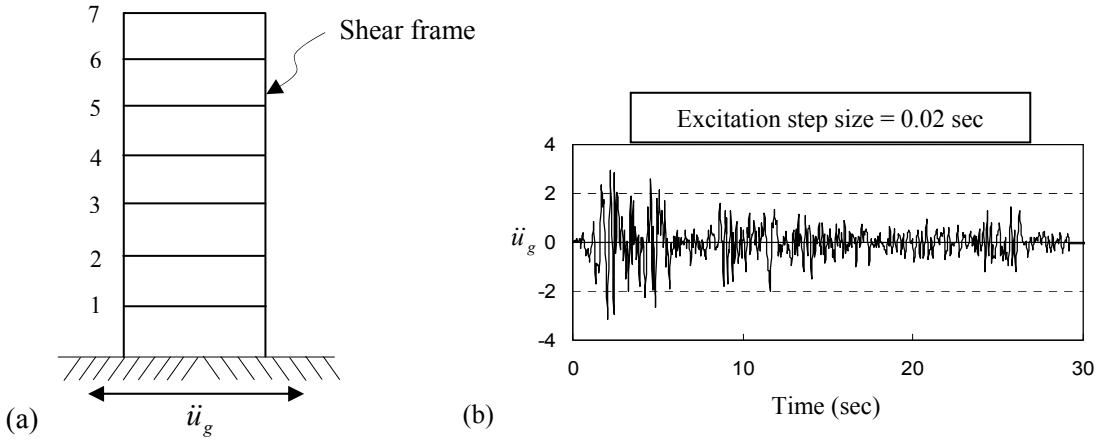


Figure 3: The structural system in the numerical study: (a) the structure, (b) the ground motion.

Floor	1	2	3	4	5	6	7
Mass $\times 10^{-3}$ (Kg)	2068	2064	2060	2056	2052	2048	2044
Stiffness$\times 10^{-6}$ (N/m)	840	820	700	680	660	640	620

Table 1: Main properties of the structure in Fig. 3(a).

and then by implementing the first order Richardson extrapolation (see Eq. (16) and [4]) to the resultant of first order Richardson extrapolation. The results together with the convergence trends are reported in Tables 2 and 3 and Figs. 4 and 5, where, the response under consideration is as noted below

$$\mathbf{U}^{a}=\begin{Bmatrix}\text{Maximum Top Displacement}\\ \text{Final Base Shear}\end{Bmatrix} \tag{32}$$

addressed as the top and bottom values in each cell of Tables 2 and 3. By comparing the last two columns in each table, it becomes apparent that Eq. (26) holds for both of the two time integration computations. This is a simple clear evidence for the validity of Eqs. (26) and (27). Similarly, and besides the theoretical discussion presented in Section 2, Figs. 4 and 5 imply the validity of the addressed accuracy basis and indeed the equivalence between the two apbases (accuracy-based and order of accuracy-based) leading to Richardson extrapolation.

Δt	$\mathbf{U}_1^a$	$\mathbf{U}_2^a$	$\mathbf{U}_3^a$	${}^{R_1}\mathbf{U}^a$	${}^{R_1}\left({}^{R_1}\mathbf{U}^a\right)$	${}^{R_2}\mathbf{U}^a$
0.02	-	-	0.26570572	-	-	-
	-	-	-6633587.8	-	-	-
0.01	-	0.26570572	0.27267107	0.27499285	-	-
	-	-6633587.8	-10209.648	2197583.07	-	-
0.005	0.26570572	0.27267107	0.27284983	0.27290942	0.27277053	0.27277053
	-6633587.8	-10209.648	848092.950	1134193.82	1063301.20	1063301.20
0.0025	0.27267107	0.27284983	0.27298766	0.27303360	0.27304188	0.27304188
	-10209.648	848092.950	500907.284	385178.729	335244.390	335244.390
0.00125	0.27284983	0.27298766	0.27303644	0.27305270	0.27305398	0.27305398
	848092.950	500907.284	-929445.92	-1406230.3	-1525657.6	-1525657.6
0.000625	0.27298766	0.27303644	0.27304956	0.27305393	0.27305401	0.27305401
	500907.284	-929445.92	-1276705.6	-1392458.8	-1391540.7	-1391540.7
0.0003125	0.27303644	0.27304956	0.27305302	0.27305417	0.27305419	0.27305419
	-929445.92	-1276705.6	-1360811.4	-1388846.7	-1388605.9	-1388605.9

Table 2: Comparison of the two approaches in computation of second order Richardson extrapolation in analysis of the system in Fig. (3) by the average acceleration method.

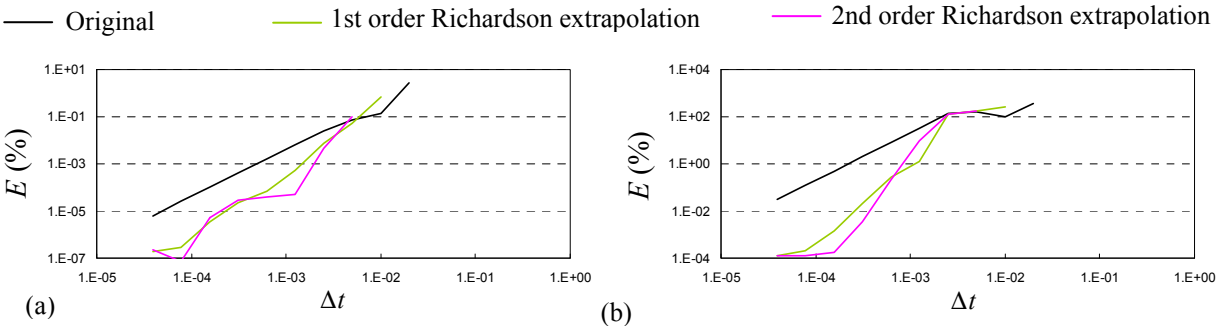


Figure 4: Convergence plots corresponding to Table 2: (a) Maximum top displacement, (b) Final base shear.

Δt	$\mathbf{U}_1^a$	$\mathbf{U}_2^a$	$\mathbf{U}_3^a$	${}^{R_1}\mathbf{U}^a$	${}^{R_1}\left({}^{R_1}\mathbf{U}^a\right)$	${}^{R_2}\mathbf{U}^a$
0.02	-	-	0.26515318	-	-	-
	-	-	-6602365.4	-	-	-
0.01	-	0.26515318	0.27265911	0.27516110	-	-
	-	-6602365.4	-1307192.0	-1307191.9	-	-
0.005	0.26515318	0.27265911	0.27284168	0.27290253	0.27257988	0.27257988
	-6602365.4	-1307192.0	413986.707	987712.942	1315556.51	1315556.51
0.0025	0.27265911	0.27284168	0.27298421	0.27303173	0.27305018	0.27305018
	-1307192.0	413986.707	600791.175	663059.330	616680.243	616680.243
0.00125	0.27284168	0.27298421	0.27303546	0.27305254	0.27305551	0.27305551
	413986.707	600791.175	-903083.30	-1404374.8	-1699722.5	-1699722.5
0.000625	0.27298421	0.27303546	0.27304931	0.27305393	0.27305412	0.27305412
	600791.175	-903083.30	-1270409.8	-1392852.0	-1391205.9	-1391205.9
0.0003125	0.27303546	0.27304931	0.27305296	0.27305418	0.27305421	0.27305421
	-903083.30	-1270409.8	-1359262.9	423470.293	682944.905	682944.905

Table 3: Comparison of the two approaches in computation of second order Richardson extrapolation in analysis of the system in Fig. (3) by the C-H ($\rho = 0.8$) method.

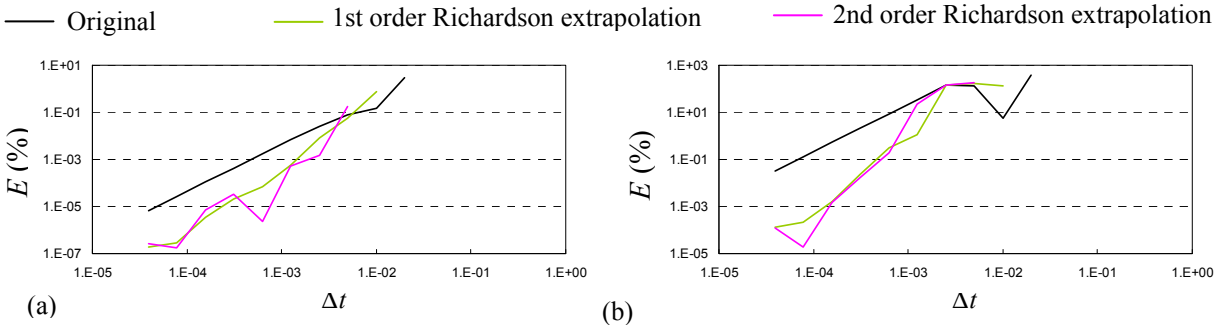


Figure 5: Convergence plots corresponding to Table 3: (a) Maximum top displacement, (b) Final base shear.

5 CONCLUSION

- Richardson extrapolation can be considered as the process of elimination of terms from the Taylor series expansion of the computation with respect to the algorithmic parameter, starting from the smallest exponents of the algorithmic parameter in the Taylor series. From this point of view, further accuracy in the proper convergence region as well as higher order of accuracy can be considered as the basis of Richardson extrapolation.
- Richardson extrapolation of order *J*, where *J* is a positive integer larger than one, can be obtained either directly or by first computing the Richardson extrapolation of a positive integer less than *J*, i.e. *m*, and then computing the Richardson extrapolation of order $J-m$ of the resultant. This can be extended; the number of the stages of implementation of Richardson extrapolation can be as large as *J*, at which case, in view of Eq. (26), Richardson extrapolation will be implemented sequentially for *J* times.

REFERENCES

[1] L.F. Richardson, J.A. Gaunt, The deferred approach to the limit. *Philosophical Transactions of the Royal Society of London*, **226**, 299–361, 1927.

[2] L.F. Richardson. The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society of London*, **210**, 307–357, 1910.

[3] C. Brezenski, M. Redivo Zagalia, *Extrapolation methods*: *theory and practice*. North-Holland, 1991.

[4] A. Soroushian, P. Wriggers, J. Farjoodi, Asymptotic upper bounds for the errors of Richardson extrapolation with practical application in approximate computations. *International Journal for Numerical Methods in Engineering*, **80**, 565-595, 2009.

[5] G. Birkhoff, G.C. Rota, *Ordinary differential equations*. John Wiley & Sons, 1978.

[6] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical recipes in Fortran The Art of Scientific Computing, 4th Edition*. Cambridge University Press, 1992.

[7] P. Henrici, *Discrete variable methods in ordinary differential equations*, Prentice-Hall, 1962.

[8] J.C. Strikwerda, *Finite difference schemes and partial differential equations*. Wadsworth & Books/Cole, 1989.

[9] C.W. Gear, *Numerical initial value problems in ordinary differential equations*. Prentice-Hall, 1971.

[10] A. Soroushian, A technique for time integration with steps larger than the excitation steps. *Communications in Numerical Methods in Engineering,* **24**, 2087-2111, 2008.

[11] A. Ralston, P. Rabinowitz, *First course in numerical analysis*. McGraw-Hill, 1978.

[12] W.L. Wood, *Practical time stepping schemes*. Clarendon, 1990.

[13] S.N. Penry, W.L. Wood, Comparison of some single-step methods for the numerical solution of the structural dynamic equation. *International Journal for Numerical Methods in Engineering*, **21**, 1941-1955, 1985.

[14] A. Soroushian, Proper convergence a concept new in science and important in engineering. D. Tsahalis ed. *4th International Conference from Scientific Computing to Computational Engineering (4th IC-SCCE)*, Athens, Greece, July 7-10, 2010.

[15] A. Soroushian, Convergence with non-integer rates. D. Tsahalis ed. *4th International Conference from Scientific Computing to Computational Engineering (4th IC-SCCE)*, Athens, Greece, July 7-10, 2010.

[16] G.H. Golub, C.F. Van Loan, *Matrix Computations*, *Third Edition*. John Hopkins University Press, 1996.

[17] A. Soroushian, Purification of convergence an approach towards reliable error evaluation. E. Onate, X. Oliver, A. Huerta eds. *11th World Congress on Computational Mechanics (WCCM XI)*, Barcelona, Spain, July 20-25, 2014.

[18] A. Soroushian, Purification of convergence: from definition to consequences. (to be submitted)

[19] J. Chung, G.M. Hulbert, A family of single-step Houbolt time integration algorithms for structural dynamics. *Computer Methods in Applied Mechanics and Engineering*, **118**, 1-11, 1994.

[20] A.K. Chopra, *Dynamics of structures: theory and application to earthquake engineering*. Prentice-Hall, 1995.

[21] M.N. Newmark, A method of computation for structural dynamics. *Journal of Engineering Mechanics (ASCE)*, **85**, 67-94, 1959.

[22] R.W. Clough, Numerical integration of equations of motion. *Lectures on Finite Element Methods in Continuum Mechanics*, University of Alabama, pp. 525-533, 1973.

[23] E.L. Wilson, *A computer program for the dynamic stress analysis of underground structures, Report No. 68-1*. University of California, Berkeley, 1968.

[24] K.J. Bathe, E.L. Wilson, Stability and accuracy analysis of direct integration methods. *Earthquake Engineering and Structural Dynamics*, **1**, 283-291, 1973.

[25] E.L. Wilson, I. Farhoomand, K.J. Bathe, Non-linear dynamic analysis of complex structures. *Earthquake Engineering and Structural Dynamics*, **1**, 241-252, 1973.

[26] A. Soroushian, J. Farjoodi, K. Bargi, M. Rajabi, A. Saaed, M. Arghavani, M.M. Sharifpour, Two versions of the Wilson-θ time integration method. J. Náprstek, J. Horáček, M. Okrouhlík, B. Marvalová, F. Verhulst, J.T. Sawicki eds. *10th Interantioanl Conferecne on Vibration Problems*. *(ICoVP 2011)*, Prague, Czech Republic, September 5-8, 2011.

[27] H.M. Hilber, T.J.R. Hughes, R.L. Taylor, Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering and Structural Dynamics*, **5**, 283-292, 1977.

[28] J. Chung, G.M. Hulbert, A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method. *Journal of Applied Mechanics (ASME)*, **60**, 371-375, 1993.

[29] T.J.R. Hughes, *The finite element method: linear static and dynamic finite element analysis*. Prentice-Hall, 1987.

[30] K.J. Bathe, *Finite element procedures*. Prentice-Hall, 1996.

[31] R.W. Clough, J. Penzien, *Dynamics of structures*. McGraw-Hill, 1993.

[32] J.L. Humar, *Dynamics of structures*. Prentice-Hall, 1990.

[33] R.R. Craig, Structural dynamics: an introduction to computer methods. John Wiley & Sons, 1981.

[34] M.N. Bismarck-Nasr, A.M. De Oliveira, On enhancement of accuracy in direct time integration dynamic response problems. *Earthquake Engineering and Structural Dynamics*, **20**, 699-703, 1991.

[35] Belytschko and Hughes T. Belytschko, T.J.R. Hughes, Computational methods for transient analysis, Elsevier, 1983.

[36] A. Soroushian, Effect of Richardson extrapolation on the convergence of some time integration methods. D. Tsahalis ed. *6th International Conference from Scientific Computing to Computational Engineering (6th IC-SCCE)*, Athens, Greece, July 9-12, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON THE FREQUENCY CONTENT OF ERRORS ORIGINATED IN A TIME INTEGRATION COMPUTATIONAL COST REDUCTION TECHNIQUE

A. Soroushian[1], M. Hosseini[1], and S.M.H. Khalkhali[2]

[1] International Institute of Earthquake Engineering and Seismology (IIEES)
S. Lavasani (Farmaniyeh), North Dibaji, West Arghavan, No. 21, Tehran 19537, Iran
e-mail: {a.soroushian,Hosseini}@iiees.ac.ir, aramsoro@yahoo.com

[2] School of Engineering, Science and Research Branch of Islamic Azad University
Hessarak, S. Satari, Tehran 14778, Iran
e-mail: khalkhali_eng@yahoo.com

Keywords: Time Integration, Digitized Excitation, Computational Cost, Accuracy, Frequency Content.

Abstract. *The true behavior of structural systems, mainly essential in seismic analyses, is nonlinear and dynamic. In analysis of nonlinear dynamic behaviors of structural systems, a versatile tool is direct time integration. In spite of the versatility, the responses obtained from time integration are approximations and the computational costs are considerable. For analyses subjected to excitation available as digitized records, a technique to reduce the computational cost with negligible sacrifice of accuracy is proposed in 2008. The technique is successfully implemented in seismic analysis of frames, buildings structures, bridges, reservoirs, power stations, silos, etc., and undergone theoretical studies from different points of view. This is a complementary study that, with the aim of reducing the small errors because of the technique and reducing the computational costs even more, numerically investigates the frequency content of the errors originated in the computational cost reduction technique. As the conclusion, more errors are expectable in lower modes and/or modes with more contribution in the response. Further extensive study is essential.*

1 INTRODUCTION

Time integration is the most versatile approach to analyze the transient behaviors of semi-discretized structural systems. Nevertheless, the computational efforts are considerable, and the responses are approximations. In a review on the time integration process, after discretization of the mathematical models in space, by methods such as finite elements, we arrive at the semi-discretized models, typically stated below:

$$\begin{aligned} &\mathbf{M}\ddot{\mathbf{u}}(t)+\mathbf{f}_{\text{int}}(t)=\mathbf{f}(t), \qquad 0<t\le t_{end} \\ &\text{Initial conditions:} \quad \left|\begin{array}{l}\mathbf{u}(t=0)=\mathbf{u}_0\\ \dot{\mathbf{u}}(t=0)=\dot{\mathbf{u}}_0\\ \mathbf{f}_{\text{int}}(t=0)=\mathbf{f}_{\text{int}_0}\end{array}\right. \\ &\text{Additional constraints: } \mathbf{Q} \end{aligned} \tag{1}$$

where, t and t_{end} imply the time and the duration of the dynamic behavior, $\mathbf{M}$ is the mass matrix, $\mathbf{f}_{\text{int}}(t)$ and $\mathbf{f}(t)$ stand for the vectors of internal force and excitation, $\mathbf{u}$(t), $\dot{\mathbf{u}}(t)$, and $\ddot{\mathbf{u}}(t)$ denote the vectors of displacement, velocity, and acceleration; $\mathbf{u}_0$, $\dot{\mathbf{u}}_0$, and $\mathbf{f}_{\text{int}_0}$, define the initial status of the model (regarding the essentiality of considering $\mathbf{f}_{\text{int}_0}$ in Eqs. (1), see [1, 2]), and $\mathbf{Q}$ implies nonlinearity restrictions, e.g. see [3, 4]. Time integration determines the status at discrete time instants $\Delta t, 2\Delta t, 3\Delta t$ …. (Δt stands for the integration step size), according to approximate formulations, introducing the integration method [5-8]. Because of the approximation, the step size Δt can not be large; however, because of the step-by-step nature, and the resulting high computational cost, the step size can also not be small. Considering this, and the unconditional stability of conventional time integration methods (see [6-9]), the integration step size recommended is basically obtained from [5, 7, 9-11]:

$$\Delta t=\begin{cases}\dfrac{T}{10} & \text{linear problem}\\ \dfrac{T}{100} & \text{nonlinear problem}\\ \dfrac{T}{1000} & \text{nonlinear problem involved}\end{cases} \tag{2}$$

where, T is the smallest period with considerable contribution in the response. In some practical cases, e.g. time integration analysis against earthquake records, where,

$$\mathbf{f}(t)=-\mathbf{M}\,\Gamma\,\ddot{\mathbf{u}}_g(t) \tag{3}$$

($\mathbf{M}$ represents the mass matrix, $\ddot{\mathbf{u}}_g(t)$ denotes the vector of ground motion at some different degrees of freedom (unless in multi-support excitations, $\ddot{\mathbf{u}}_g$ reduces to a function of time) and Γ is a matrix implying the effect of ground motion at different degree of freedom [12]), $\mathbf{f}(t)$ is available as a vector of digitized records, and Eq. (2) is being replaced with

$$\Delta t=\min\left({}_f\Delta t,\frac{T}{\chi}\right) \tag{4}$$

where, $_f\Delta t$ implies the step size, by which, $\mathbf{f}(t)$ is digitized, and in view of Eq. (2);

$$\chi = \begin{cases} 10 & \text{linear problem} \\ 100 & \text{nonlinear problem} \\ 1000 & \text{nonlinear problem involved in impact} \end{cases} \tag{5}$$

When, $_f\Delta t$ governs Eq. (4), some additional computational cost, e.g. about

$$\frac{T-\chi_f\Delta t}{\chi_f\Delta t}>0 \tag{6}$$

for linear problems, is being dictated to the analyses, merely, in order to take into account the total information of $\mathbf{f}(t)$; see also [13].

In 2008, a technique is proposed for reducing the above-mentioned additional computational cost, while taking into account the total earthquake record, and not sacrificing considerable accuracy (compared to analyses with steps obtained from Eq. (4)) [13-16]. The past experiences were successful; see Table 1. Nevertheless, with the rapid scientific improvements, the value of χ, as well as the value of $_f\Delta t$ in Eqs. (4) and (5) is likely to be decreased;

System(s) analyzed	The cost reduction (%)	Source
SDOF system	75	[14]
2-DOF nonlinear system	49.27	[14]
Eight storey shear frame	80	[17]
Thirty-storey building	50	[18]
3-component earthquakes	66.7	[19]
Silo	77.65	[20, 21]
Water tank	66.7	[16, 22]
Building in pounding	12.7	[23]
Bridge with linear and nonlinear behaviors	45-80	[24-26]
Power stations	> 50	[27]
Regular residential buildings	50-87	[28, 29]
Bridges with pre-stressed elements, subjected to multi-support excitation and nonlinearities	30-70	[30]
Residential building with irregularities in height	50-80	[31]
Space Structures	>50	[32]
A telecommunication tower	>50	[33]
A cooling tower	>50	[34]

Table 1. Experiences on the computational cost reduction technique proposed in [14].

see also [35]. Accordingly, in view of Eq. (6), more enlargement of digitization step size would be desirable, and study towards improvement of the technique proposed in 2008 [14], especially for decreasing the even small inaccuracy induced by the technique, can cause considerable practical achievements. Considering this and in view of a recent research reported in [16], this paper presents an attempt to point out whether there exists any special distribution of the errors in different oscillatory modes. In the next section a brief theoretical discussion is presented. Numerical evidences, regarding the discussion in Section 2, are presented in Section 3, and the paper is concluded in Section 4.

2 BRIEF THEORETICAL DISCUSSION

In a review on the formulation of the technique proposed in 2008 [14] (for more detailed formulation, see [13, 15]), the replacement of the excitation $\mathbf{f}(t)$ with the excitation $\widetilde{\mathbf{f}}(t)$, digitized at larger steps, is based on convergence and specifically on convergence of second order. In more detail, in view of the Eqs. (35), (38), and (39), in [14], also repeated below:

$$\begin{aligned} t_i = 0: &\quad \widetilde{\mathbf{f}}_i = \mathbf{f}(t_i) \\ 0 < t_i \le t_{end}: &\quad \widetilde{\mathbf{f}}_i = \frac{1}{2}\mathbf{f}(t_i) + \frac{1}{4n'}\sum_{k=1}^{n'}\left[\mathbf{f}(t_{i+k/l}) + \mathbf{f}(t_{i-k/l})\right] \\ t_i = t_{end}: &\quad \widetilde{\mathbf{f}}_i = \mathbf{f}(t_i) \end{aligned} \tag{7}$$

$$\begin{aligned} t = \Delta t: &\quad n' = n-1 \\ \Delta t < t \le t_{end} - \Delta t: &\quad n' = \begin{cases} \dfrac{n}{2} & n = 2j,\ j \in \mathrm{Z}^+ \\ \dfrac{n-1}{2} & n = 2j+1,\ j \in \mathrm{Z}^+ \end{cases} \\ t = t_{end} - \Delta t: &\quad n' = n-1 \end{aligned} \tag{8}$$

$$\begin{aligned} \Delta t = n_f \Delta t \le \min(h, \frac{T}{10}) &\quad \text{for linear system} \\ \Delta t = n_f \Delta t \le \min(h, \frac{T}{100}) &\quad \text{for nonlinear system} \\ \Delta t \le t_{end} & \end{aligned} \tag{9}$$

and the derivation of this formulation, there is no consideration about the dynamic nature of the phenomenon. Accordingly, it sounds reasonable to simply expect more contribution in the additional errors because of the technique, from the oscillatory modes contributing more in the dynamic behaviors, e.g. the first modes for many behaviors. This is also in consistence with Eq. (4), and also the notions of amplitude decay and period elongation [5-9], according to which, the higher modes, with lower values of T, are being time integrated more inaccurately and hence the additional errors because of the technique proposed in 2008 [14] are comparatively not considerable at the higher modes. Since, this claim is in agreement with a recent observations reported in [16], as a step towards profound study and clearing the distribution of the additional errors caused by the technique in different natural modes, some numerical examples are presented, in Section 3.

3 NUMERICAL EXAMPLES

3.1 2-D Frame

The 2D frame introduced in Fig. 1(a) and Table 2 is subjected to the ground acceleration displayed in Fig. 1(b) and studied considering linear behavior. The histories of top displacement and base shear reported in Fig. 2, together with Fig. 1(b), lead to

$$n = 2 \quad (\equiv \quad \Delta t = 2_f \Delta t) \tag{10}$$

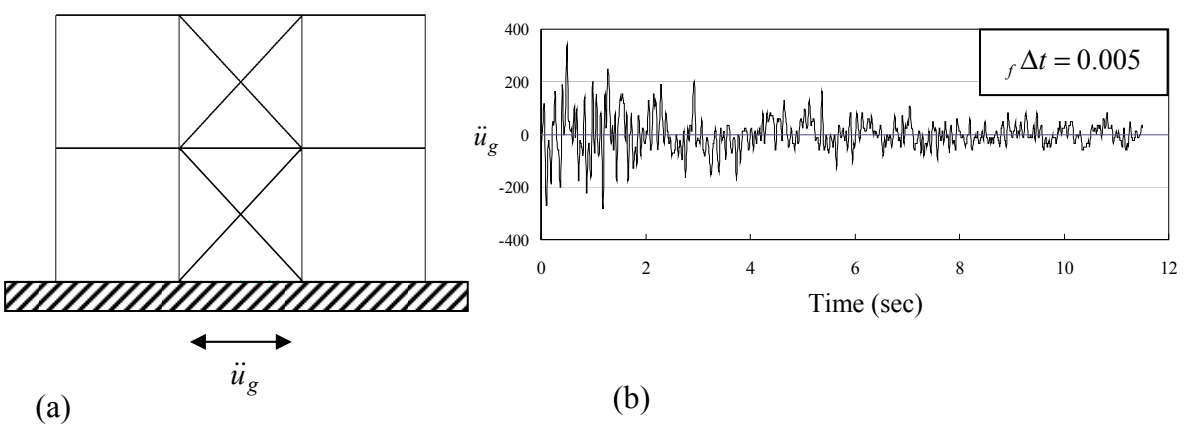


Figure 1: Structural system in the first example: (a) 2D frame (b) Earthquake record.

Members	**Profiles**
Columns	BOX 8x18x1.25
Beams	IPE 360
Braces	BOX 8x8x0.6

Table 2: Profiles of the structural members in Fig. 1(a).

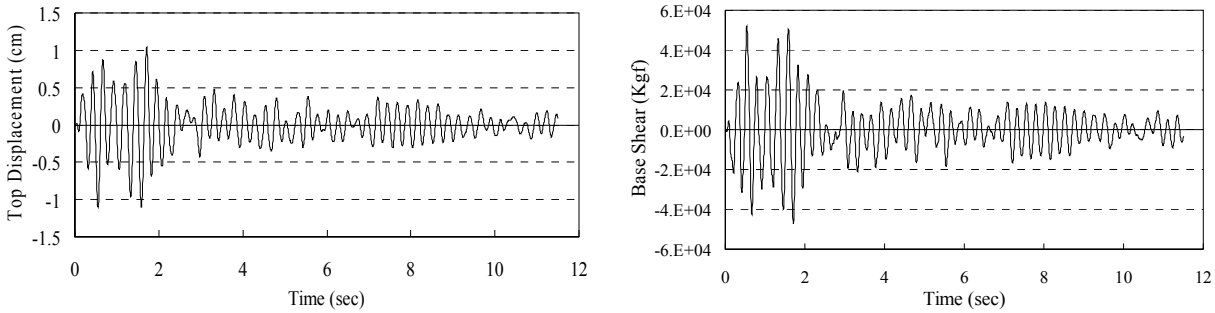


Figure 2: Responses for the structural system in Fig. 1.

in implementation of the technique proposed in [14]. In Figs. 3 and 4, the responses obtained from time integration analyses with the average acceleration method [5-7, 12, 36] and steps satisfying Eq. (4) are compared, with the responses obtained, when changing the step sizes to step sizes obtained from Eq. (4) after eliminating $_f\Delta t$, i.e.

$$\Delta t = n_f \Delta t \tag{11}$$

where, n is to be as stated in Eq. (10). As apparent in Fig. 3, the technique has successfully halved the computational cost, in the price of negligible loss of accuracy. Meanwhile, though hardly recognizable in Fig. 4, the errors because of the technique are more in lower oscillatory modes (higher values of T); in this example (and for many systems), these are the modes with more contribution in the response.

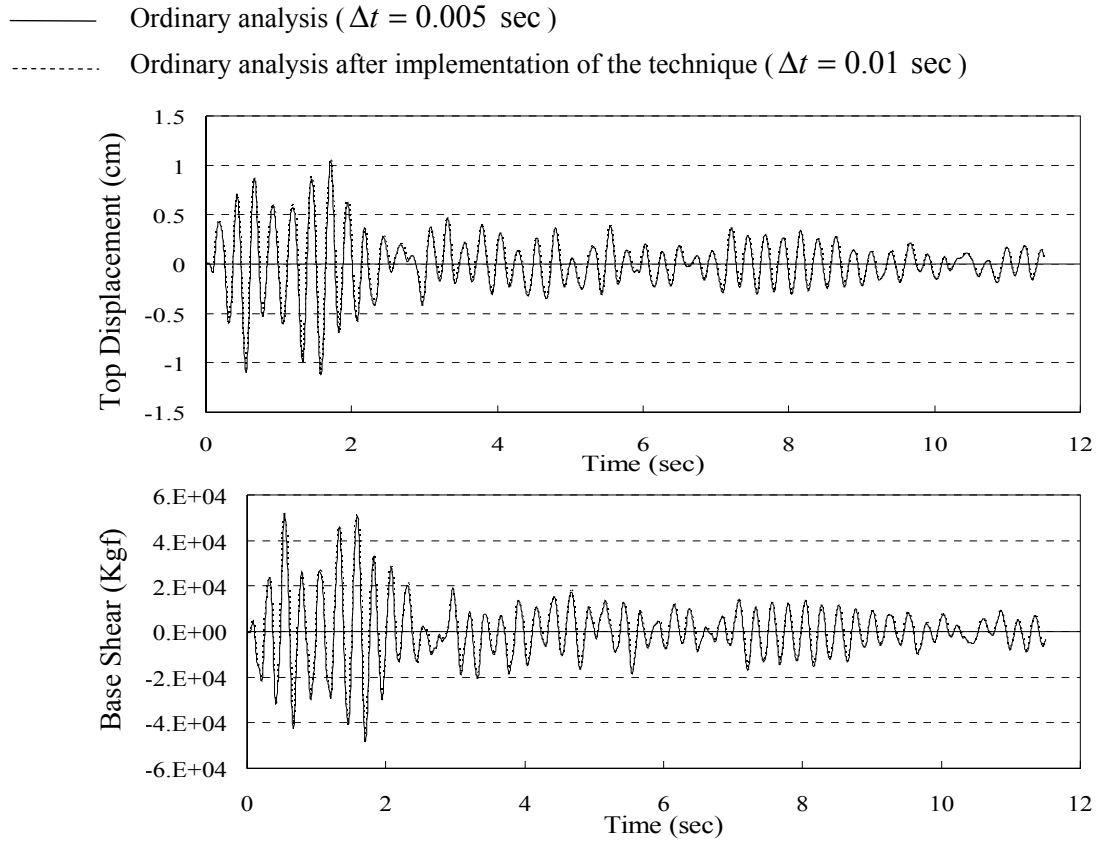


Figure 3: Responses computed for the structural system in Fig. 1 by the average acceleration method.

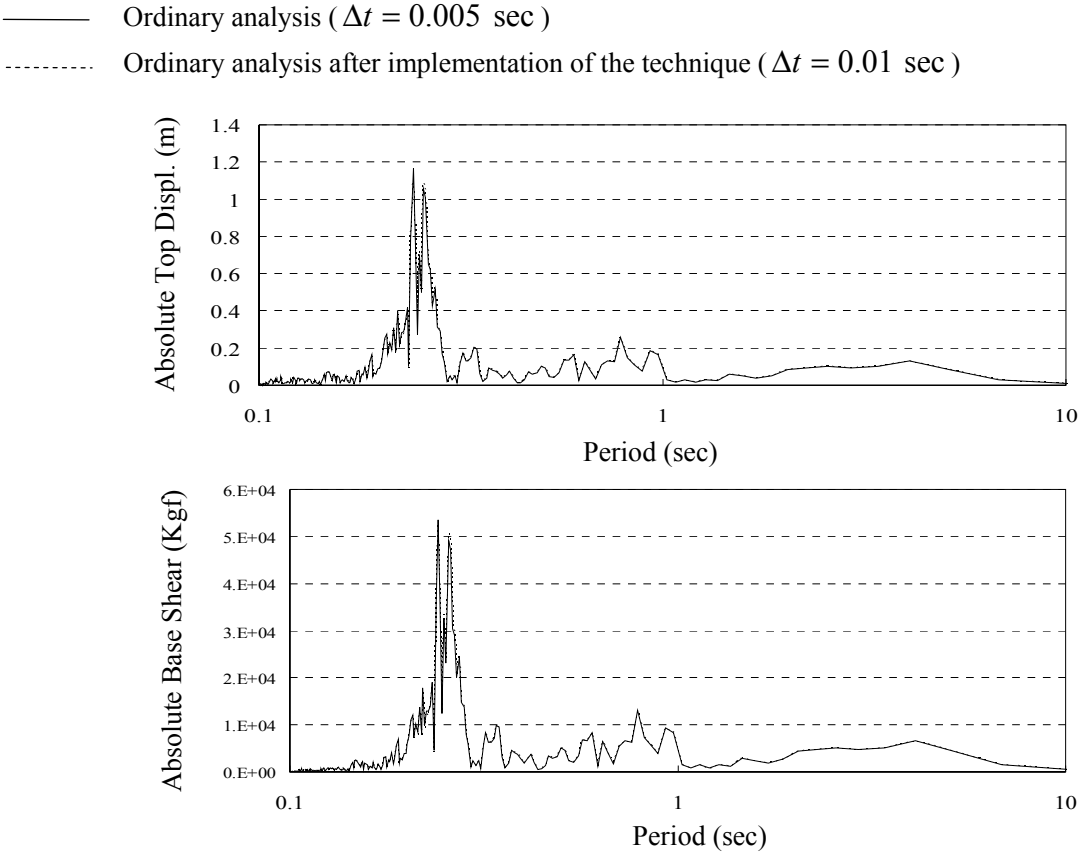


Figure 4: Frequency content of the response displayed in Fig. 3.

3.2 3-D Frame

The study reported in Section 3.1 is repeated for the 3-D frame introduced in Fig. 5(a) and Table 2, against the earthquake records in Fig. 5(b) applied in the main directions ($_f \Delta t = 0.005$ sec). The structural members are briefly introduced in Table 3; the structural system is dual (moment resisting frame plus bracings) with two-sided slabs at floors; the floors height are 3 meters and the vertical loads are distributed uniformly on the floors, while the earthquake records are applied in the main directions, and, in separate studies, the analysis methods are the HHT ($\gamma = 0.6, \beta = 0.3025, \alpha = -0.3$) and the average acceleration [36, 37]. The almost exact responses are as displayed in Fig. 6. In view of Figs. 5(b) and 6 (also see [24]),

$$n = 5 \tag{12}$$

is an appropriate selection in implementation of the technique proposed in 2008 [14]. The resulting responses are as reported in Figs. 7 and 8, respectively, in time and frequency domains, once again, implying the good performance of the technique and the more errors of the technique in lower modes.

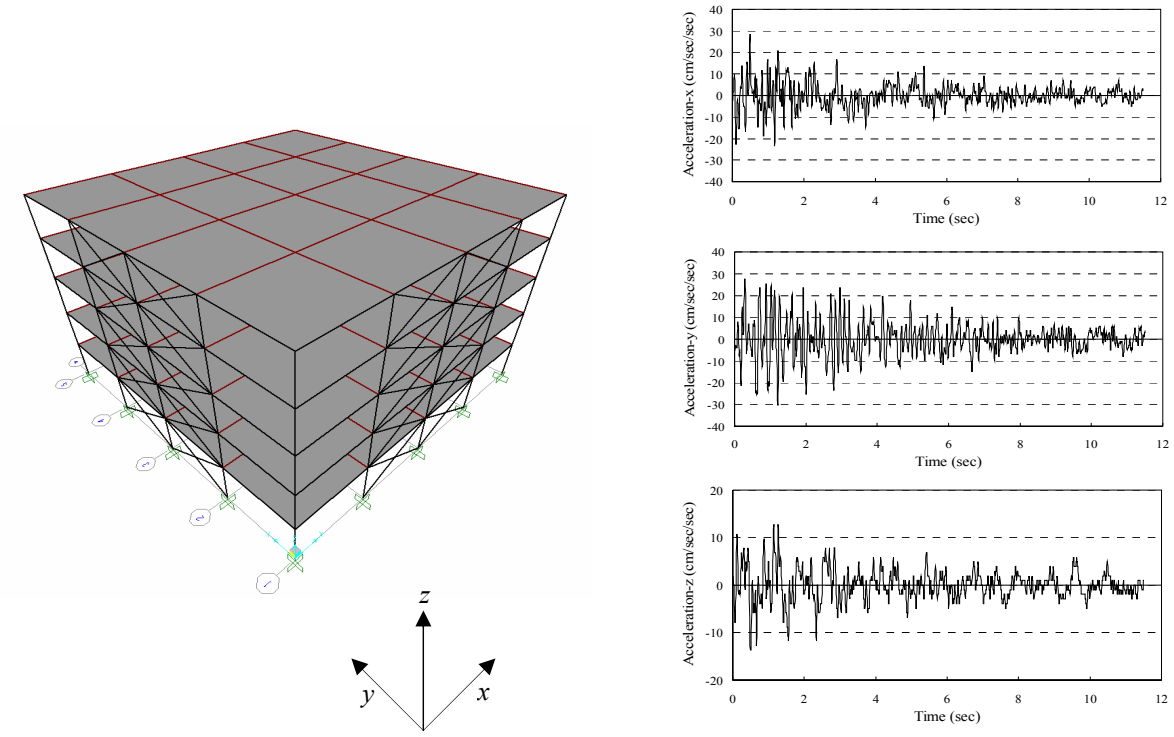


Figure 5: Structural system in the second example: (a) 3D frame, (b) Ground motion.

Members	Profiles
Beams	IPE 270, IPE 300, IPE 360, IPE 400
Columns	BOX 20x20x1.25, BOX 22x22x1.25, BOX 22x22x1.6, BOX 24x24x1.75, BOX 26x26x2.22
Braces	BOX 12x12x1, BOX 14x14x1, BOX 9x9x1

Table 3: Profiles of the structural members in Fig. 5(a).

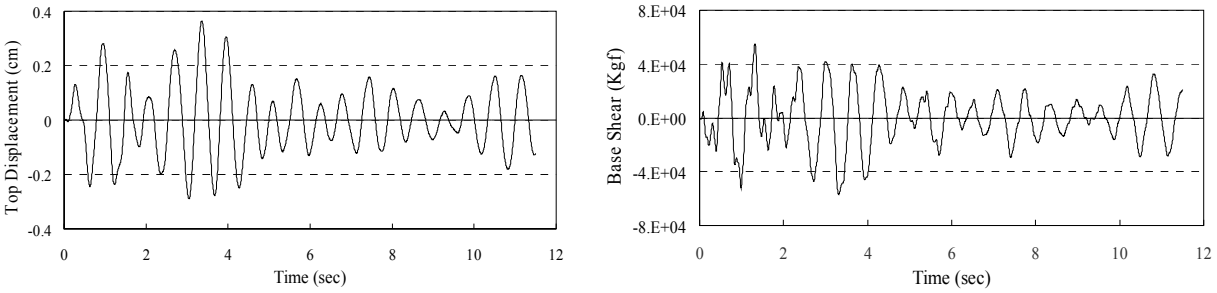


Figure 6: Top displacement and base shear histories for the second example.

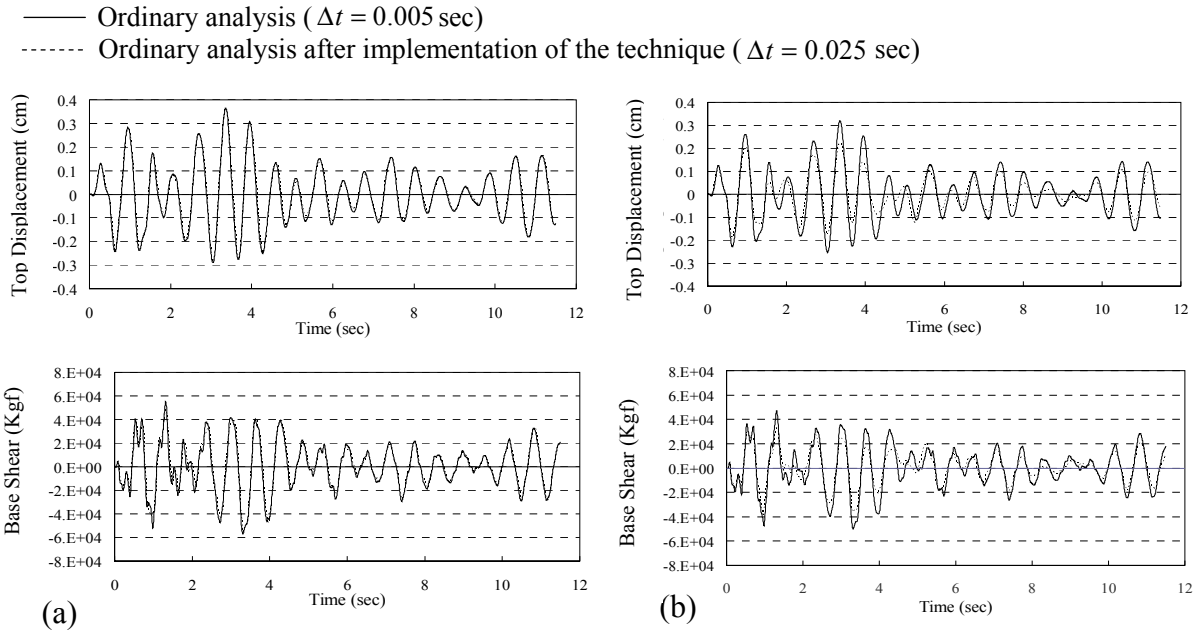


Figure 7: Time histories of the responses computed for the structural system introduced in Fig. 5 and Table 3 by: (a) Average acceleration, (b) HHT.

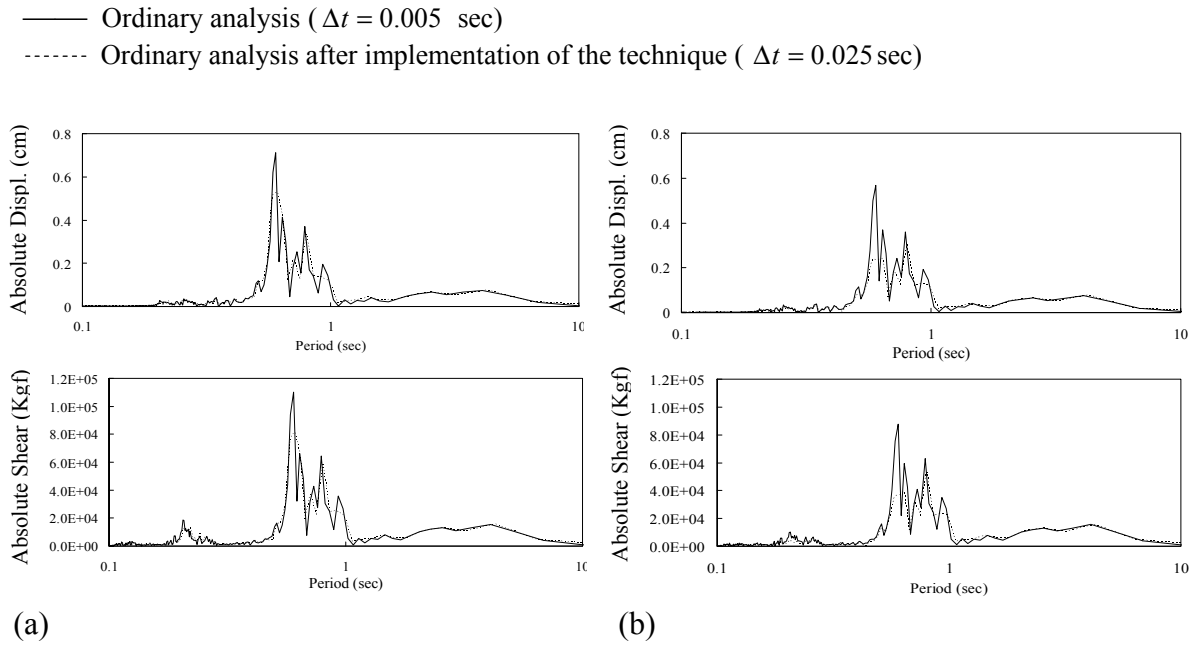


Figure 8: Frequency content of the response reported in Fig 7: (a) Average acceleration, (b) HHT.

The study is repeated taking into account the nonlinear behavior of the joints as well as the P-Δ effect. Merely the P-Δ actually affected the behavior and that only slightly. Accordingly, considering the difference between the linear and nonlinear cases in Eq. (2),

$$n = 1.25 \tag{13}$$

is an appropriate selection in the implementation of the technique, proposed in [14], in analysis of the system introduced in Fig. 5 and Table 3 (see also [38]). The results reported in Figs. 9 and 10, once again, are evidences for the claim in Section 2.

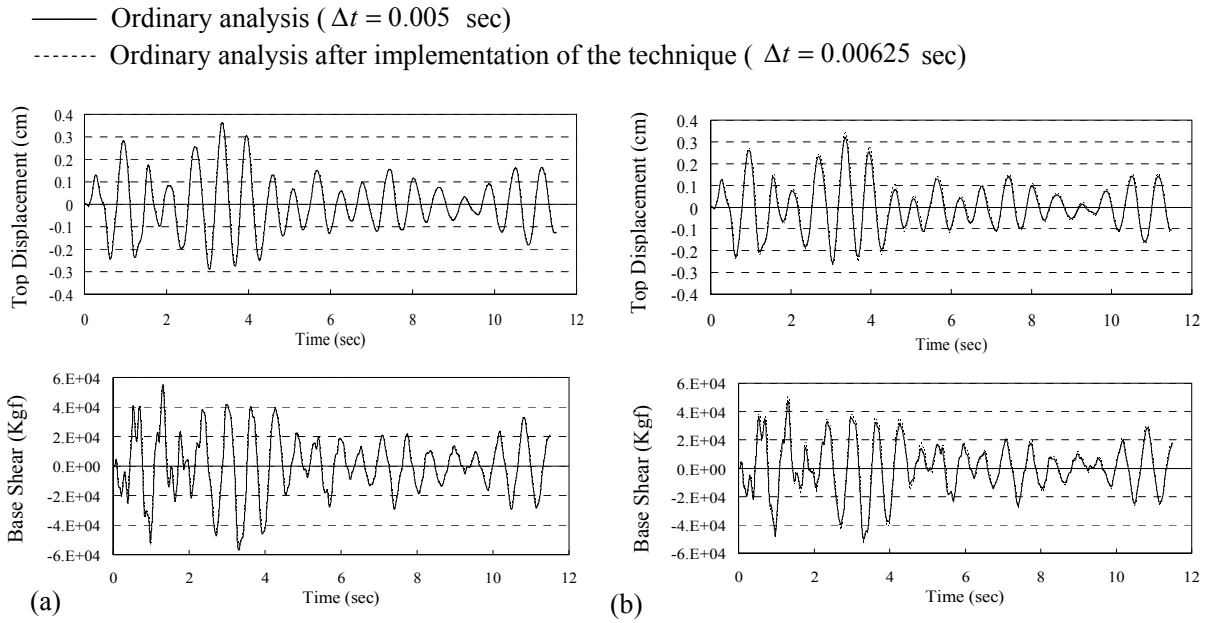


Figure 9: Time history of the response computed for the nonlinear case of the structural system introduced in Fig. 5 and Table 2 by: (a) Average acceleration, (b) HHT.

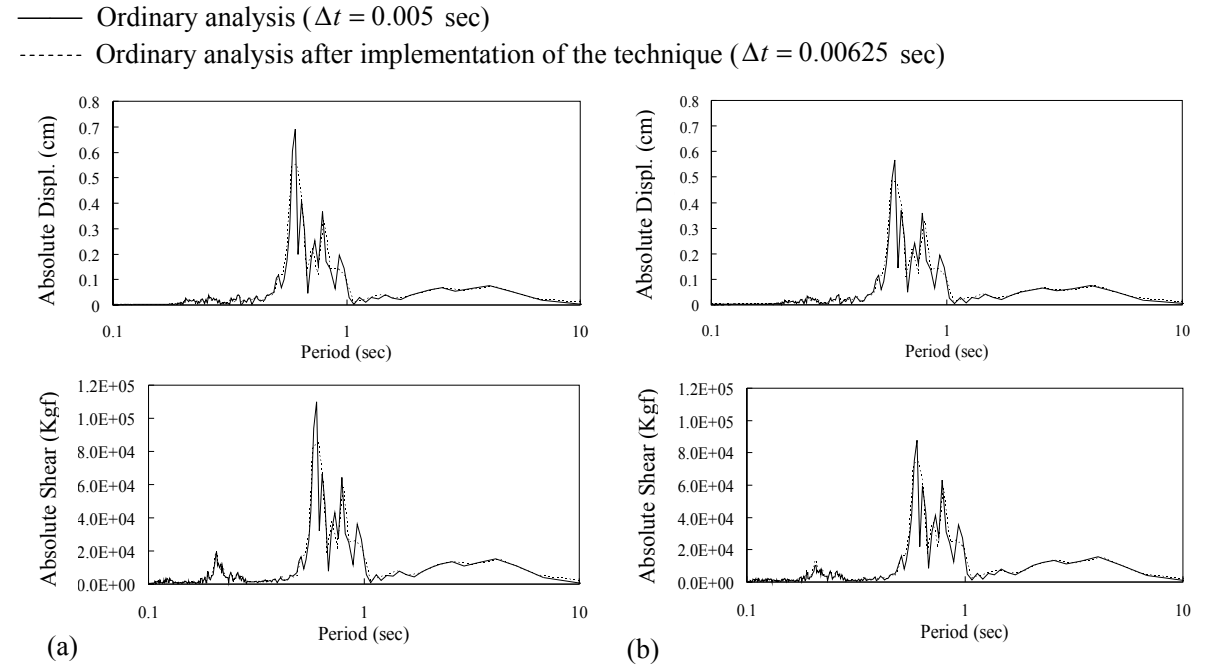


Figure 10: Frequency content of the response reported in Fig 9: (a) Average acceleration, (b) HHT.

4 CONCLUSIONS

The performance of a time integration computational cost reduction technique in frequency domain is studied via brief theoretical discussion and three examples. Accordingly, it sounds reasonable to expect more errors either in lower modes or in modes with high contribution in the response; this is not restricted to a specific time integration method and to linear behaviors. Towards more efficiency for the technique proposed in [14], further study is essential and highly recommended.

REFERENCES

[1] H. Gavin, *Structural dynamics*. Class Notes CE 283, Duke University, 2001.

[2] A. Soroushian, *New methods to maintain responses' convergence and control responses' errors in the analysis of nonlinear dynamic models of structural systems*. PhD Dissertation, University of Tehran, 2003. (in Persian)

[3] P. Wriggers, *Computational contact mechanics*. John Wiley & Sons, 2002.

[4] T.J.R. Hughes, K.S. Pister, R.L. Taylor, Implicit-explicit finite elements in nonlinear transient analysis. *Computer Methods in Applied Mechanics and Engineering,* **17/18**, 159-182, 1979.

[5] R.W. Clough, J. Penzien, *Dynamics of structures*. McGraw-Hill, 1993.

[6] J. Argyris, J.P. Mlejnek, *Dynamics of structures*. Elsevier, 1991.

[7] K.J. Bathe, *Finite element procedures*. Prentice-Hall, 1996.

[8] T. Belytschko, T.J.R. Hughes, *Computational methods for transient analysis*. Elsevier, 1983.

[9] T.J.R. Hughes, *The finite element method: linear static and dynamic finite element analysis*. Prentice-Hall, 1987

[10] J.F. McNamara, Solution schemes for problems of nonlinear structural dynamics. *Journal of Pressure Vessels (ASME)*, **96**, 149-155, 1974.

[11] NZS, *Structural design actions, Part 5: Earthquake actions-New Zealand*. New Zealand Standard, 2004.

[12] A.K. Chopra, *Dynamics of structures: theory and application to earthquake engineering*. Prentice-Hall, 1995.

[13] A. Soroushian, Integration step size and its adequate selection in seismic dynamic analysis of structural systems. *In* Computational methods in earthquake engineering, vol. 3. *Edited by* M. Papadrakakis and N. Lagaros and V. Plevris. Springer. (accepted for publication)

[14] A. Soroushian, A technique for time integration with steps larger than the excitation steps. *Communications in Numerical Methods in Engineering*, **24**, 2087-2111, 2008.

[15] A. Soroushian, *Direct time integration with step larger than the steps by which the excitations are digitized*, *Report 7510*. International Institute of Earthquake Engineering and Seismology (IIEES), 2011. (In Persian)

[16] A. Soroushian, P. Farshadmanesh, S. Azad, On the essentiality of techniques to enlarge analysis steps in time integration against digitized excitations. *International Journal for Earthquake Engineering and Seismology (JSEE)*, **17**, 2015. (accepted for publication)

[17] A. Soroushian, On the performance of a conventional accuracy controlling method applied to linear and nonlinear structural dynamics. M. Crocker ed. *17th International Congress on Sound & Vibration (ICSV17)*, Cairo, Egypt, July 18-22, 2010.

[18] A. Soroushian, A. Aziminejad, A more efficient seismic analysis of tall buildings by implementing a recently proposed technique. A.A. Tasnimi, A. Kalantari eds. *6th International Conference Seismology and Earthquake Engineering (SEE6)*, Tehran, Iran, May 16-18, 2011.

[19] O. Bahar, S. Ramezani, Faster time integration analysis for building structures subjected to 3-component earthquakes. M. Papadrakakis, M. Fragiadakis, V. Plevris eds. *3rd ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMDYN 2011)*, Corfu, Greece, May 25-28, 2011.

[20] F. Nateghi, M. Yakhchalian, On less computational costs for analysis of silos seismic behaviors by time integration. M. Papadrakakis, M. Fragiadakis, V. Plevris eds. *3rd ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2011)*, Corfu, Greece, May 25-28, 2011.

[21] F. Nateghi, M. Yakhchalian, (2011) An investigation into the effectiveness of a technique proposed for reducing computational cost of time integration in the analysis of silos seismic behaviors. K.K. Tamma eds. K.K. Tamma *11th US National Congress on Computational Mechanics (USNCCM 2011)*, Minneapolis, USA, July 25-29, 2011.

[22] A. Soroushian, A. Saaed, M. Arghavani, M. Rajabi, M.M. Sharifpour, Less computational costs in the analysis of reservoirs seismic behaviours by time integration. J. Náprstek, J. Horáček, M. Okrouhlík, B. Marvalová, F. Verhulst, J.T. Sawicki eds. *10th Interantioanl Conferecne on Vibration Problems*. *(ICoVP 2011)*, Prague, Czech Republic, September 5-8, 2011.

[23] A. Soroushian, On the performance of a recent technique for more efficient time integration in severe seismic conditions. C-K. Choi ed. *1st International Conference on Advances in Structural Engineering and Mechanics (ASEM'11)*, Seoul, South Korea, September 18-22, 2011.

[24] S. Azad, *A study on accelerating time history analysis of bridges*, M.Sc. Dissertation, International Institute of Earthquake Engineering and Seismology (IIEES), 2015. (In Persian)

[25] A. Soroushian, M. Hosseini, A. Vasseghi, On the performance of a technique for more efficient time integration when applied to bridge structures seismic analysis. M. Papadrakakis, M. Fragiadakis, V. Plevris eds. *3rd ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2011)*, Corfu, Greece, May 25-28, 2011.

[26] A. Soroushian, A. Vasseghi, M. Hosseini, On practical performance of a technique for more efficient dynamic analysis in view of real seismic analysis of a bridge structural system. *In* Computational methods in earthquake engineering, *Edited by* M. Papadrakakis and M. Fragiadakis and V. Plevris. Springer, 2013.

[27] M. Bastami, A technique for more efficient time integration applied to seismic analysis of power substation equipment. E. Onate, X. Oliver, A. Huerta eds. *11th World Congress on Computational Mechanics (WCCM XI)*, Barcelona, Spain, July 20-25, 2014.

[28] A. Sabzei, *On the performance of a recent technique for seismic analyses computational cost reduction when applied to buildings structural systems*. M.Sc. Dissertation, International Institute of Earthquake Engineering and Seismology (IIEES), 2013. (In Persian)

[29] A.Y. Reziakolaei, A.Sabzei, A. Soroushian, On the performance of a structural analysis cost reduction technique when applied to residential buildings. M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris eds. *4th ECCOMAS Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2013)*, Kos Island, Greece, June 12-14, 2013.

[30] S. Azad, A. Soroushian, S. Malek, On the performance of a computational cost reduction technique when applied to bridges multi-support excitations. M.K. Jafari, K. Amini-Hosseini eds. *7th International Conference on Seismology and Earthquake Engineering (SEE7)*, Tehran, Iran, May 18-21, 2015. (in Persian)

[31] A.A. Hadad, *Reducing computational costs in time integration analyses of buildings with irregularities in height because of mass*, M.Sc. Dissertation, International Institute of Earthquake Engineering and Seismology (IIEES), 2015. (In Persian)

[32] A. Garakaninezhad, R.K. Moghadas, On the performance of a technique to accelerate time integration when applied to space structures analyses. M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris eds. *5th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2015)*, Crete, Greece, May 25-27, 2015.

[33] A. Soroushian, S.R. Mirghaderi, Y. Zarabimanesh, A.J. Mehrnoush, A. Z. Zadeh, M.H. Ghondaghsaz, On reduction of computational costs in transient analysis of telecommunication towers. S-K. Yuan eds. *12th World Congress on Computational Mechanics (WCCM XII*), Seoul, South Korea, July 24-29, 2016. (accepted)

[34] A. Soroushian, A.J. Mehrnoush, Y. Zarabimanesh, M.H. Ghondaghsaz, A. Baiani, A.Zakizadeh, On the performance of computational cost reduction technique applied to cooling towers. M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris eds. *7th European Congress on Computational Methods in Applied Science and Engineering (ECCOMAS Congress 2016)*, Crete, Greece, July 5-10, 2016. (accepted)

[35] J. Havskov, G. Alguacil, *Instrumentation in earthquake seismology (modern approaches in geophysics)*. Springer, 2005.

[36] N.M. Newmark, A method of computation for structural dynamics. *Journal of Engineering Mechanics (ASCE)*, **85**, 67-94, 1959.

[37] H.M. Hilber, T.J.R. Hughes, R.L. Taylor, Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering and Structural Dynamics*, **5**, 283-292, 1977.

[38] A. Sabzei, A.Y. Reziakolaei, A. Soroushian, On more versatility for an integration step size enlargement technique in time integration analysis. M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris eds. *4th ECCOMAS Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPDYN 2013)*, Kos Island, Greece, June 12-14, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL SIMULATION OF TRANSONIC BUFFET AND ITS CONTROL USING TANGENTIAL JET BLOWING

Ksenia Abramova[1], **Kamil Khairullin**[1], **Alexander Ryzhov**[1], **Vitaly Soudakov**[1]

[1] Central Aerohydrodynamic Institute named after prof. N.E. Zhukovsky (TsAGI)
Russia, Zhukovsky, Zhukovsky str. 1
{aleksandr.ryzhov, vit_soudakov}@tsagi.ru

Keywords: buffet, URANS, flow control, tangential jet blowing, transonic flow.

Abstract. *The unsteady interaction between a shock wave and boundary layer over a transonic wing can lead to the generation of considerable separation zones and periodical shock wave motion along the airfoil chord. The phenomenon is called transonic buffet. In the present work the buffet phenomenon on a transonic airfoil is simulated numerically using second order finite-volume method. Unsteady Reynolds Averaged Navier-Stokes (URANS) equations with different turbulence models are integrated. Transonic airfoil P-184-15SR of TsAGI is used as a reference geometry with Reynolds number ~2.6×10^6. The buffet onset boundary is calculated for different free-stream Mach numbers. Active flow control method is proposed to alleviate the buffet onset. In this method, jet is blown tangentially near the shock wave on the upper airfoil surface. Parametric studies are performed. In all cases tangential blowing leads to significant lift growth and buffet alleviation.*

1 INTRODUCTION

Airfoils at transonic speeds are subjected to the buffet phenomenon, associated with unsteady shock-boundary layer interaction inducing self-sustained oscillations of separation zone under the shock foot on the upper surface of the wing. That leads to significant changes of the airfoil performance, i.e. lift, drag and moments. This phenomenon can further lead to structural vibrations called buffeting. Wing design standards impose margins between the buffeting onset and the cruise condition. As a consequence, a delay in buffeting onset could lead to improved aerodynamic performance characteristics that can result in the reduction of wing area and hence the friction part of drag. One of the ways of buffet alleviation is flow control.

The buffet is associated with the shock-boundary layer interaction (SBLI). Up to date a number of methods of SBLI control are developed. One of them is a passive method, in which a cavity covered with a perforated plate is placed under the shock foot [1]. Grooves and stream-wise slots are also well-known passive devices [2]. In all the methods the shock is weakened and wave drag is reduced. 2D bumps also lead to the wave drag reduction, but also result in too high penalty under off-design conditions [3]. 3D bumps are thoroughly investigated to diminish the penalty [4].

Another group of methods is based on the boundary layer energizing to prevent its separation. Among the most popular are mechanical and fluidic vortex generators (VGs). Mechanical VGs were investigated in [5-6] and proved their efficiency. The main drawback of the method is drag increase under cruise condition. Fluidic air-jet VGs and tangential jet blowing with position at 15% of the chord length to control SBLI were also considered in [5]. Boundary layer suction [7] and application of plasma actuators [8-9] were also considered. These concepts have multiple advantages, such as optional turning on during cruise regime and operation in a closed-loop strategy to optimize flow control. However, they require additional equipment resulting in a weight growth.

The concepts outlined can be used for buffet control. Mechanical trailing edge device (TED) which can change rear loading of an airfoil was considered in [10]. Fluidic VGs with air jets along with fluidic TED, where jet was blown on the pressure side, were studied in [11]. VGs are shown to be able to delay the buffet onset in the angle of attack domain by suppression of a separation zone downstream of the shock. In the case of fluidic TED, the separation was not suppressed and buffet was alleviated in the lift domain only.

In the present study buffet is simulated on the transonic supercritical airfoil P-184-15SR of TsAGI with the Reynolds number based on the chord $Re=2.6\times10^6$. Steady and unsteady Reynolds averaged Navier-Stokes equations are used with Spalart-Allmaras (SA) and shear stress transport (SST) turbulence models. The buffet onset boundary is calculated in a range of free-stream Mach numbers from 0.72 to 0.75. Flow control method based on fluidic air-jet blowing is proposed to alleviate the buffet onset. The jet is blown tangentially near the shock wave on the upper airfoil surface. Buffet alleviation is obtained for these cases.

2 PROBLEM STATEMENT

2D geometry investigated is a supercritical airfoil P-184-15SR developed at TsAGI with thickness 15% and chord length c=0.2 m. Reynolds number based on free-stream parameters and chord length is $Re=2.6\times10^6$. The baseline geometry corresponds to the smooth airfoil (figure 1).

To simulate the jet blowing, slot is added at x-coordinate X_j=60% of the chord with height h=0.15 mm (figure 2). The global mesh (figure 3) is not changed. The only grid nodes downstream of slot nozzle are added (figure 4). Computational grid consists of approximately 200 000 cells. Grid nodes are clustered normal to the surface inside the boundary layer so that

Y^+_1<1. Grid convergence study showed that the grid size is sufficient for numerical simulations [12].



Figure 1: Smooth airfoil P-184-15SR.

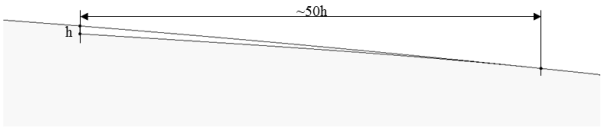


Figure 2: Slot geometry at X_j=0.6.

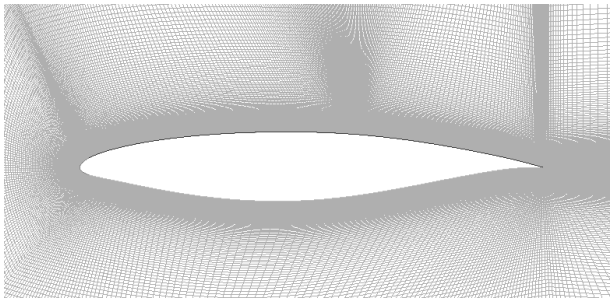


Figure 3: Grid near the airfoil.

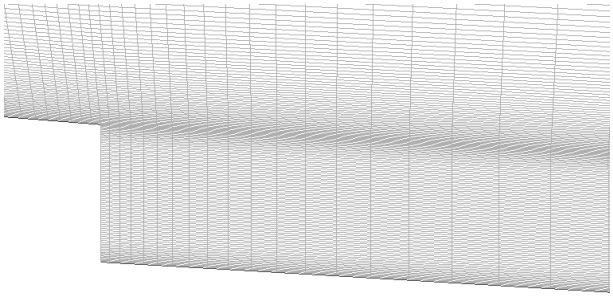


Figure 4: Grid near slot nozzle.

RANS and URANS equations are used for simulations. The calculations are carried out for the ideal compressible gas with laminar Prandtl number Pr=0.72. Laminar viscosity-temperature dependence is approximated by Sutherland law with Sutherland constant 110.4 K.

Numerical solutions are obtained using an implicit finite-volume method. The equations are approximated by a second-order shock-capturing scheme. The flux vector is evaluated by an upwind flux-difference splitting of Roe. Second order upwind scheme is used for spatial discretization of convective terms. Central-differencing scheme is used for diffusion terms. The second order time discretization is used for transient simulations. Dual time stepping scheme is used. Time step equals to $\Delta t=2x10^{-6}$ s with internal iterations converging up to the error $\sim 10^{-6}$. Further time steps and error decreasing showed no impact on solutions.

Two turbulence models were used: Spalart-Allmaras and SST. In RANS simulations, the difference between the models for pressure coefficient C_p is relatively small (figure 5). For these simulations, free-stream Mach number is M=0.73, angle of attack AoA=4 degrees, laminar-turbulent transition was fixed at x/c=0.15. In URANS simulations, SST model showed no buffet at all considered regimes so that SA model was used for further simulations.

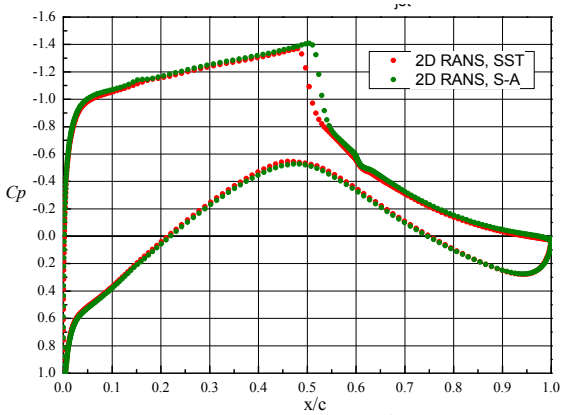


Figure 5: Cp distribution: M=0.73, AoA=4°, geometry with a slot.

Some preliminary RANS calculations were also done to compare the flowfield over the baseline geometry and the airfoil with a slot without a jet for M=0.77 and AoA=6 deg. Slot was placed at x/c=60% and its height was of the order of magnitude less than the boundary layer thickness h=0.15 mm. The difference was small enough (figure 6).

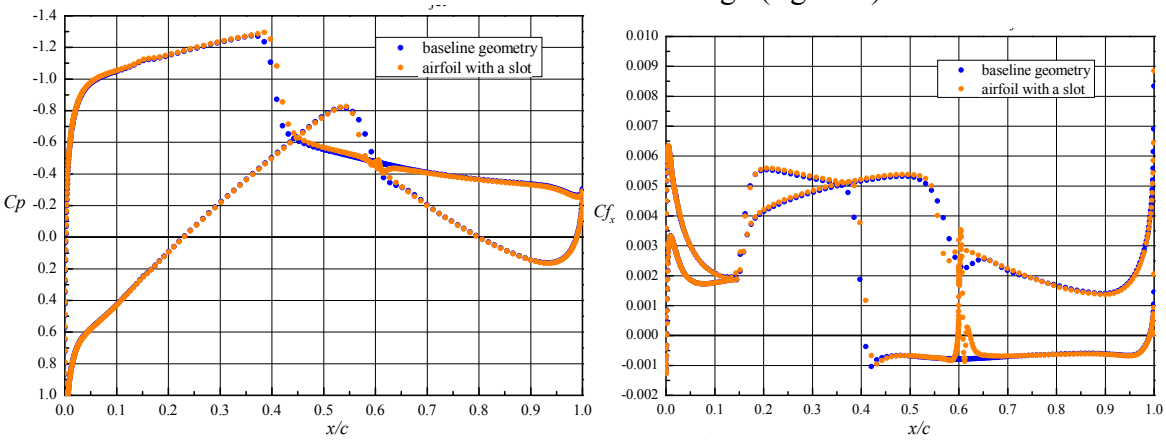


Figure 6: Cp and x-component of Cf distributions: M=0.77, AoA=6°, baseline geometry and airfoil with a slot.

3 BUFFET SIMULATION: BASELINE GEOMETRY

Buffet characterization on the P-184-15SR airfoil on the base of 2D URANS studies are carried out for the cases without jet blowing for the baseline geometry. Summary of these results on (M, α) plane is shown in Figure 7. Regimes below solid line correspond to regimes without buffet, while regimes under solid line correspond to buffet onset. Lift coefficient C_L convergence history for the case without and with buffet are shown in Figure 8. Buffet frequency was calculated using fast Fourier transformation. It varies from 99 Hz for M=0.72 and AoA=5° to 118 for M=0.74 and AoA=4.5°.

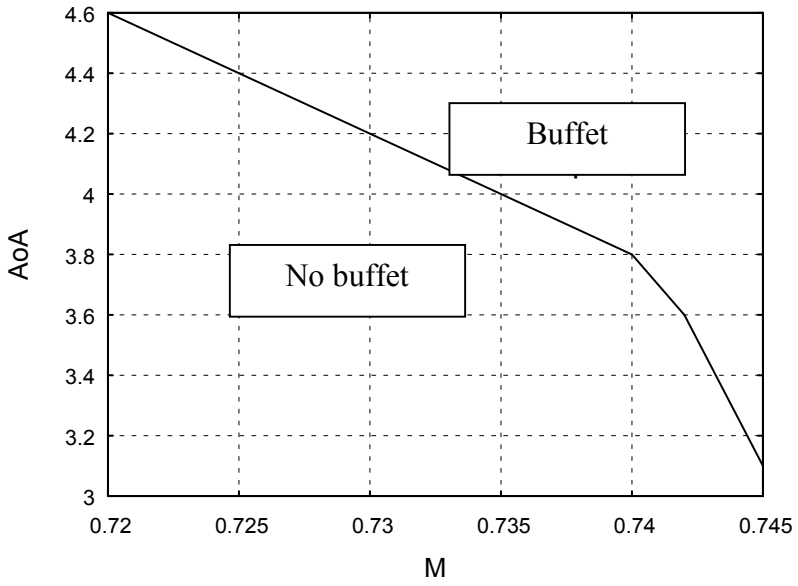


Figure 7: Buffet onset on (AoA, M) plane: baseline geometry.

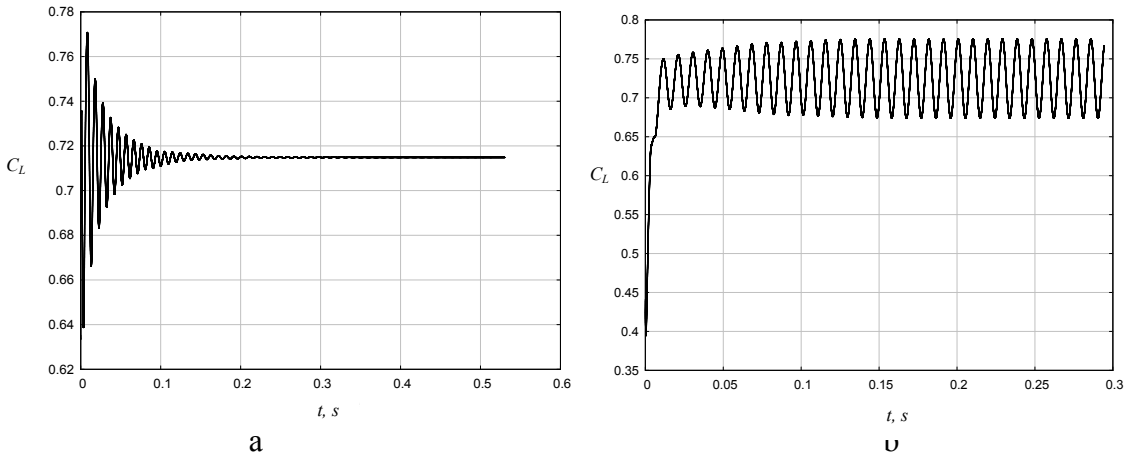


Figure 8: C_L convergence history for regimes without (a, M=0.73, AoA=4°) and with buffet (b, M=0.73, AoA=4.5°).

During the buffet period shock reaches its most upstream and downstream locations (figure 9). Figure 10a shows Cp distributions on airfoil surface. Blue curve corresponds to the mean C_p distribution where C_p is averaged for several periods of shock oscillation. In this case the shock is smoothed between the most upstream and downstream locations. Red curve corresponds to Cp distribution at time moment in which lift coefficient equals to the mean value. It is seen that there is a sufficient difference between these curves.

It should be noted that the shock motion also leads to the local maximum of the root mean square (RMS) of Cp in the region of shock oscillation (figure 10b).

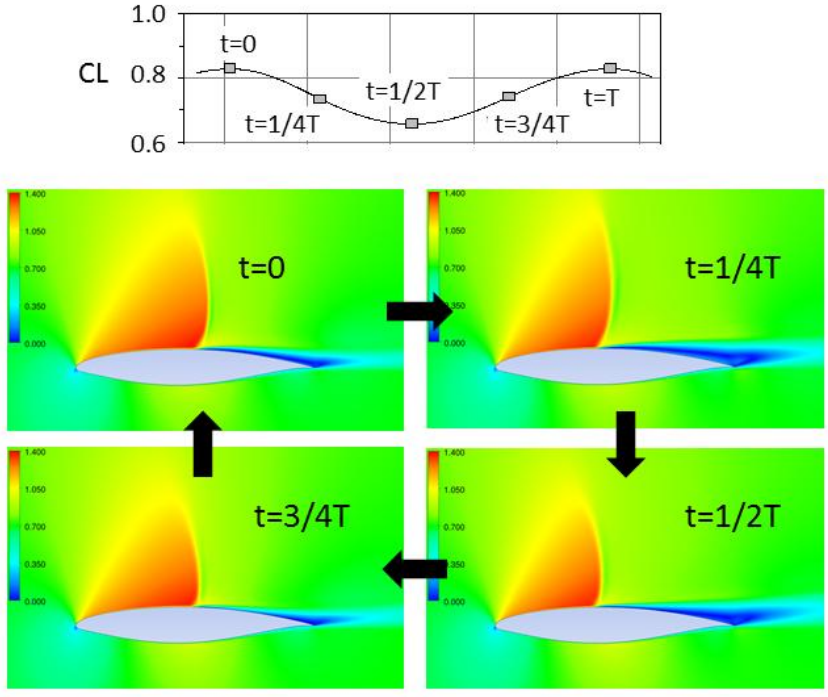
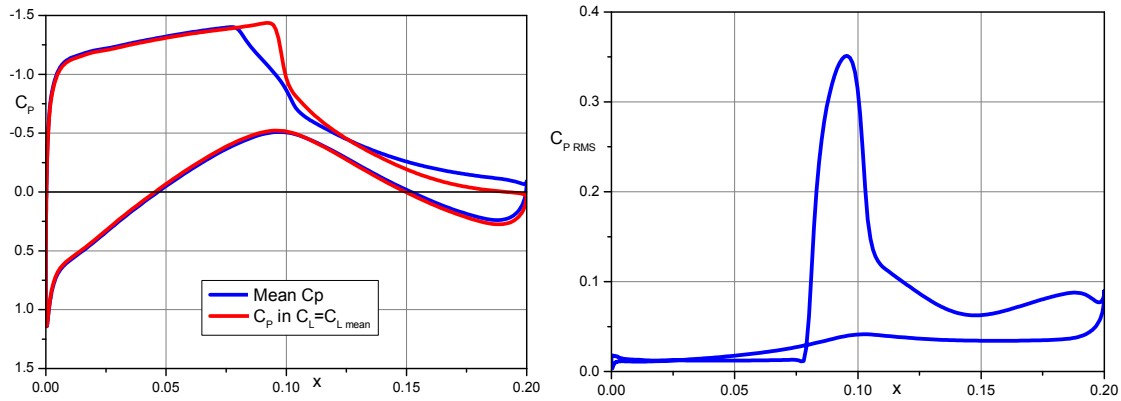


Figure 9: Mach number field evolution during the buffet period: M=0.73, AoA=5°.



a b

Figure 10: Mean and instantaneous Cp distributions (a) and RMS of Cp (b), M=0.73, AoA=5°.

4 TANGENTIAL JET BLOWING EFFECT

To control the buffet supersonic jet is blown out of the slot (figure 11) with momentum coefficient C_μ=0.0066. The mechanism of buffet alleviation and lift increase is associated with the separation zone elimination. As seen in figure 12, longitudinal component of wall friction coefficient is above zero downstream the shock for the case with jet blowing, while being negative for the baseline geometry.

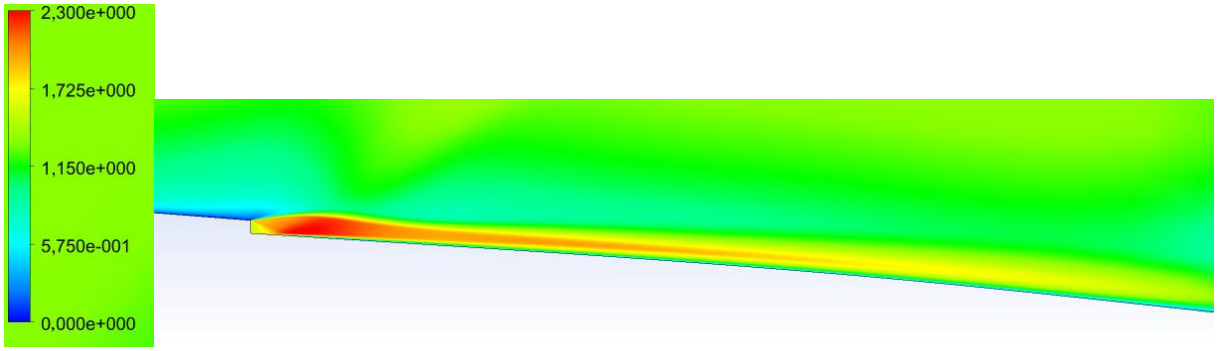


Figure 11: Mach number field near the slot nozzle, M=0.74, AoA=2°, X_j=60%, Cμ=0.0066.

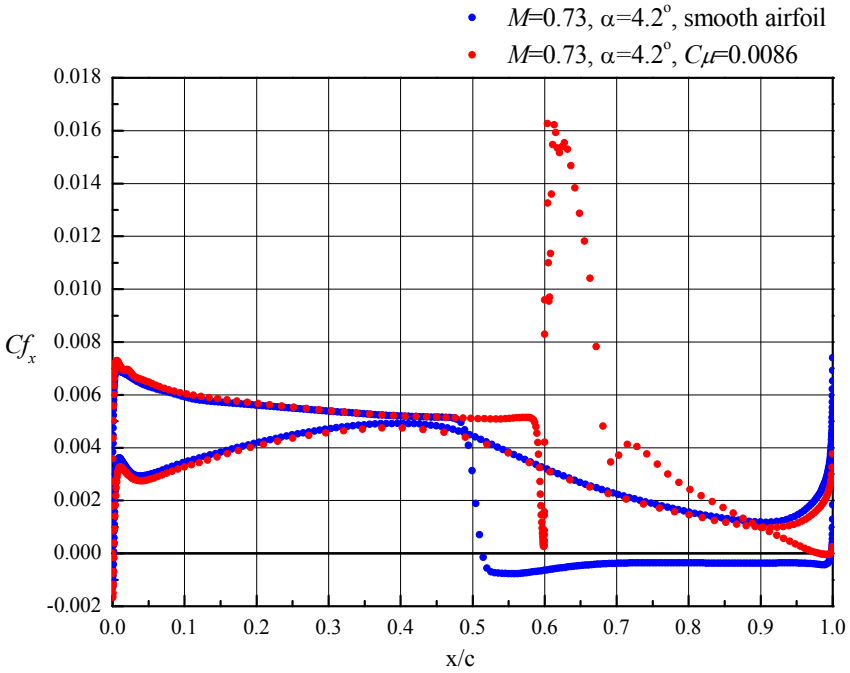


Figure 12: X-component of skin friction coefficient: smooth airfoil and airfoil with jet blowing: M=0.73, AoA=4.2°.

The mean values of C_L and standard deviations of C_L are plotted in figure 13. Both for the baseline geometry and for the jet blowing, RANS results close to URANS mean values. For the baseline case, deviation of lift curve from the linear regime is near AoA=2° while buffet onset regimes in URANS begin from AoA=4.2°. For the jet blowing with Cμ=0.0086, there is no buffet at all. The C_L increase reaches 0.2 on the linear stage and increases up to 0.4 at AoA=6°.

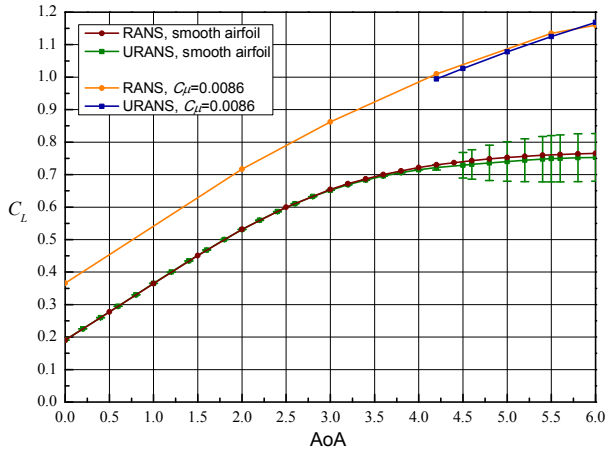


Figure 13: Lift curve for M=0.73 with and without tangential jet blowing.

Additional calculations were performed for low-momentum jet with Cμ=0.00069 (figure 14). In this case, there is no impact on C_L but AoA of buffet onset is delayed to 0.5-0.8° relative to the baseline geometry.

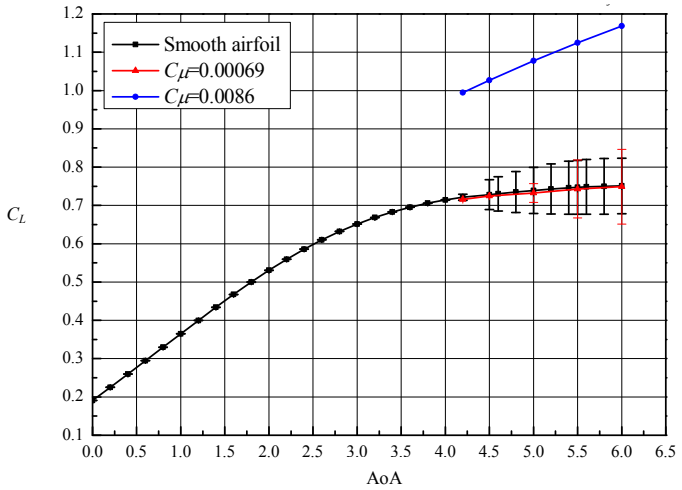


Figure 14:Lift curve for M=0.73, URANS, red curve – low momentum jet with Cμ=0.00069.

5 CONCLUSIONS

- Two-dimensional numerical simulations are carried out to characterize the buffet phenomenon on transonic supercritical airfoil P-184-15SR. The buffet boundary in AoA – Mach number is calculated.
- SST model showed no buffet while SA model was used for URANS simulations.
- Tangential jet blowing in the shock region is investigated to delay buffet. Numerical simulations showed that tangential jet blowing suppresses shock-induced separation and significantly increases lift coefficient.
- URANS results showed buffet onset delay both in the AoA and CL domain.

ACKNOWLEDGEMENTS

This work is supported by BUTERFLI Project of the 7th Framework Program of the European Commission (Grant Agreement 605605).

REFERENCES

[1] L. Bahi, J.M. Ross, T. Nagamatsu, Passive shock wave/boundary layer control for transonic airfoil drag reduction. *AIAA Paper* 1983-0137.

[2] S. Raghunathan, Passive control of shock-boundary layer interaction. *Progress in Aerospace Science*, 25: 271-296, 1988.

[3] J. Birkemeyer, H. Rosemann, E. Stanewsky, Shock Control on a Swept Wing. *Aerospace Science and Technology*, 4:147-156, 2000.

[4] H. Ogawa, H. Babinsky, M. Patzold, T. Lutz, Shock-Wave/Boundary-Layer Interaction Control Using Three-Dimensional Bumps for Transonic Wings. *AIAA Journal*, 46(6):1442-1452, 2008.

[5] H.H. Pearcey, Shock-induced separation and it's prevention by design and boundary layer control. *In: Boundary layer and flow control – it's principles and application. Ed. by Lachmann, Vol. 2, London, Pergamon Press,* 1170-1361, 1961.

[6] V.M. Gadetskiy, Ya.M. Serebriyskiy, V.M. Fomin, Investigation of the influence of vortex generators on turbulent boundary layer separation. *NASA TT* F-16056, 1974.

[7] P. Krogmann, E. Stanewsky, P. Thiede, Effects of suction on shock/boundary layer interaction and shock-induced separation. *J. Aircraft,* 22(1): 37-42, 1985.

[8] S.B. Leonov, D.A. Yarantsev, V.G. Gromov, A.P. Kuriachy, Mechanisms of Flow Control by Near-Surface Electrical Discharge Generation. *AIAA-Paper* 2005-780, 2005.

[9] A. Marino, P. Catalano, C. Marongiu, P. Peschke, C. Hollenstein, R. Donelli, Effect of High Voltage Pulsed DBD Plasma on the Aerodynamic Performances in Subsonic and Transonic Conditions. *AIAA-Paper* 2013-2752, 2013.

[10] D. Caruana, A. Mignosi, C. Robitaille, M. Correge, Separated Flow and Buffeting Control. *Flow, Turbulence and Combustion,* 71: 221-245, 2003.

[11] J. Dandois, P. Molton, A. Lepage, A. Geeraert, V. Brunet, J.-B. Dor, E. Coustols, Buffet Characterization and Control for Turbulent Wings. *Aerospace Lab Journal,* 6: AL06-01, 2013.

[12] K.A. Abramova, M.A. Brutyan, S.V. Lyapunov , A.V. Petrov, A.V. Potapchik, A.A. Ryzhov, V.G. Soudakov, Investigation of buffet control on transonic airfoil by tangential jet blowing. 6-th European Conference for Aeronautics and Space Sciences *(EUCASS), Poland, Krakow, June 25-July 3, 2015.*

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SHOCK WAVE BOUNDARY LAYER INTERACTION IN PROFILE CASCADES REPRESENTING ROTOR BLADINGS OF THE LAST STAGE OF LARGE OUTPUT STEAM TURBINES

Martin LUXA[1], David ŠIMURDA[1], Jaromír PŘÍHODA[1] and Jana VÁCHOVÁ[2]

[1] Institute of Thermomechanics, AS CR, v.v.i.
Dolejskova 5, Prague 8, The Czech Republic
e-mail: luxa@it.cas.cz, simurda@it.cas.cz, prihoda@it.cas.cz

[2] Doosan Skoda Power Co. Ltd.
Tylova 57/1, Pilsen, The Czech Republic
e-mail: jana.vachova@doosan.com

Keywords: shock wave-boundary layer interaction, long turbine blade, large output steam turbine, last stage

Abstract: *This paper is mainly focused on selected interaction phenomena occurring in the root, middle and tip section of the last stage rotor blading consisting of titanium blades Doosan Skoda 1375mm (54") and steel blades Doosan Skoda 1220mm (48"). Measurements and CFD simulations of the flow in selected sections along the blading radius show different types of interaction. Typically for the parts of the blade closer to the hub, only interaction of internal branch of exit shock wave with the suction side boundary layer takes place. At the blade tip where the flow is supersonic at both inlet and outlet, interaction of internal branch of inlet shock wave with the pressure side boundary layer occurs. The "reflected", additional shock wave arising from this interaction affect the flow field close to the suction side of the adjacent profile. The structure of interaction often depends on the flow conditions (nominal conditions and off-design conditions of flow, incl. different inlet flow angles).*

1 INTRODUCTION

Major part of modern, ultra long turbine rotor blades, which are designed for the last stages of low pressure cylinder in steam turbines of large output, operate in transonic and supersonic regimes of flow. The shock wave/ boundary layer interaction takes place along the major part of the blade (or even along the whole blade). The situation of Mach number distribution in the last stage of the low pressure cylinder of large output steam turbine is shown the Fig. 1. The inlet relative velocity for the titanium blade 54" reaches sonic condition approximately at the 85% of the blade height and relative exit velocities downstream the titanium blade are supersonic along the whole blade. The whole interblade channel is under nominal conditions aerodynamically choked. A small closed local region of supersonic velocity in the interblade channel, which occurs at the root section, extends upstream progressively along the blade height and covers larger part of interblade channel (see Fig. 2).

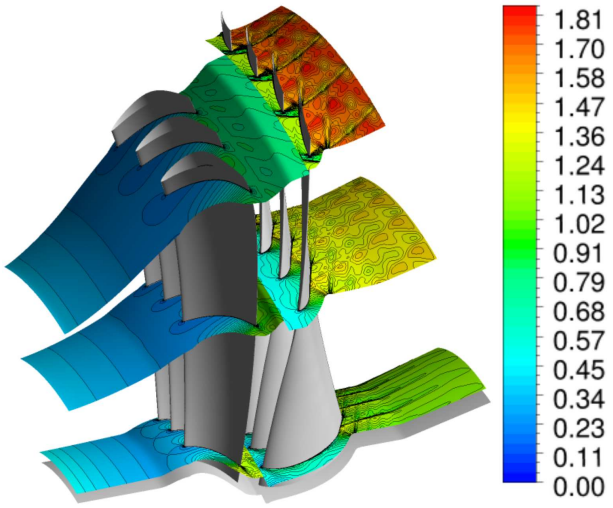


Figure 1: Mach number distribution along the blades in the last stage of low pressure cylinder with rotor titanium blade 1375mm (54").

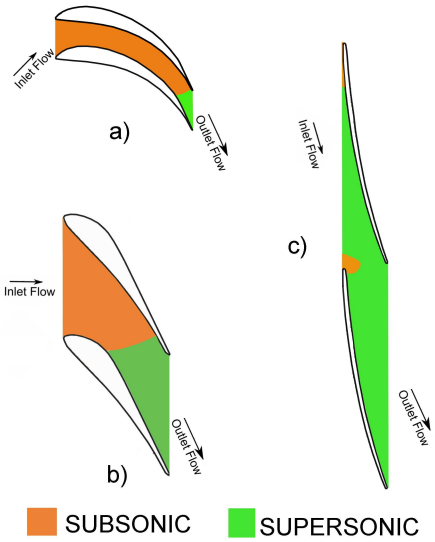


Figure 2: Proportion of supersonic and subsonic parts of flow field in long rotor turbine blade 54" (hub - a, midsection - b , tip - c).

In the tip region, the flow field is completely supersonic (except a small subsonic region near the leading edge downstream the inlet shock wave). Typical shock wave/boundary layer interaction, which occurs approximately along the whole blade height (considering the blade root), is interaction of internal branch of exit shock wave with boundary layer on the suction side of neighbouring blade (first interaction region) - see sign 1 in Fig. 3. Situation is much more complex in the tip region. In addition to this interaction, the interaction of internal branch of inlet shock wave upstream the profile leading edge with boundary layer on pressure side of previous profile often takes place. It is typical for this second interaction region, that new "reflected" shock wave arises which interacts with boundary layer on the opposite side of interblade channel, i.e. with boundary layer on the suction side of the blade (third interaction region). In some cases of the tip profile cascades (for higher inlet Mach numbers), the oblique internal branch of inlet shock is so significantly bent, that it misses suction side of the previous profile and propagates through the wake downstream the blade row and thus only the interaction of internal branch of exit shock wave with boundary layer on the suction side of neighbouring blade occurs.

Interaction with turbulent boundary layer in the first interaction region is often observed in experiments. However, the character of the interaction can be changed under different off-

design flow conditions. This is mainly due to changes of inlet flow angle, which generate not only different boundary layer formation and development on the suction side of the blade, but - logically - also different position of the boundary layer transition. Characteristic for the second interaction region in investigated tip cascades is the interaction with laminar or transitional boundary layer. It is therefore very useful to apply proper transitional models of turbulence in the CFD simulations. These models predict the situation in the flow field, which is observed at optical and pneumatic experiments in aerodynamic wind tunnel in Aerodynamic laboratory of Institute of Thermomechanics in Nový Knín significantly better. Finally, in the third region, interaction with turbulent boundary layer appears in interferograms and schlieren pictures.

Parallel application of both different methods of measurements and commercial and in-house developed codes of CFD effectively leads not only to significantly more complex information regarding the transonic complex flow field in closed curved channel but also to rapid development of numerical and experimental techniques.

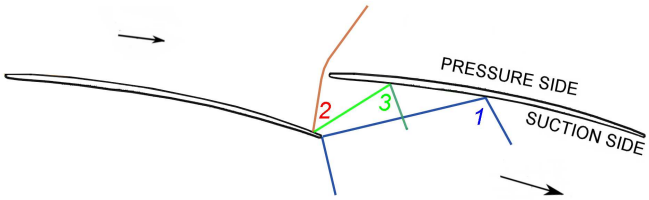


Figure 3: Three regions of shock wave - boundary layer interaction in the tip section of long turbine blade with supersonic inlet velocity

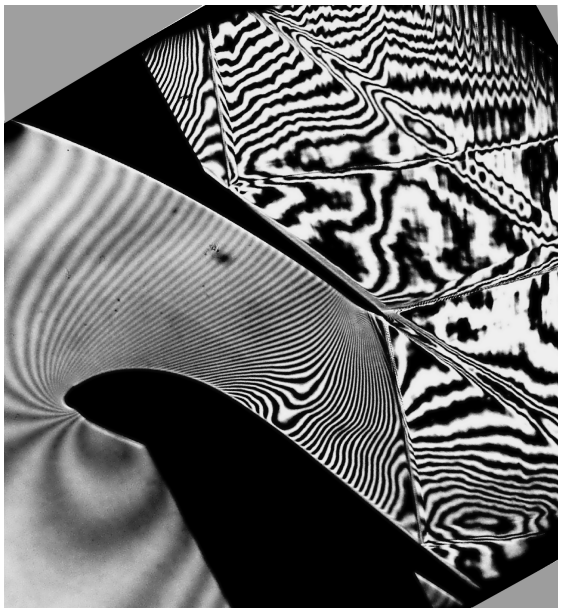


Figure 4: Interferogram showing shock wave/boundary layer interaction in the mid section of rotor turbine blade 48´´ long at isentropic exit Mach number M_{2is} = 1.348, incidence angle ι = + 30°.



Figure 5: Schlieren picture showing shock wave/boundary layer interaction in the mid section of rotor turbine blade 48´´ long at isentropic exit Mach number M_{2is} = 1.354, incidence angle ι = - 20°.

2 INTERACTION WITH THE BOUNDARY LAYER ON THE SUCTION SIDE

Interaction of a shock wave with the boundary layer on suction side of the blade occurs usually along the whole rotor blade in the last stage. The cambering of central lines of particular profiles of the blade changes along the blade height. This cambering depends on the changes of velocity triangles. Therefore, the surface of the suction side at hub or at the mid section is formed as convex and at the blade tip as concave.

The first example of shock wave/boundary layer interaction on the convex suction side of the profile is shown in Figures 4 and 5. Both photographs were taken under off-design condition of the flow in the tested cascade. The interferometric visualisation of the flow field in the mid section of the blade shows typical structure of interaction without flow separation when the boundary layer is most probably turbulent (see Fig. 4) [1]. Completely different situation regarding the structure of interaction is shown in the schlieren picture in Fig. 5. This interaction results in local flow separation and according to the length of the separated flow region, the boundary layer prior to interaction is probably laminar. During the interaction, the boundary layer changes to turbulent. These phenomena have impact on overall flow structure and the blade cascade losses. The difference between these two observed regimes lies in the great difference of the angle of incidence ($\Delta\iota = 50°$). Detail of the recompression area downstream the sonic fringe in interferograms in Figures 6 and 7 show stronger compression near the boundary layer in case of the overloaded blade ($\iota = +30°$, Fig. 6) in comparison with the situation in the Figure 7, where we can see situation in the flow field near the suction side at large negative incidence angle. The difference between these two situations in the recompression area is better apparent from the distribution of the Mach number on the suction side (Fig. 8) evaluated from the two interferograms (Figs. 6 and 7). This phenomenon together with both the different situation in the process of boundary layer forming at the leading edge area and the different length of stream line on the suction side upstream the interaction with shock wave might be the reasons for the position of transition and so for different behaviour of the boundary layer in the process of interaction with the internal branch of the exit shock wave.

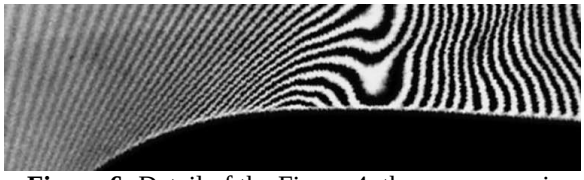


Figure 6: Detail of the Figure 4: the recompression area on the suction side, mid section of rotor turbine blade 48´´, isentropic exit Mach number $M_{2is} = 1.348$, incidence angle $\iota = + 30°$.

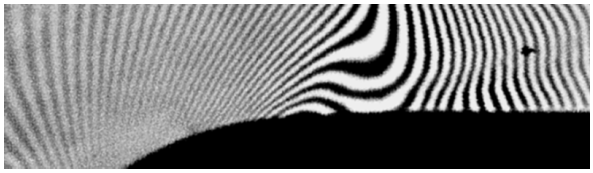


Figure 7: Interferogram related to schlieren picture in the Fig. 5: the recompression area on the suction side, mid section of rotor turbine blade 48´´, isentropic exit Mach number $M_{2is} = 1.354$, incidence angle $\iota = - 20°$.

The numerical simulations of flow through the mid section of turbine blade 48" were carried out by Rudas [2] or by Vachová [3]. In publication [3], the numerical simulations were carried out by the commercial code NUMECA using γ-Re transition model connected with the SST k-ω/k-ε turbulence model. Simulations of laminar and fully turbulent flow were also carried out for comparison. The calculated distribution of the skin friction coefficient along the profile chord with application of transition model is depicted in the diagram in Figure 9. The calculation shows the local boundary layer separation in the area of boundary layer interaction with shock wave, x/c ~ (0.65 ÷ 0.7). Although CFD predicted local flow separation, extent of the separated flow

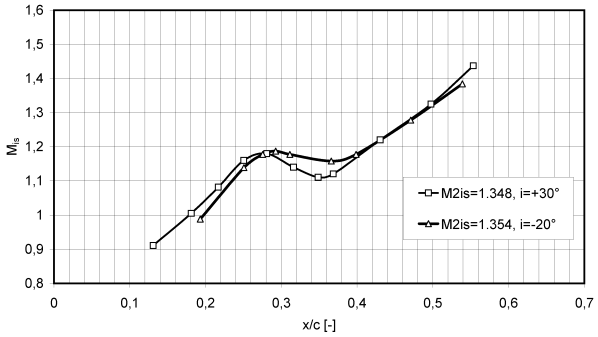


Figure 8: Isentropic Mach number distribution along the suction side near the recompression region at regimes shown in Figures 6 and 7.

region is under predicted (see Fig. 10 and [3]). Not even laminar calculation provided comparable extent of the separated flow region [3].

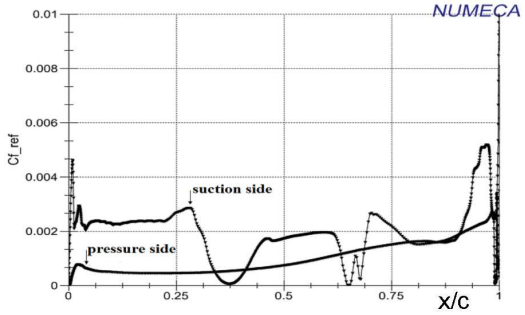


Figure 9: Distribution of skin friction coefficient along the blade profile - CFD simulation, transitional flow (Váchová [3])

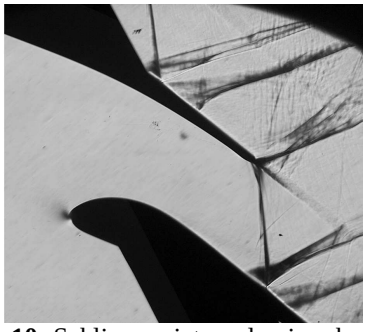


Figure 10: Schlieren picture, laminar boundary layer interaction with shock wave, incidence angle i = 0°, M_{2is} = 1.206

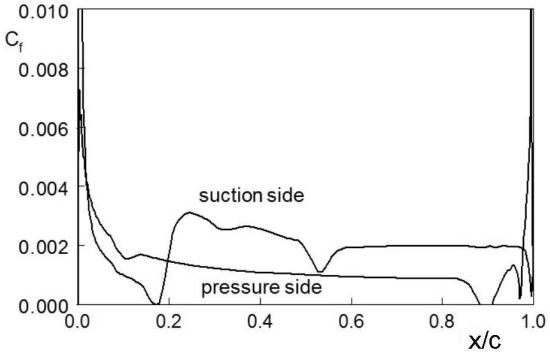


Figure 11: Distribution of friction coefficient C_f along the tip profile (long rotor blade 54"), M_1 = 1.45, M_{2is} = 2.0, CFD simulation (Straka, Příhoda [4]).

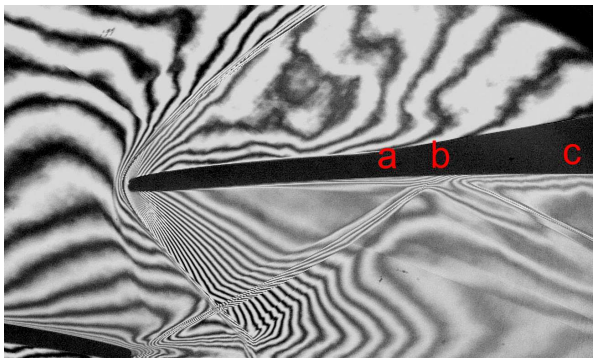


Figure 12: Interaction of internal branch of exit shock wave with boundary layer on suction side of neighbouring profile, (long rotor blade 60"), M_1 = 1.886, M_{2is} = 2.096.

Completely different situation occurs in case of the shock wave/boundary layer interaction on concave suction side in the tip section of the long turbine blade. The boundary layer is turbulent in the interaction area, nevertheless, the local separation bubble often occurs in the flow field near the point of interactions under off design conditions. Very strong expansion near the leading edge on suction side is followed by stagnation region in which the interaction takes place. Insufficient or negative gradient of static pressure along the suction side in the point just upstream of the interaction is typical for studied geometries of these tip profiles. According to the CFD simulations that were carried out by Straka and Příhoda [4], the boundary layer transition takes place commonly upstream the point of interaction (see Fig.11). The numerical simulation was carried out by means of the EARSM turbulence model according to Hellsten [5] completed by the bypass transition model with the algebraic equation for the intermittency coefficient proposed by Straka and Příhoda [6] and implemented into the in-house numerical code. The bypass transition on the suction side is

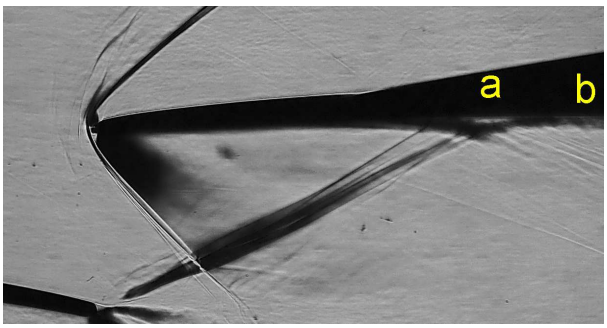
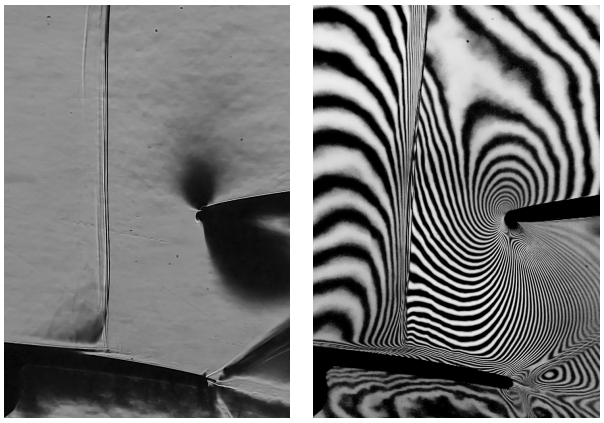


Figure 13: Interaction of internal branch of exit shock wave with boundary layer on suction side of neighbouring profile with separation bubble, (long rotor blade 60´´) , M_1 = 2.012, M_{2is} = 2.149.

evident from the distribution of C_f in the Fig. 11 (x/c ~ 0.15). The interaction of the boundary layer with inner branch of the exit shock wave can be seen in the C_f distribution at the distance x/c ~ 0.5 from the leading edge. The interferogram in the Figure 12 shows typical situation in the interaction area. In this region, thickness of the boundary layer (a) grows due to the adverse pressure gradient, the interaction with turbulent boundary layer (b) is more complex, and downstream the interaction, relatively very thick viscous shear layer (c) occurs. Transonic structures appearing in flow field in cascades representing tip sections of long turbine blades are generally extremely sensitive to changes of inlet and outlet parameters of flow. The schlieren picture in the Figure 13 shows situation in flow field at different inlet Mach number (or different incidence angle - unique incidence rule [7]) and different back pressure (in comparison with the flow parameters in Fig. 12). Local boundary layer separation (a) takes place in the area of interaction and very thick viscous shear layer develops downstream the interaction (b).



Schlieren picture Interferogram

Figure 14: Interaction of internal branch of detached inlet normal shock wave with boundary layer on pressure side of neighbouring profile (tip section of long rotor blade 60") , M_1 = 1.124, M_{2is} = 1.592.

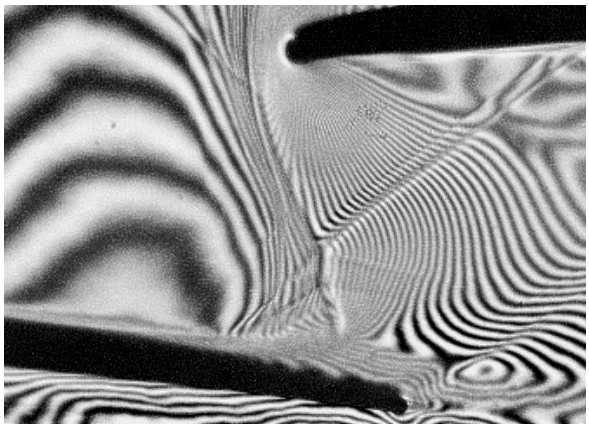


Figure 15: Complex interaction of internal branch of oblique inlet shock wave with boundary layer on pressure side of neighbouring profile (tip section of long rotor titanium blade 54") , M_1 = 1.416, M_{2is} = 1.746.

3 INTERACTION WITH THE BOUNDARY LAYER ON THE PRESSURE SIDE

This type of interaction appears commonly only in the tip region of ultra long turbine rotor blades (48", 54", 60"), where both the proper supersonic inlet velocity and suitable back pressure occurs. Interaction of internal branch of inlet shock wave with boundary layer on the pressure side in the tip section of rotor turbine blade 54" is described for example in [8]. It is typical for the pressure sides of investigated tip profiles, that the laminar/turbulent boundary layer transition takes place in the region of interaction with the inlet shock wave. An example of calculated distribution of friction coefficient along the pressure side is depicted in Fig. 11.

The interaction of detached normal inlet shock wave with boundary layer is shown in Fig. 14. These photographs were taken under off design condition at the cascade inlet and outlet. Both the schlieren picture and interferogram show local boundary layer separation and thick dissipative shear layer downstream the interaction. Increasing of the inlet Mach number leads to the attaching of inlet shock wave closely to leading edge and (by sufficiently low value of back pressure) to the bending of its inner branch. The point of interaction moves towards the trailing edge. The test results show that the interaction is more complex and local separation takes place once again (see Fig. 15). Impact of this complex interaction is that it suppresses the expansion area on the suction side of the neighbouring profile. Strong unsteady

behaviour of the flow field in the area of the interaction has been also observed. The sequence of interferograms, which were obtained by application of high speed CCD camera during interferometric measurement of tip cascade designed for blade 60" long, is shown in the Figure 16. Pictures were taken with frame frequency 6 kHz. The base of the shock wave moves cyclically upstream and downstream and the thickness of the shear layer downstream the interaction grows and decreases. This unsteady behaviour affects not only the cyclical changes of pressure distribution on pressure side but also the cyclical changes in the expansion in the region downstream the leading edge of the neighbouring profile (see the next chapter). As a result, cyclical torsion can occur in the region of the tip section of ultra long blade.

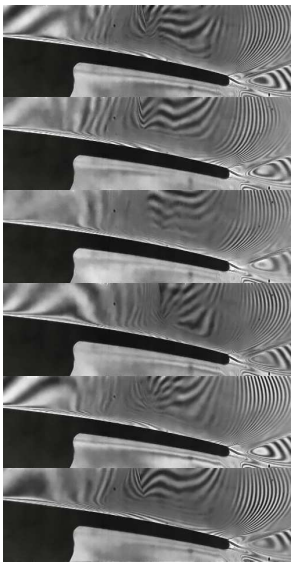


Figure 16: Unsteady interaction on pressure side (tip section of long rotor blade 60"), $M_1 = 1.124$, $M_{2is} = 1.797$, frame frequency 6 kHz.

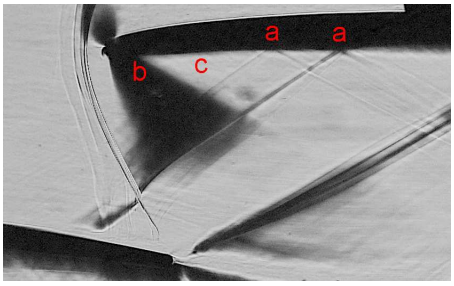


Figure 17: Complex interaction on the pressure and suction side (tip section of long rotor titanium blade 54"), schlieren picture , $M_1 = 1.45$, $M_{2is} = 2.0$.

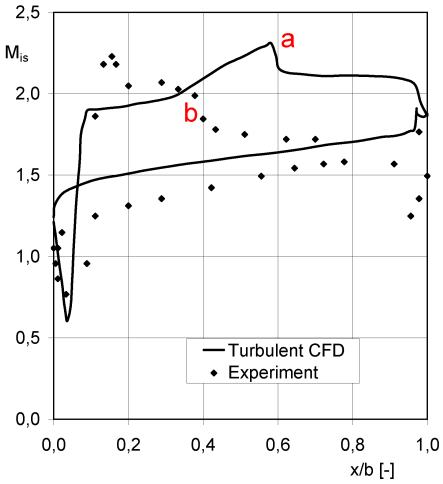


Figure 18: Distribution along the tip profile (tip section of long rotor titanium blade 54") obtained from experiment and by turbulent CFD simulation, $M_1 = 1.45$, $M_{2is} = 2.0$.

4 THE THIRD REGION OF THE INTERACTION

The possibility of the interaction of “reflected” shock wave (which arises from the interaction of the inner oblique part of inlet shock wave with the boundary layer on the pressure side) with boundary layer on the suction side (the third region of interaction – see Fig.3) is limited to the flow regimes in the tip region of long turbine blades with such an inlet Mach number and back pressure, which guarantees this interaction on the pressure side (or in the near wake area downstream the profile), and subsequently the formation of reflected wave.

Simple interaction without flow separation on the pressure side would guarantee expected additional supersonic expansion upstream the point (sign "a" in Fig. 17) where internal branch of exit shock wave interacts with boundary layer on the suction side. This solution (obtained by application of turbulent CFD procedure) is described by Rudas [2]. The additional supersonic expansion in experiment is suppressed, because the interaction of oblique part of inner branch of inlet shock wave with the laminar boundary layer on pressure side of neighbouring profile is much more complex [4] - see interferogram in Figure 15. Therefore, downstream the interaction of the inner branch of exit shock wave with boundary layer on suction side, the values of Mach number up to the trailing edge are significantly lower than in

the case of turbulent interaction on the pressure side (Fig. 18). This is also reason why the interaction of exit shock with the suction side boundary layer is located further upstream in experiment (sign "b" in Fig. 18). This fact has an essential negative influence on the aerodynamic loading of the studied profile. The interactions in the schlieren picture (sign "a" in the Figure 17) are interactions with turbulent boundary layer. Downstream the second interaction (and upstream the interaction with internal branch of exit shock wave) relatively thick dissipative layer develops. The region of the boundary layer transition is located next to the very strong supersonic expansion (sign "b" in Fig. 17) in the area (see "c" in Fig. 17) with practically neutral pressure gradient. This is in very good agreement with the result of numerical simulation (Straka, Příhoda [4]), where the transitional model was applied (see Fig. 11 in the paragraph 2). The distribution of the pressure coefficient C_p along the suction side of the profile obtained from experiment, turbulent CFD simulation (Fluent) and simulation, which was carried out by the in-house numerical code, are compared in the Fig. 19. Used in house code was based on the finite-volume solver for the RANS equations closed by the explicit algebraic model of Reynolds stresses completed by the modified algebraic transition model. The experimental data are evaluated from interferometric measurement, so the values of pressure coefficient are calculated from isoentropic values of static pressure distribution on the profile. Although boundary conditions were not identical in these three cases, the differences between particular approaches show conclusively the complexity of the processes in the transonic flow field in investigated tip sections of ultra long turbine blades.

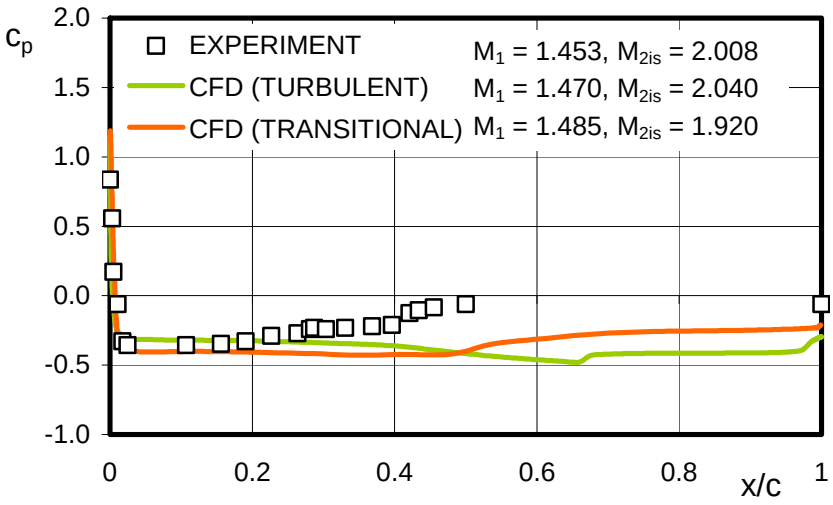


Figure 19: Distribution of the pressure coefficient along the suction side of tip profile obtained by in house code, commercial code and experiment, tip section of long rotor titanium blade 54".

5 CONCLUSIONS

The paper summarizes particular problems of boundary layer/shock wave interaction that is typical for modern ultra long rotor turbine blades and plays important role in case of recent designs of these blades.

Three basic situations have been shown, which appear in the inter blade channel: the interaction of internal branch of exit shock wave with boundary layer on suction side of the neighbouring profile; the interaction of internal branch of inlet shock wave with boundary layer on pressure side of the neighbouring profile; finally the interaction of "reflected" oblique

internal branch of inlet shock with boundary layer on suction side of the profile that generates the inlet shock wave.

Experimental data obtained from measurements on models of mid and tip section under design and off design conditions provide important information regarding complex and unsteady structures appearing in the transonic flow field within the boundary layer – shock wave interaction.

The application of different CFD procedures - commercial or developed "in house" proved the to predict the flow through such blade cascades correctly, it is necessary to employ transitional turbulence model when RANS is to be used. Results published in the paper show that the EARSM model performs reasonably well. The application of different experimental and numerical methods helps not only to validate numerical models and procedures but also to better understand physical phenomena, which appear in the measured flow fields as information provided by experiments is limited.

ACKNOWLEDGEMENTS

The work was supported by the Technology Agency of the Czech Republic under the grant TA03020277 and by the Czech Science Foundation under grant P101/12/1271. Institutional support RVO 61388998 is also gratefully acknowledged. The authors also thank DOOSAN ŠKODA POWER Co., Ltd., who made this research possible.

REFERENCES

[1] Délery J., Marvin J.G., (1986), Shock wave/boundary layer interactions, AGARDograph No. 280 Neuilly sur Seine, France, 222 p.

[2] Luxa, M., Šimurda, D., Šafařík, P., Synáč, J., Rudas, B., (2013), High-speed aerodynamics investigation of the midsection of 48´´ Rotorblade for the last stage of steam turbine, In *10th European Conference on Turbomachinery - Fluid Dynamics and Thermodynamics*, Lappeenranta, pp360-369 : University of technology Lappeenranta, pp. 360-369.

[3] Váchová, J., Luxa, M., Příhoda, J., Šimurda, D., (2015), Transition Model Application on Mid-Section Turbine Blade Cascade, In *Proceedings of the 12th International Symposium on Experimental Computational Aerothermodynamics of Internal Flows*, Lerici, pp. 1-8.

[4] Luxa, M., Příhoda, J., Šimurda, D., Straka, P., Synáč, J., (2015) Investigation of the Compressible Flow through the Tip-Section Turbine Blade Cascade with Supersonic Inlet, In *Proceedings of the 12th International Symposium on Experimental Computational Aerothermodynamics of Internal Flows*. Lerici, pp. 1-8.

[5] Hellsten, A., (2005), A new advanced k-omega turbulence model for high-lift aerodynamics, AIAA Jour., 43, pp.1857-1869.

[6] Straka, P., Příhoda, J., (2010), Application of the algebraic bypass-transition model for internal and external flows, Proc. Conf. Experimental Fluid Mechanics 2010, Liberec, pp.636-641.

[7] Luxa, M., Rudas, B., Synáč, J., Šafařík, P., Šimurda, D., (2014), Experimental Technique for Setup of Supersonic Inlet Flow in a Profile Blade Cascade. In *The Application of Experimental and Numerical Methods in Fluid Mechanics and Energy 2014 - Proceedings of the International Conference*, Žilina, pp.149-152.

[8] Luxa, M., Šimurda, D., Fořt, J., Fürst, J., Šafařík, P., Synáč, J., Rudas, B., (2015), Aerodynamic Investigation of Tip Section for Titanium Blade 54", Proceedings of the 11th European Conference on Turbomachinery Fluid Dynamics and Thermodynamics (ETC11), Madrid.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SHOCK WAVE BOUNDARY LAYER INTERACTION INVESTIGATION: CONTROL BY ROD VORTEX GENERATORS AND THEIR APPLICATION ON HELICOPTER ROTOR BLADES

F. Tejero E., P. Doerffer, P. Flaszyński

The Szewalski Institute of Fluid-Flow Machinery
Polish Academy of Sciences
Fiszera 14, 80-231 Gdańsk, Poland
{fernando, doerffer, pflaszyn}@imp.gda.pl

Keywords: shock wave, flow separation, flow control, vortex generator, helicopter rotor blades.

Abstract. *The paper describes the numerical simulations of a new passive flow control device (Rod Vortex Generator - RVG) and the possibilities to induce streamwise vortices for controlling of a shock wave boundary layer interaction. Firstly, a row of rods were placed at the lower wall of a curved wall nozzle with negative pressure gradient. The lower wall is designed in order to mimic similar flow conditions as for wings or helicopter rotor blades. The severe flow conditions (M = 1.43) result in a strong reverse flow which is controlled by the proposed vortex generators. On the other hand, the application of RVGs on helicopter rotor blades was analyzed. In the case of forward flight conditions, flow separation is usually present at the advancing side due to shock wave boundary layer interaction while dynamic stall appears at the retreating side. The numerical results of the flow past the AH-1G helicopter rotor blade were validated against flight test data and the flow separation on the advancing side was reduced by the proposed technology improving the aerodynamic performance (ratio thrust / torque).*

.

1 INTRODUCTION

Nowadays it is required to design devices of the highest efficiency leading to minimum emissions. For this reason, flow control devices are becoming more popular over the last years. For example, for energy consuming applications (e.g. rotorcrafts), the use of a proper flow control system might reduce the fuel consumption, thus limiting the CO_2 emissions.

The main feature of the aerodynamic of a helicopter is the unsteadiness of the flow due to the rotating parts (i.e. main and tail rotors). There have been several investigations concerning the flow separation reduction at the rear of the helicopter fuselage (non-rotating) which is the major component of the parasitic drag of the helicopter [1]. On the other hand, several flow control devices might be suitable at the rotating parts of the helicopter. It is important to remark that in forward flight conditions, the forward velocity is added to the blade's rotating velocity on the advancing side ($0^\circ < \psi < 180^\circ$), while the forward velocity is subtracted from the blade rotation on the retreating side ($180^\circ < \psi < 360^\circ$). This unsteadiness of the flow passing the rotor blade leads with compressibility effects on the advancing side due to severe Mach number and dynamic stall on the retreating side caused by high inflow angles. For this reason, a list of different flow control devices suitable for the rotating parts of the helicopter have been identified: vortex generators, gurney flaps, movable flaps, synthetic jets, surface suction, passive porous surfaces, riblets, etc. Working principle of each device, with the main benefits for rotorcraft application was recently published by Kenning *et al.* [2].

This paper presents the numerical investigation of separation reduction on helicopter rotor blades by Rod Vortex Generators (RVGs). As a previous step, a basic configuration of a curved wall nozzle with a strong shock wave ($M = 1.43$) which induces flow separation was studied. The application of RVGs induced streamwise vortices, limiting the reverse flow. Besides, their implementation on a wing configuration suggested that the ratio lift/drag was also possible to improve with this passive system [3]. RVGs are capable to induce streamwise vortices which enforce an exchange of momentum in the direction normal to the wall. High-momentum is transferred to the low-momentum area close to the wall, becoming fuller the boundary layer profile at the low-momentum area (separation control). Several parameters play an important role in the effectiveness of RVGs: diameter (ϕ), height (h), and pitch (Θ) and skew (α) angles, see Fig. 1. For all simulations presented in this paper, the RVGs were designed according to the boundary layer thickness (δ) of each case. According to previous investigations, the rods were designed with the following parameters: $\phi = 0.5 \cdot \delta$ and $h = 0.25 \cdot \delta$. In the case of pitch and skew angles, it was concluded during an optimization investigation that the strongest streamwise vortices are created for 30° and 45°, respectively.

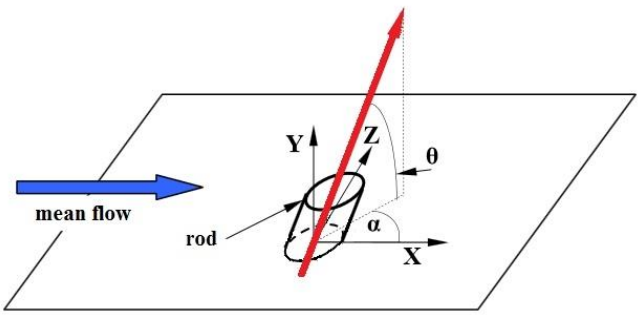


Figure 1. Schematic view of the Rod Vortex Generators

2 PHYSICAL AND NUMERICAL MODELLING

The numerical investigation was carried out with the FLOWer solver from DLR [4]. It is a modern, block-structured, aviation oriented, cell-centered code with different turbulent models closures. The ROT version of the code allows the usage of the chimera overlapping grids technique [5] and moving meshes. This approach allows to easily create grids for complex configurations (e.g. rotor blades). Besides, the implementation of RVGs on the rotor blade is solved with a single grid (RVG grid) which is added to the reference chimera set-up. Due to the capabilities of separation prediction, the two equations, low-Reynolds $k-w$ turbulence model LEA (Linear Explicit Algebraic Stress Model) [6] was chosen. The numerical algorithm uses a finite-volume, central scheme of 2nd order for the spatial discretization. The same explicit, Runge-Kutta method (CFL = 2.5) of time integration, was used for the steady simulations of the curved wall nozzle, as for the internal iterations of the implicit dual-time-stepping scheme of 2nd order applied for the helicopter rotor blade in forward flight conditions.

3 CURVED WALL NOZZLE

The experimental investigations of supersonic nozzles were carried out in flat and curved test sections. Previously, both nozzles were investigated within the UFAST project (Unsteady Effects in Shock Wave Induced Separation) [7] providing the largest available database concerning unsteady shock wave boundary layer interaction. The present numerical investigation is only focused in the curved wall nozzle (negative pressure gradient which mimics flow conditions around airfoils/wings or helicopter rotor blades). Its main feature is that in the curved duct the supersonic area is generated at the convex wall. The Mach number upstream of the shock wave increases with the increasing mass flow rate in the nozzle, until it is choked.

Prediction of the flow separation, its location and reattachment lines or size of the separation bubble is very important for accurate prediction of aerodynamic performance. In the case of channel flows, the proper prediction of flow structure is highly dependent on the appropriately resolved corner flows. The CFD results were compared with the measurements gathered during UFAST project: boundary layer profiles, static pressure on the wall, oil visualization and schlieren picture.

Figure 2 presents the Mach number contour map for the curved wall nozzle. The strong shock wave at the lower wall induces flow separation.

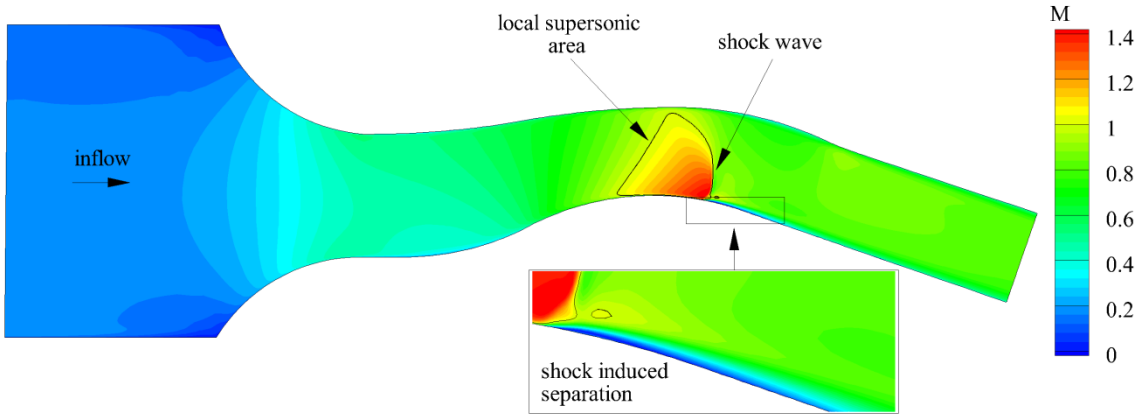


Figure 2. 2D view of the curved wall nozzle

In total, 16 RVGs are placed in spanwise direction covering the full span of the nozzle. Each rod is embedded in a rectangular box and a butterfly topology within the box is applied (see Fig. 3). The application of the RVGs increases significantly the number of mesh cells:

from $4.86 \cdot 10^6$ for the reference case to $21.7 \cdot 10^6$ for the flow control case. For all simulations the non-dimensional of the first layer of cells from the wall is of the order of $y^+ = 1$ (sufficient for resolving the laminar part of the turbulent boundary layer).

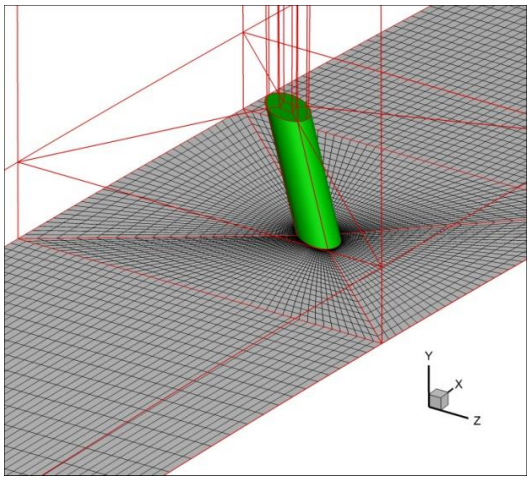


Figure 3. RVG detail and grid topology

Figure 4 presents the boundary layer comparison for the reference case at three cross sections: X = 15 mm (upstream of the flow separation), 67 mm (area of reverse flow) and 145 mm (downstream of the flow reattachment). It is noticeable that the predicted incoming boundary layer is overpredicted in reference to the measurements which results in an overprediction of the separation bubble. The separation bubble height is slightly larger in the simulation than in the experimental data. Lastly, the last cross section allows for a comparison of the boundary layer thickness for the reattached flow, showing its overestimation by the CFD simulations.

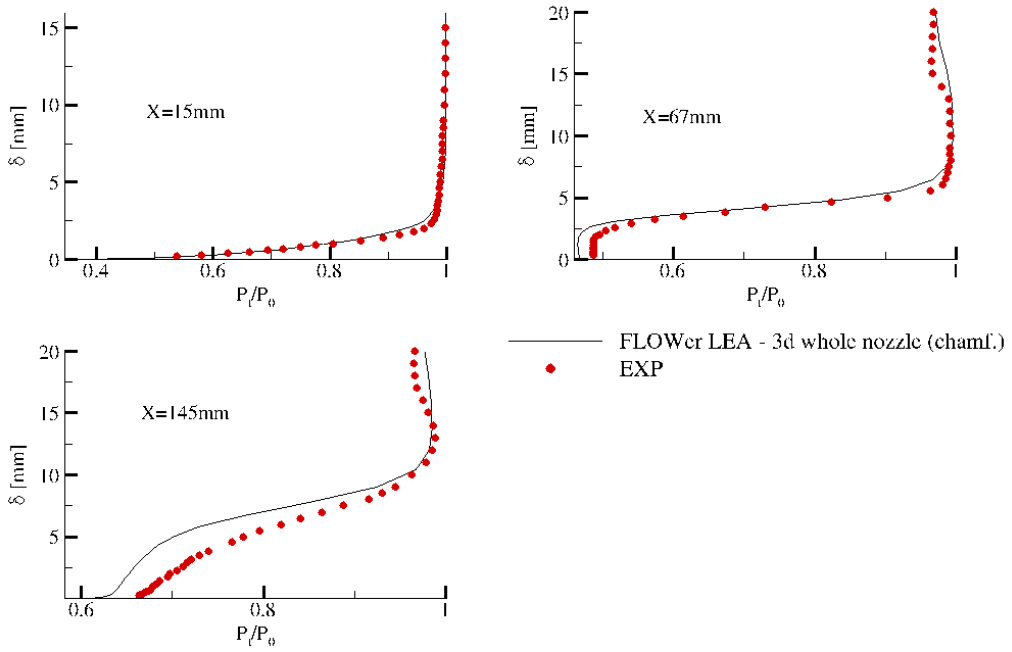


Figure 4. Boundary layer comparison – Reference case

Figure 5 presents the boundary layer comparison when the RVGs are applied. Both cross-sections are located downstream of the shock wave. It is visible that some differences of the boundary layer thickness and triple point location exist. More detailed results with different CFD solvers (Ansys – Fluent and Numeca Fine/Turbo) and turbulence models (Spalart-Allmaras and SST) can be found in [8].

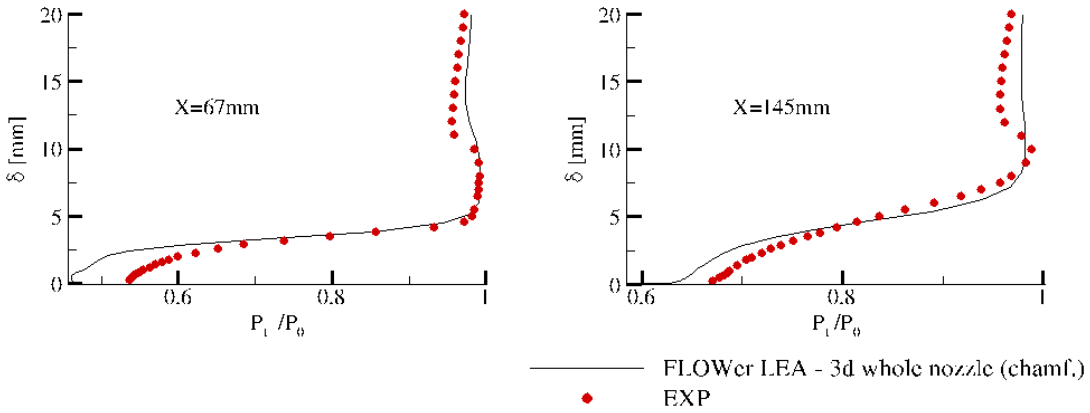


Figure 5. Boundary layer comparison – RVG case

Comparison of the separation bubbles size for the reference and flow control cases is shown in Figure 6. Although the reverse flow is not fully eliminated, the large separation bubble seen in the reference case is reduced and splitted into smaller bubbles, proving the effectiveness of this proposed passive flow control device.

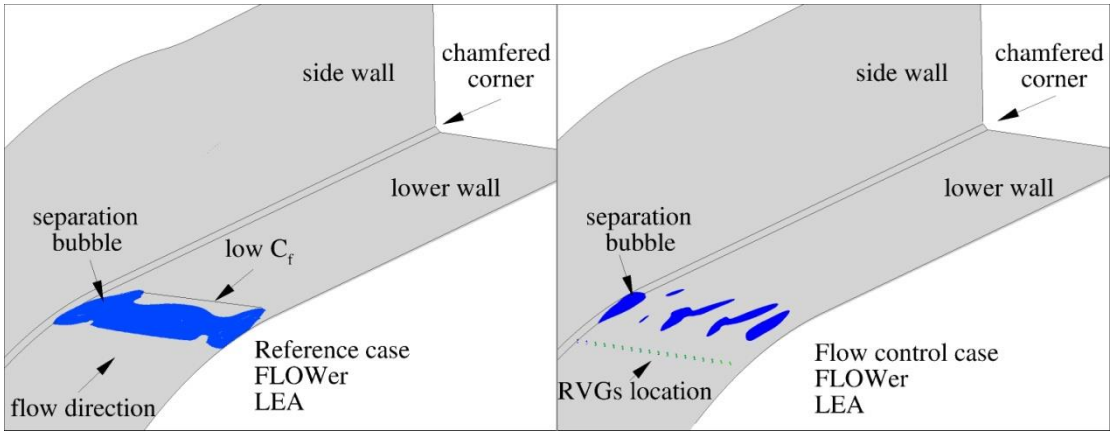


Figure 6. Separation control by RVGs on the curved wall nozzle

4 AIRFOIL CONFIGURATION

The aerodynamic enhancement by RVGs on an airfoil/wing configuration was investigated as previous step to helicopter rotor blades. For this reason, the flow around the NACA 0012 for severe conditions was studied. The Mach number was set to M = 0.8 and Reynolds number Re = $9 \cdot 10^6$. These flow conditions are representative at the advancing side of the flow passing the helicopter rotor blade in high-speed forward flight. Due to the severe inflow conditions, supersonic areas with strong shock waves induce flow separation. The angle of attack was increased to intensify the shock wave and therefore the reverse flow. The numerical vali-

dation of the reference case can be found in [3] as well as more detailed information concerning the geometry.

The onset of the reverse flow for the studied conditions is AoA = 1.4˚ (the flow is reattached) and above AoA = 2.0˚ the flow separation is strong enough to prevent the reattachment. The figure 7 presents an exemplary picture (AoA = 4.0˚) of the suction side of the airfoil with an array of four rod vortex generators. The shock wave induces flow separation at 40% of the chord. The application of RVGs leads to the increment of the skin friction downstream of the rods (due to the formation of streamwise vortices) and a delay of the separation line to 44% of the chord.

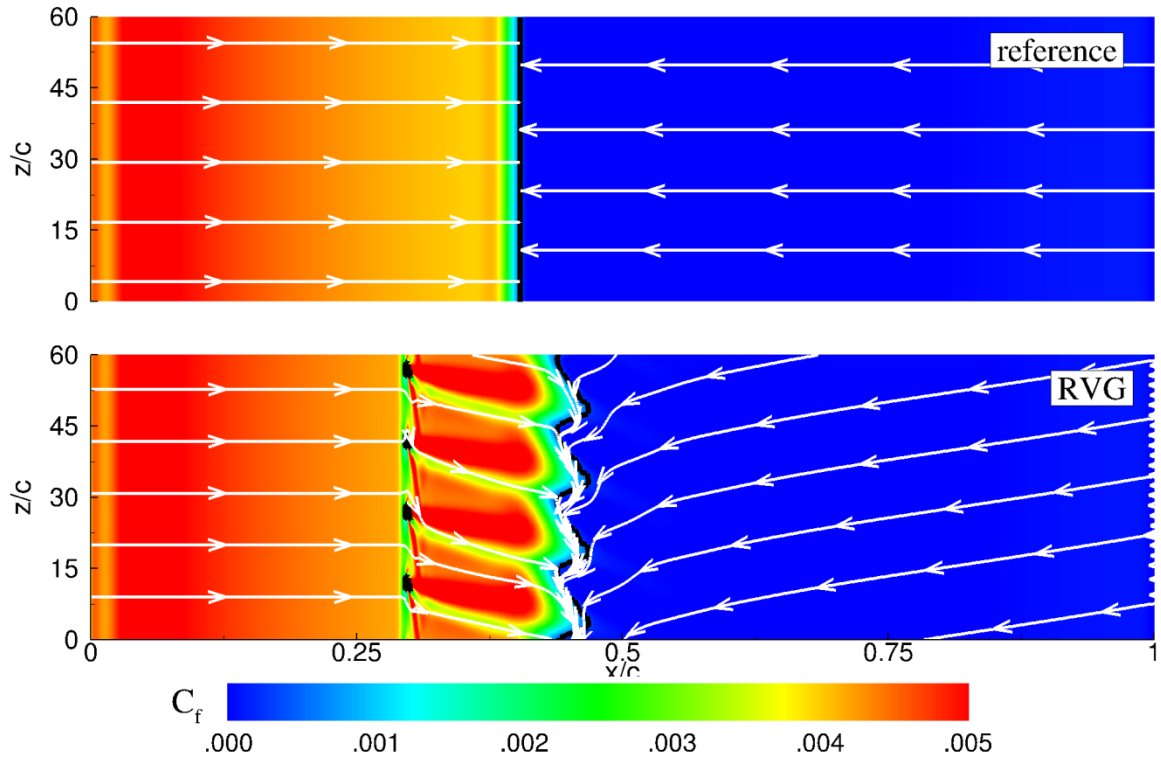


Figure 7. Friction coefficient and streamlines (M = 0.8, Re = $9 \cdot 10^6$, AoA = 4.0˚)

Figure 8 presents the effect of the RVGs on the normal force (C_n), drag (C_d) and pitching moment (C_m) coefficient. Below AoA = 1.4˚ for the polar graphs of the flow control device, no data is shown since the flow is fully attached to the profile, therefore it cannot be expected a positive effect of the RVGs. With increasing angles of attack, the reverse flow is stronger and the positive effect on terms of C_n appears. Besides, the stall angle of the airfoil is increased by 2.0˚ with the flow control system. The increment of C_n results in a drag penalty which is acceptable if the ratio C_n / C_d is compared. For the same drag, the application of the RVGs provides a higher C_n (better aerodynamic performance). Only a slight difference appears for high angles of attack when the C_m is compared. Therefore, it can be concluded that the application of RVGs on airfoils lead with a separation control which is related to the aerodynamic enhancement presented here.

5 HELICOPTER ROTOR BLADE IN FORWARD FLIGHT

The application of the passive rod vortex generators on the helicopter rotor blades was investigated for two different states of flight: hover and forward flight. In the case of hover conditions, the model helicopter rotor blade of Caradonna – Tung [9] was chosen due to the

availability of experimental data for severe conditions which induced flow separation and could be controlled by the proposed flow control system. For validation purposes, the transonic case with a tip Mach number of 0.87, collective angle equal to 8˚ and tip Reynolds number of $3.9 \cdot 10^6$ was chosen. A review of previous RANS simulations for this severe test case reveals a significant scatter with the measurements [10, 11]. The implementation of 14 RVGs at the tip of the blade lead to a separation control and aerodynamic enhancement (thrust coefficient was increased by 2.2% with a power consumption penalty of 1.2%) [12, 13].

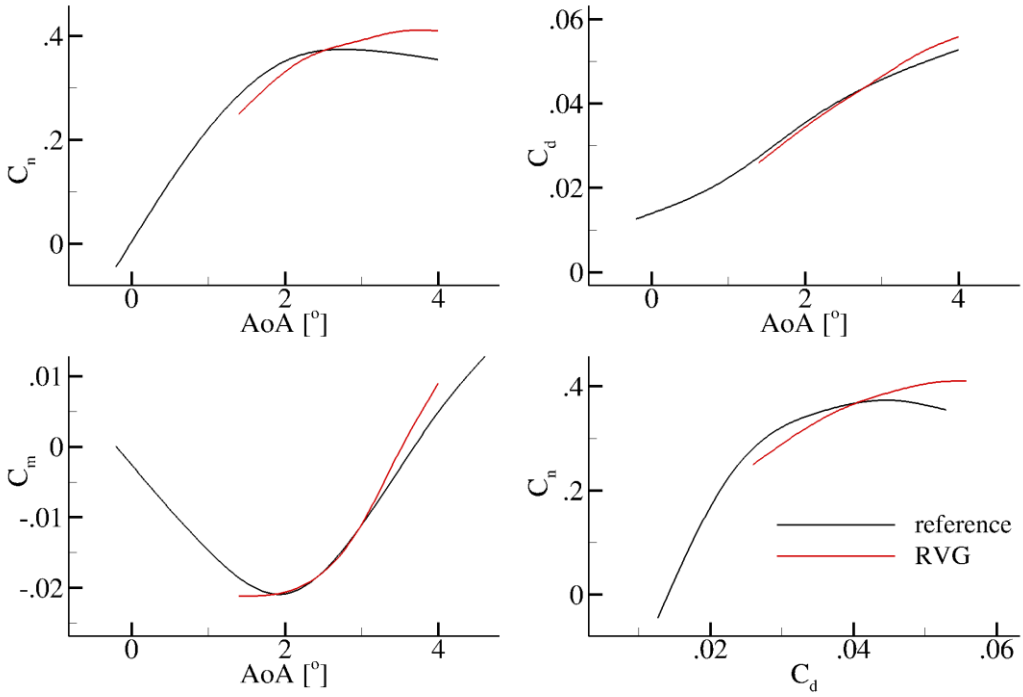


Figure 8. Comparison of C_n, C_d and C_m (M = 0.8, Re = $9 \cdot 10^6$, AoA = 4.0˚)

On the other hand, the numerical model for the AH-1G helicopter rotor blade in forward flight was validated against the flight test data gathered during the Tip Aerodynamics and Acoustics Test (TAAT)[14]. The AH-1G is a two person, single-engine, 2-bladed, teetering rotor with a 540 symmetrical airfoil section and a linear twist of -10˚ from the shaft to the tip. During the TAAT test different forward speeds were investigated and reported. The numerical model for three different markers (low-[15, 16], medium- and high-speed) was recently published in [17, 18] where it is also possible to find a detailed description of the geometrical model applied (chimera overlapping grids technique). Only the high-speed case reveals significant flow separation which might be controlled by the proposed flow control system. The reverse flow is caused by the severe flow conditions passing the rotor blade: tip Mach number of 0.63 and advance ratio equal to 0.38. Figure 9 presents the change of normal force coefficient (C_n) with the azimuthal position at the cross section r/R = 0.99 for different speed markers. The agreement with flight test data is acceptable despite of the numerical model simplifications (e.g. the influence of the fuselage is neglected and the blades are considered as rigid).

The rod vortex generators were designed according to the previous investigations (nozzle flows and airfoils) and the flow properties (boundary layer thickness) present at the advancing side. The application of this passive flow control device on helicopter rotor blades in high-speed forward flight leads with two different behaviors. Firstly, the shock wave boundary lay-

er interaction at the advancing side is controlled which is reflected in an enhancement of the normal force coefficient. Secondly, the dynamic stall present at the retreating side is not controlled by the RVGs due to the flow is separated already from the leading edge (rods are located in the reverse flow area and streamwise vortices are not created).

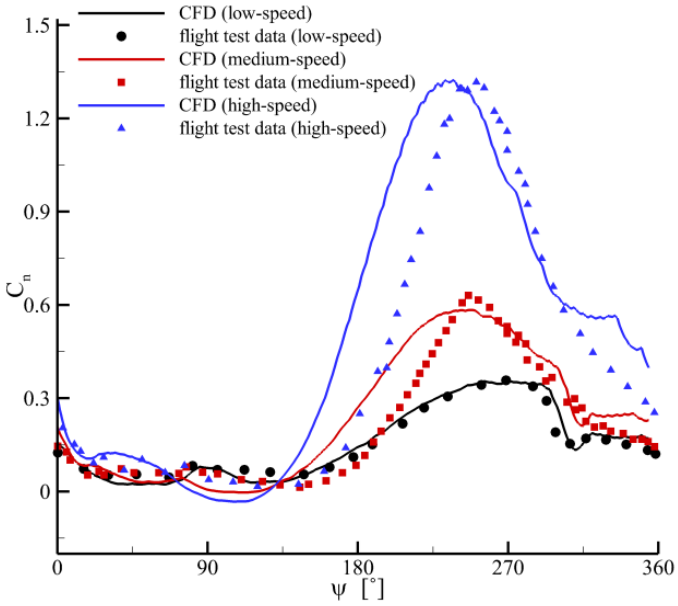


Figure 9. Instantaneous normal force coefficient C_n at r/R = 0.99 [17]

Figure 10 presents the rotor disk colored with the difference of the normal force coefficient (C_n) of the flow control configuration respect the reference case [18]. It is noticeable how the majority of the disk is colored with green (value of 0) due to the local influence of the rods in the global performance. Only their effects are present between the inner and outer rods with red color (positive effect) on the advancing side and blue color (negative effect) on the retreating side. The other azimuthal positions show a positive effect due to the low skin friction present for these positions.

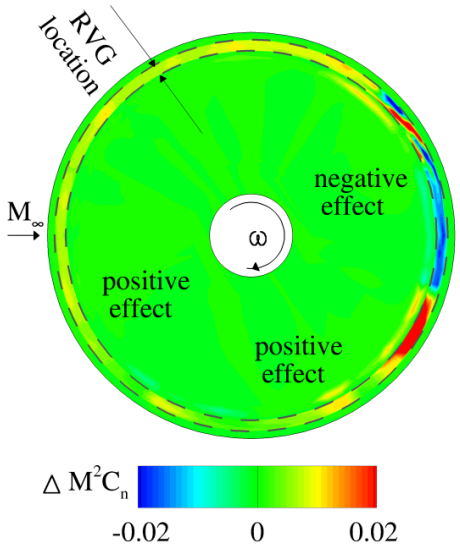


Figure 10. Normal force coefficient difference (flow control - reference) for the AH-1G helicopter rotor blade in high-speed forward flight

The flow separation induced by the strong shock wave is controlled by the proposed flow control device (see fig. 11). As shown at the selected cross-section and azimuthal positions for which reversed flow exists in case of the reference case, the flow is attached to the blade surface for the RVG configuration.

The positive effect of the RVGs on the advancing side leads with an increment of the thrust coefficient by 2.6% with a power consumption penalty of 1.1%.

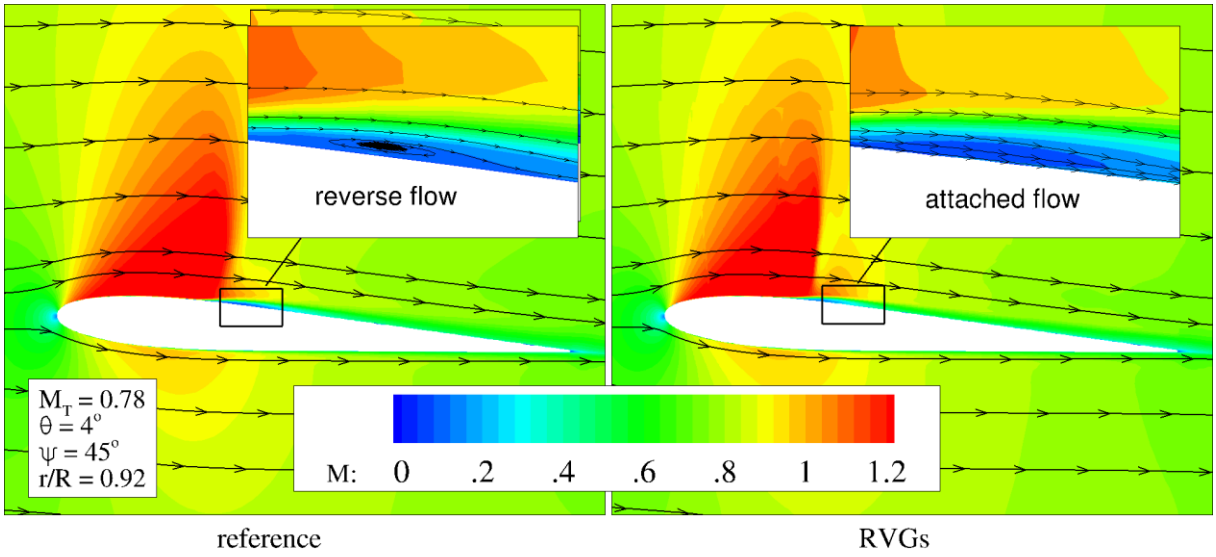


Figure 11. Contour map of Mach number at r/R = 0.92

6 CONCLUSIONS

The paper presents the numerical investigation of a passive flow control device (Rod Vortex Generators - RVGs) and the application on helicopter rotor blades. First of all, the flow separation control was investigated in more simple configurations as channel flows or airfoils. In both cases, the flow separation was reduced by means of streamwise vortices. Besides, the possibilities of aerodynamic enhancement on airfoils were studied with success.

The numerical simulations confirmed that the application of RVGs on helicopter rotor blades leads with a separation control which is reflected in an increment of thrust coefficient with a minor power consumption penalty.

REFERENCES

[1] B.G. Allan, N.W. Schaeffler, Numerical investigation of rotorcraft fuselage drag reduction using active flow control. *American Helicopter Society 67th Annual Forum*, Virginia Beach, VA, May 3-5, 2011.

[2] O.C. Kenning, I.W. Kaynes, J.V. Miller, The potential application of flow control to helicopter rotor blades. *30th European Rotorcraft Forum*, Marseilles, France, September 14-16, 2004.

[3] F. Tejero E., P. Doerffer, O. Szulc, Shock wave induced flow separation by means of air-jet and rod vortex generators, *TASK Quarterly*, **19**(2), 167-180 2015.

[4] C.C. Rossow,, N. Kroll, D. Schwamborn, The MEGAFLOW project – numerical flow simulations for aircraft. *Progress in Industrial Mathematics at ECMI 2004, edited by A. Di Bucchianico, M. Mattheij and M.A. Peletier, Sptringer,* 2006.

[5] T. Schwarz, The overlapping grid technique for the time accurate simulation of rotorcraft flows, *31st European Rotorcraft Forum,* Florence, Italy, September, 13-15, 2005.

[6] T. Rung, H. Lubcke, M. Franke, L. Xue, F. Thiele, S. Fu, Assessment of explicit algebraic stress models in transonic flows, *4th International Symposium on Engineering Turbulence Modelling and Measurements,* Ajaccio, France, May, 24-26, 1999.

[7] P. Doerffer, C. Hirsch, J.P. Dussaugem H. Babinski, G.N. Barakos, Unsteady effects of shock wave induced separation. *Notes on Numerical Fluid Mechanics and Multidisciplinary Design, Springer,* 2010.

[8] F. Tejero E. Numerical investigations of the possibilities of separation reduction on a helicopter rotor blade by means of streamwise vortices, *PhD thesis,* Gdansk, 2016.

[9] F.X. Caradonna C. Tung, Experimental and analytical studies of a model helicopter rotor in hover. *NASA Technical Memorandum 812322,* 1981.

[10] H.J. Kang, O.J. Kwon, Unstructured mesh Navier-Stokes calculation of the flowfield of a helicopter rotor in hover. *Journal of the American Helicopter Society, **47**(2), 90-99, 2002.*

[11] H. Sun, S. Lee, Response surface approach to aerodynamic optimization design of helicopter rotor blade, *Journal of Numerical Methods in Engineering,* **64**, 125-142, 2005.

[12] F. Tejero E., P. Doerffer, O. Szulc, Aerodynamic analysis of potential use of flow control devices on helicopter rotor blades, *Journal of Physics: Conference Series, **503** (012067), 2014.*

[13] F. Tejero E., P. Doerffer, P. Flaszynski, O. Szulc, Numerical investigation of rod vortex generators on hovering helicopter rotor blade, *6th European Conference on Computational Fluid Dynamic,* Barcelona, Spain, July, 20-25, 2014

[14] J.L. Cross and M.E. Watts, Tip aerodynamics and acoustics test. *NASA Reference Publication 1179,* 1988.

[15] F. Hernandez, W. Johnson, Correlation of airloads on a two-bladed helicopter rotor, *NASA Technical Memorandum 103982,* 1993.

[16] J. Ahmad, P.N Duque, Helicopter rotor blade computation in unsteady flows using moving overset grids, *Journal of Aircraft,* **33**(1), 54-60, 1996.

[17] F. Tejero E., P. Doerffer, O. Szulc, J.L. Cross, Numerical simulation of the tip aerodynamics and acoustics test, *Journal of Thermal Science,* **25**(2), 153-60, 2016.

[18] F. Tejero E., P. Doerffer, O. Szulc, Application of a passive flow control device on helicopter rotor blades, *Journal of the American Helicopter Society,* **61,** 012001-1 – 012001-13, 2016.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ANALYSIS OF EFFECTS OF SHAPE AND LOCATION OF MICRO-TURBULATORS ON UNSTEADY SHOCKWAVE-BOUNDARY LAYER INTERACTION IN TRANSONIC FLOW

Janusz Sznajder[1], **Tomasz Kwiatkowski**[2]

[1] Institute of Aviation
Aleja Krakowska, 110/114, 02-256 Warsaw, Poland
e-mail: janusz.sznajder@ilot.edu.pl

[2] Institute of Aviation
Aleja Krakowska, 110/114, 02-256 Warsaw, Poland
tomasz.kwiatkowski@ilot.edu.pl

Keywords: Laminar Flow, Tripping laminar boundary layer, Computational Fluid Dynamics, Transonic flow

Abstract. *Tripping of laminar boundary layer on V2C laminar airfoil designed by Dassault Aviation was investigated computationally using two types of turbulators: simulated array of delta-shaped, counter-rotating vurtex generators producing chordwise-oriented vortices and by vanes perpendicular to airfoil surface and to flow producing spanwise-oriented vortices. Unsteady Reynolds-Averaged Navier-Stokes equations were solved using as a closure the four-equation Transition SST turbulence model implemented in ANSYS Fluent solver. The assumed flow conditions involved strong shockwave producing flow separation at Mach number of 0.70. Both types of turbulators produced their specific effects when tripping laminar boundary layer, which varied, depending on their placement in different chordwise positions. Both solutions for tripping laminar boundary layer have reduced flow unsteadines, resulting from oscillations of shockwave. Using delta-shaped vortex generators it was possible to obtain slightly improved lift-to-drag ratio for one combination of angle of attack and Mach number and the turbulator composed of parallel perpendicular-vane turbulator has shown strong buffet-damping properties.*

1 INTRODUCTION

Interaction of laminar boundary layer with shockwave leads to occurrence of laminar separation bubble at the foot of the shockwave and of subsequent transition of the boundary layer to turbulent over the separation bubble when flow is crossing the shockwave. This phenomenon occurs for internal and external flows, with oblique and normal shockwaves. Interaction of laminar boundary layer with shockwave on aircraft wing with laminar airfoil in transonic flow may lead to separation of the flow behind shockwave, large drag increase and also to buffet phenomenon - oscillations of shockwave, resulting from pressure waves reflected from wake behind wing. In FP7 TFAST project different methods of tripping of laminar boundary layer in front of the shockwave were tested with the aim of preventing the laminar boundary layer being penetrated by shockwave. One of questions in the project is how close to shockwave tripping of laminar boundary.layer may be conducted in order to obtain turbulent shockwave-boundary layer interactions (SWBLI) . Another question, investigated by the authors of the paper is how to obtain improvement of aerodynamic characteristics of the laminar airfoil in transonic flow with shockwave terminating the supersonic region.

2 CONDITIONS AND METHODS OF INVESTIGATIONS

The object of investigations was V2C laminar airfoil designed by Dassault Aviations especially for investigatuions of SWBLI in the TFAST project. The conditions of investigations were defined as Mach number Ma=0.7 and Reynolds number Re=2.67, resulting from the chord length c=0.2m, being equal to chord of the airfoil model investigated in the transonic wind tunnel at Institute of Aviation. In these conditions, depending on the value of angle of attack the flow is either steady, with stationary shockwave, or unsteady, with low-amplitude buffet oscillations of shockwave. The flow analysis was conducted by solution of Unsteady Reynolds-Averaged Navier-Stokes Equations (URANS) using ANSYS Fluent solver [1]. As a closure of the system of equations the four-equation Transition SST turbulence model was applied having as variables k – turbulent energy, ω - specific dissipation rate of k, γ - intermittency (probability of boundary layer being turbulent) and Re_θ - Reynolds number based on boundary layer momentum thickness. Second-order spatial and temporal discretisation of URANS equations was applied. In conditions of transonic flow with shockwave on the upper surface two methods of tripping of laminar boundary layer were investigated, one involving delta-shaped vortex generators producing vortices with chordwise-oriented rotation axes and with another type of vortex generators – micro-vanes, perpendicular to flow and airfoil surface, generating vortices with spanwise-oriented rotation axes.

3 MODELLING OF TRIPPING OF LAMINAR BOUNDARY LAYER

Tripping of laminar boundary layer (b.l.) was simulated by resolving in the computational grid two technical solutions submerged in the laminar b.l., generating vortices: delta-shaped vortex generators and micro-vanes, perpendicular to airfoil surface and to flow, shown in Figure 1. The first concept is classical vortex generator inclined to flow at an angle of 14°, similar to inclination used by other researchers in transonic flow [2], [3]. Its height was chosen according to an experimentally obtained criterion for minimum Reynolds number, based on turbulator height, capable of tripping laminar boundary layer [4]. The other concept may be considered as a series of micro-Gurney tabs generating vortices in the internal locations on airfoil chord, as opposed to classical Gurney tab, which is located at airfoil trailing edge. The first concept was investigated in computational domain involving lower and upper wind-tunnel walls in order to make easier comparisons with results of wind-tunnel investigations in test section of 0.6m height. On the side walls of the computational domain, located at the dis-

tance from generators' side vertices equal 10% of the width of span fragment occupied by the device, the boundary condition of symmetry was applied. This solution was applied for modelling of an array of counter-rotating vortex generators. Inlet boundary condition was located approximately six airfoil chords in front of it, as a distance where physical inlet to test section of wind tunnel was located. On the inlet surface the pressure inlet boundary condition was imposed, and on the outlet surface, located also approximately six airfoil chords behind airfoil pressure outlet boundary condition was imposed. The values of static and dynamic pressure on inlet and outlet surfaces were chosen based on tests in the wind-tunnel. The other concept of turbulator was investigated using open-space boundary conditions, pressure far-field and pressure outlet, located approximately 50 airfoil chords in front of and behind the airfoil. This decision was taken because of the planned test case for this turbulator – buffet oscillations of shockwave which proved easier to simulate in open-space flow than in simulated conditions of solid-wall test section of wind tunnel.

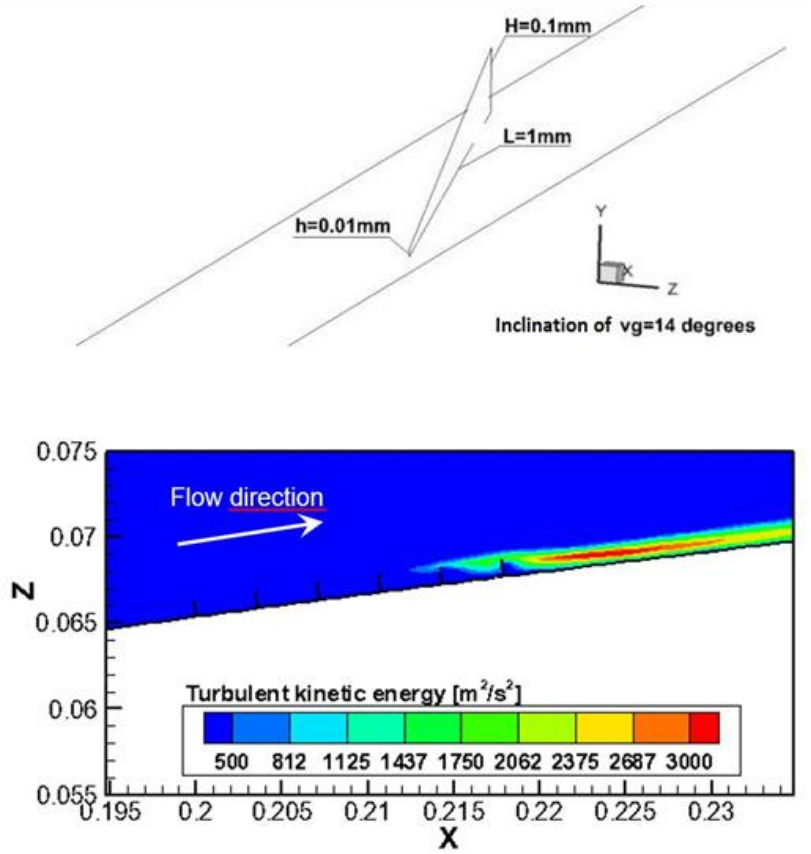


Figure 1. General ideas of concepts of proposed laminar boundary-layer tripping devices: delta-shaped vortex generators on upper part and perpendicular micro-vanes on lower part.

4 INVESTIGATED CASES

4.1 Baseline case – V2C airfoil with natural laminar-turbulent transition

Before simulating laminar-turbulent transition induced by the tripping devices, flow simulations were conducted for the clean airfoil in order to obtain numerical solution of natural

transition with the applied computational methods. Figures Figure **3** - Figure **5** show the subsequent stages of laminar-turbulent transition determined by the flow solver. The transition takes place in front of the shockwave, in the region of pressure plateau at 40% – 45% chord shown in Figure 2. In Figure 3 is shown rise of intermittency in the boundary layer in this region, and in plot of tangential stress on airfoil surface in Figure 4 it can be seen that rise of intermittency is preceded by the separation of laminar boundary layer visible as change of sign of tangential stress.

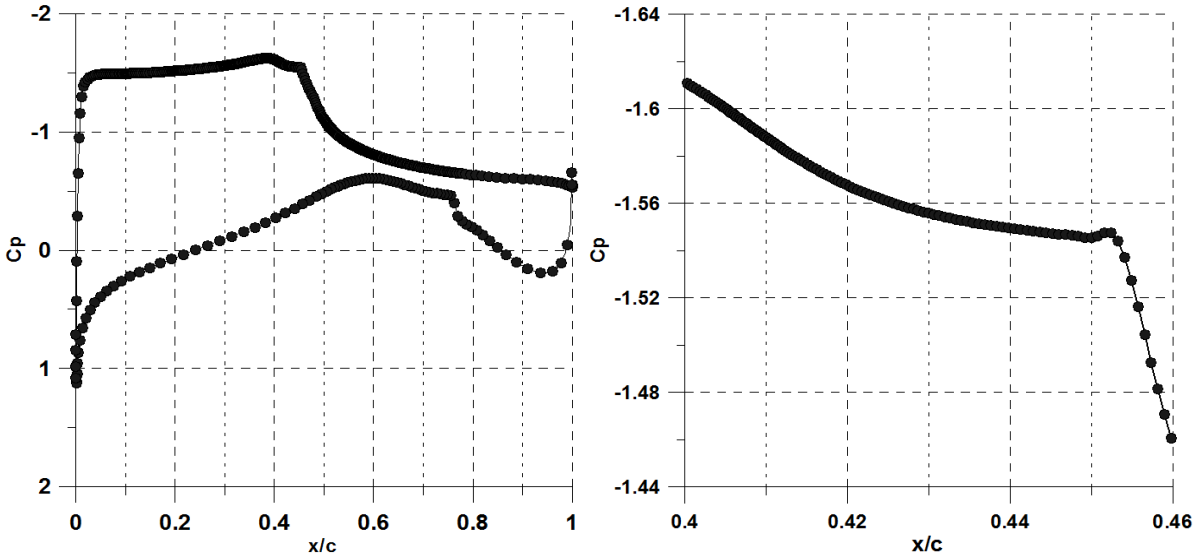


Figure 2. Plots of pressure coefficient over the complete airfoil surface (left) and for the region of plateau in front of the shockwave where laminar-turbulent transition takes place (right). α=7°, Ma=0.70, Re=2.67.

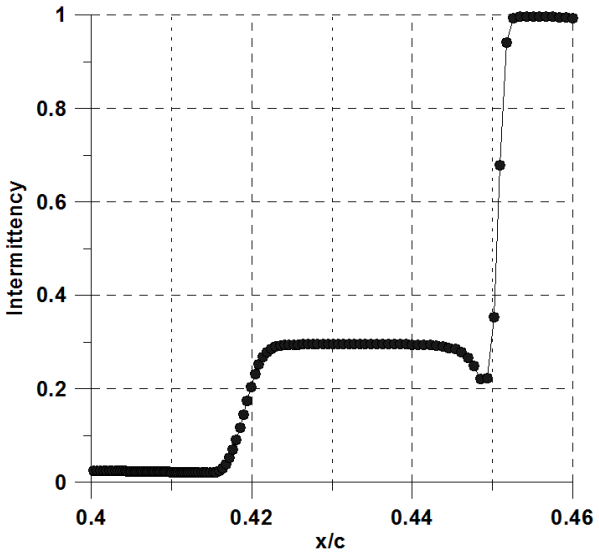


Figure 3. Plot of intermittency at control height of 0.2mm in transition region α=7°, Ma=0.70, Re=2.67.

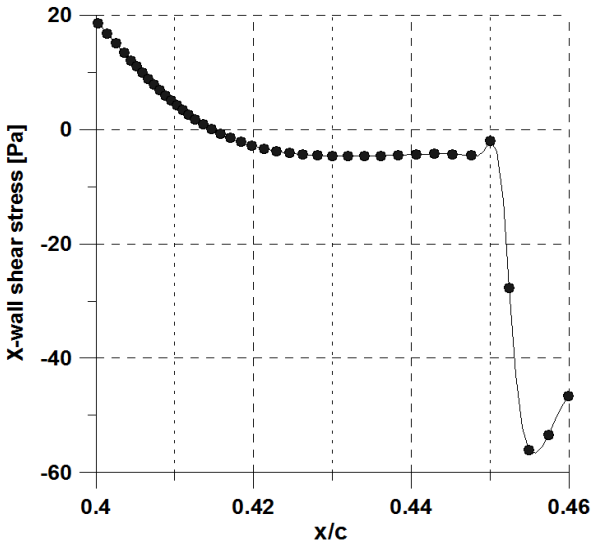


Figure 4. Plot of tangential stress on airfoil surface in transition region α=7°, Ma=0.70, Re=2.67.

In Figure 5 is presented contour of intermittency in boundary layer in the symmetry plane of the computational domain. On the left side of the contour flow in the boundary layer is laminar with intermittency approachning zero. Outside boundary layer intermittency is set to

unity, in order to model flow in wind-tunnel with prescribed turbulence intensity. Start of transition is visible as increase of intermittency at a distance of approximately 0.2mm from the surface, at which height it grows to unity in a distance shown in Figure 3.

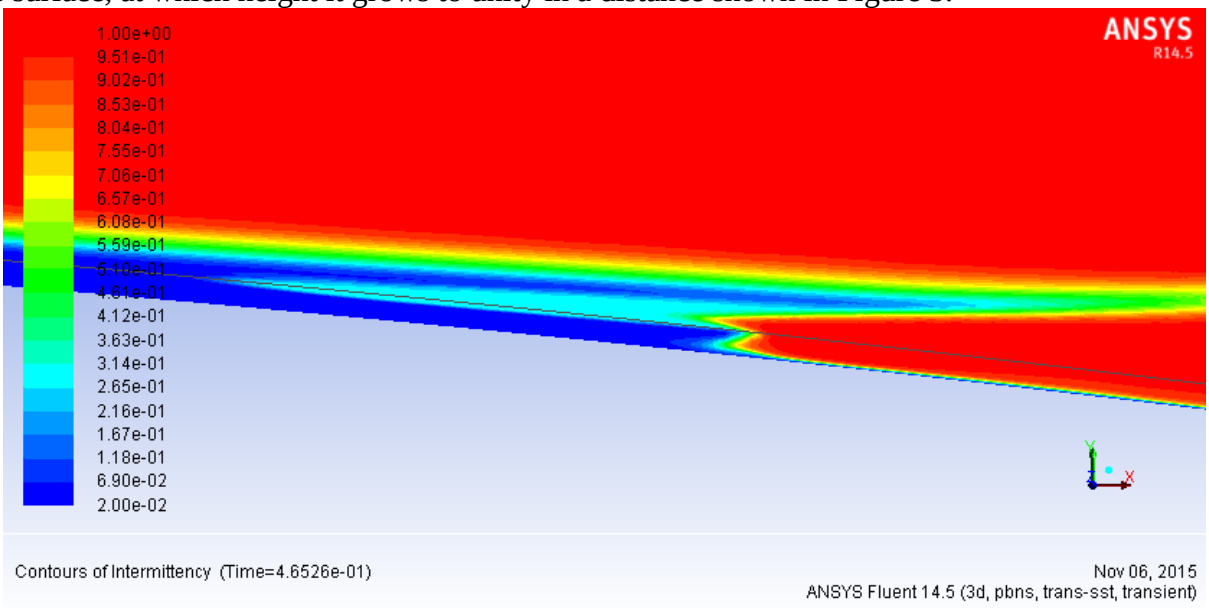


Figure 5. Contour of intermittency in the region of natural laminar-turbulent transition. Black line above airfoil surface represents the control height of 0.2mm applied in Figure 3. α=7°, Ma=0.70, Re=2.67.

4.2 Transition produced by delta-shaped vortex generators

Simulations of forced transition were conducted for two values of angle of attack: 4° and 7°. At an angle of 4° in flow with natural transition there occur low-intensity oscillation of shockwave with Strouhal number of 0.08, characteristic for buffet phenomenon. For this reason the chordwise comparisons of distributions of flow quantities such as pressure coefficient are done for median position of oscillating shockwave. In cases with tripped laminar boundary layer the flow was steady. Tripping of laminar boundary.layer was simulated for height of delta-shaped vortex-generators equal to 0.10mm. Length of the vortex.generator was equal to 1mm. For both heights of vortex generator their alternative locations in 20%c, 30%c, 40%c and 50%c (c standing for airfoil chord) were simulated. As it can be seen in Figure 6 the effects of vortex generators are very local and are visible only in the area of interaction of shockwave with boundary layer. The details of these effects, are shown in greater scale exclusively for this region in Figure 7. It can be seen that moving vortex generators closer to the shockwave extends the region of rising suction, where for clean airfoil the laminar separation formed, creating the pressure plateau shown with black line. On the other hand, for vortex generators located close to the shockwave, at 50%c, in the region of rising pressure a smaller plateau starts to form. This, as it can be seen in the plots of intermittency in Figure 8, may be linked to lower effectiveness of the vortex generators in thicker boundary layer in the central region of the airfoil. In Figure 9 is shown comparison of tangential stress on airfoil surface for the analysed cases. It can be seen that for vortex generators (vgs) located in 50%c there is no rise of tangential stress behind the vortex generator, characteristic for turbulized boundary layer. On the other hand, the increase of suction peak is the greatest of all investigated cases, which suggests that the increase of suction peak is an effect of vorticity added by the vortex generators and not the effect of tripping laminar boundary layer.

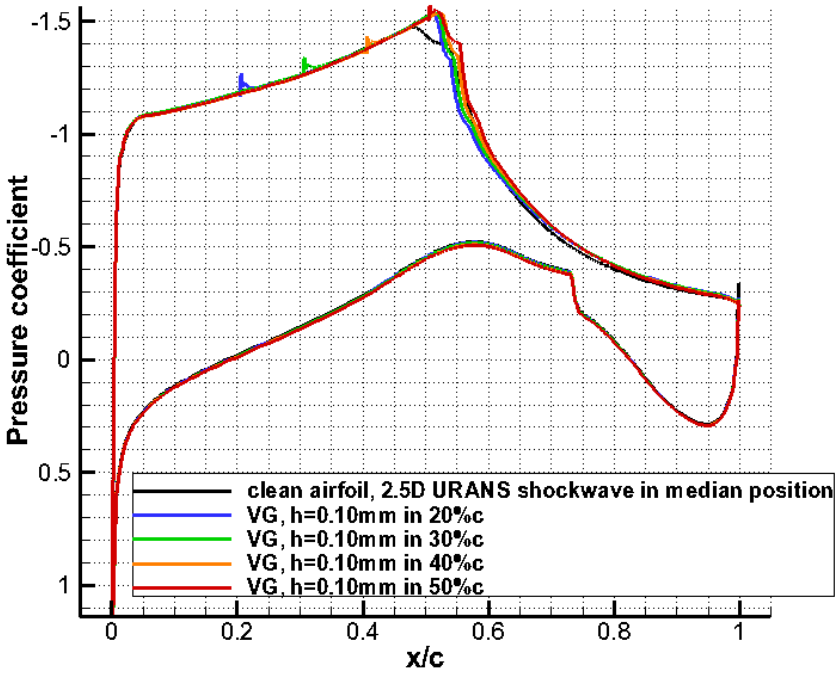


Figure 6. Comparison of distribution of pressure coefficient for clean airfoil case and for cases with simulated arrays of counter-rotating vortex generators 0.10mm-height in different positions of airfoil chord. α=4°, Ma=0.70, Re=2.67.

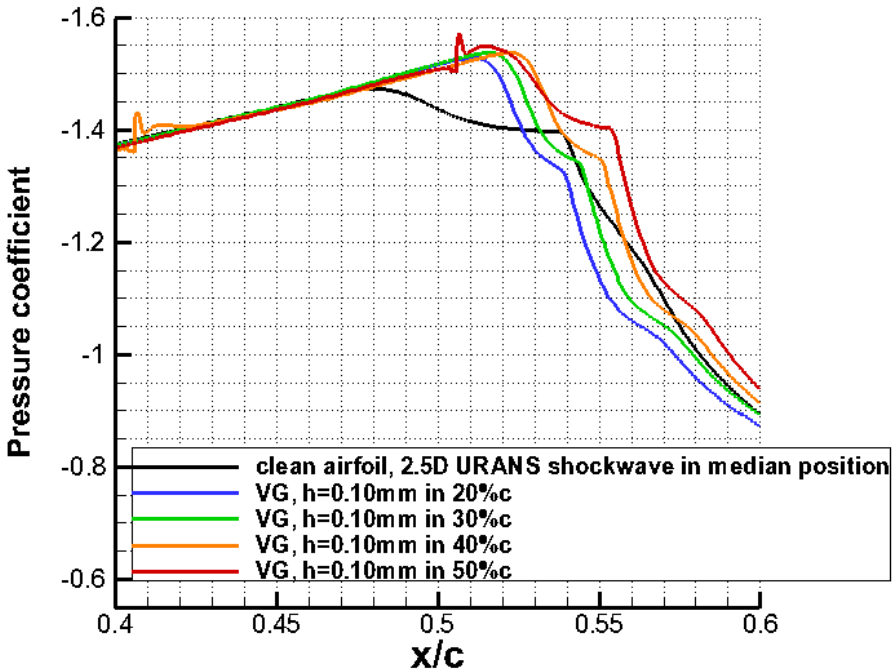


Figure 7. Details of distribution of pressure coefficient for clean airfoil and cases with vortex generators 0.10mm-height in region of shockwave-bundary layer interactions. α=4°, Ma=0.70, Re=2.67.

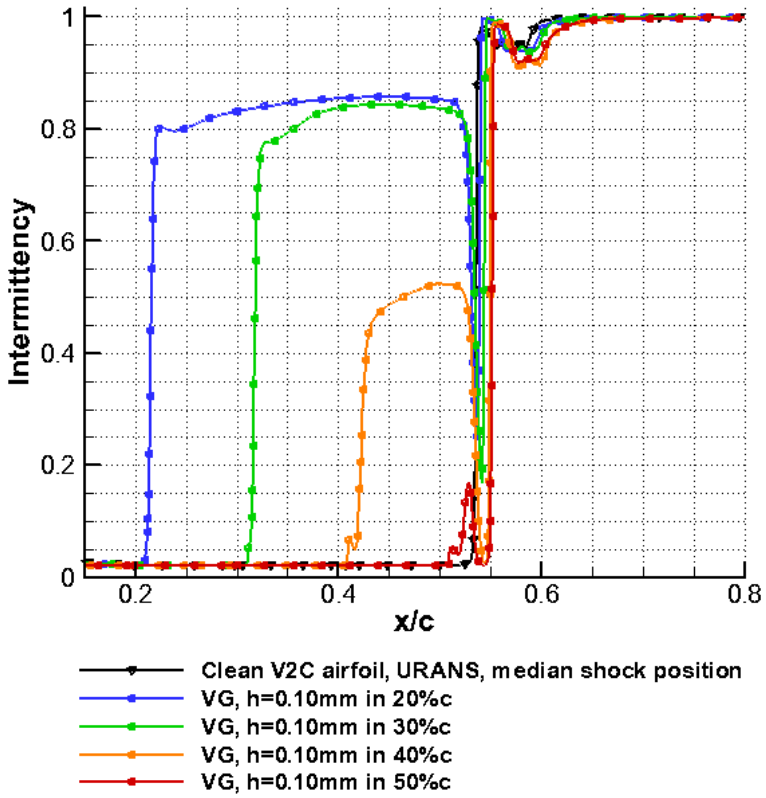


Figure 8. Comparison of distribution of intermittency in the boundary layer for clean airfoil case and for cases with simulated arrays of counter-rotating vortex generators of 0.10mm-height in different positions of airfoil chord. α=4°, Ma=0.70, Re=2.67.

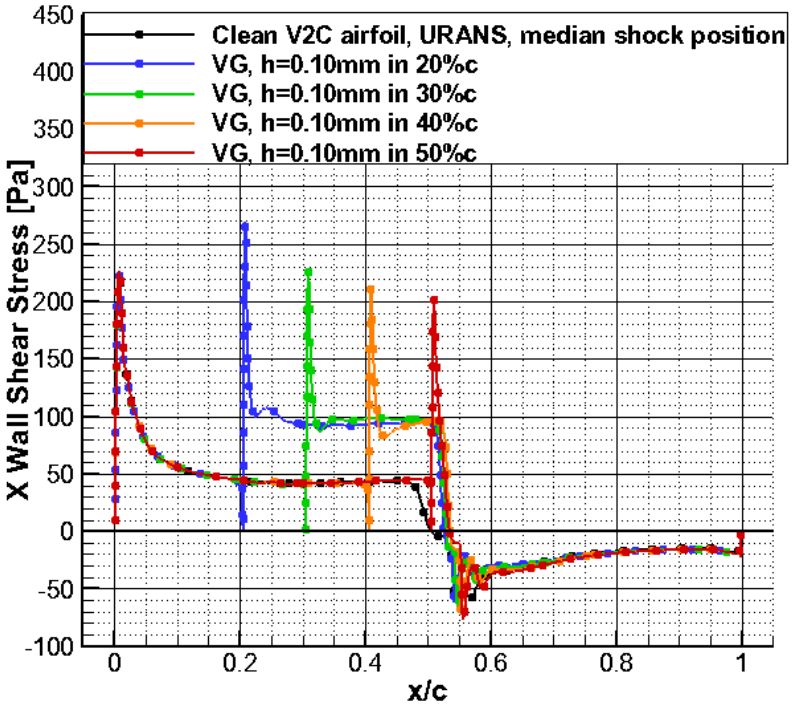


Figure 9. Comparison of distribution of tangential stress on airfoil surface for clean airfoil case and for cases with simulated arrays of counter-rotating vortex generators of 0.10mm-height in different positions of airfoil chord. α=4°, Ma=0.70, Re=2.67.

The most important effects of tripping laminar boundary layer by the delta-shaped vortex generators consist in stabilising flow, which is shown in Figure 10 to Figure 12 as solution converging to a steady-state value with stationary shockwave. Such stabilization of shock-wave was observed also in earlier work [6], where BAY methods was used for modelling of vortex generators. Other effects found at angle of attack $\alpha=4°$ were small increase of drag and reduction of lift-to-drag ratio, shown in Figure 11 and in Figure 12. Changes in lift depend on the chordwise position of the vgs which has influence on shape of suction peak in front of the shockwave. For vgs located in the frontal region of the airfoil, which tend to shift shockwave upstream, there occurs reduction of lift, and for the vgs located closest to the shockwave which create larger suction peak, lift slightly increases over the average value for clean airfoil. Increase of drag is also positive, and, contrary to expectations,

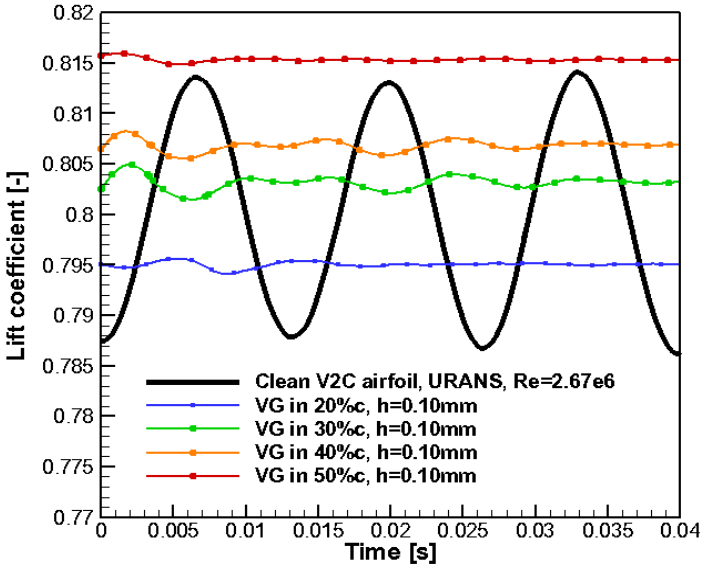


Figure 10. Effects of tripping laminar boundary layer with delta-shaped vortex generators located in 20%c, 30%c, 40%c and 50%c on lift coefficient.

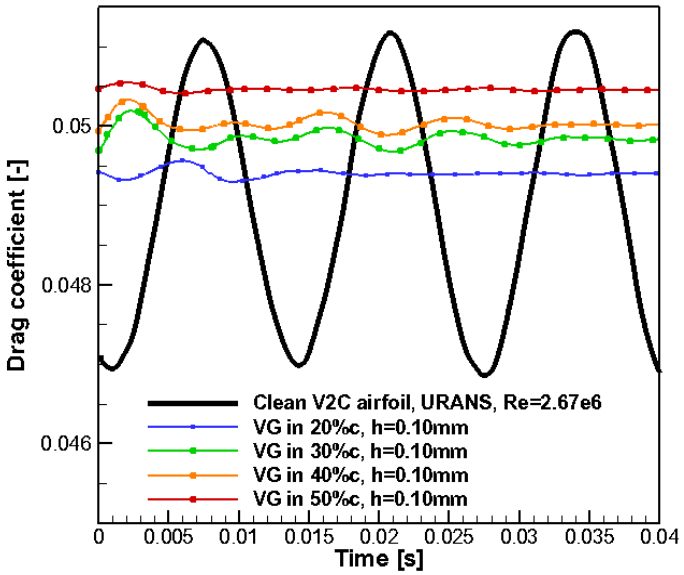


Figure 11. Effects of tripping laminar boundary layer with delta-shaped vortex generators located in 20%c, 30%c, 40%c and 50%c on drag coefficient.

drag increase is the highest for vgs located closest to the shockwave,in spite of shorter length of turbulised region with increased tangential stress.This is due to bild-up of suction peak in location behind 50%c, where negative slope of surface leads to increase of pressure drag. When analysing the changes of lift-to-drag ratio (L/D) due to presence of vortex generators it can be seen that although for all locations of vgs there occurs reduction of L/D with respect to average value for clean configuration, this reduction is the lowest for rearward-placed vortex generators. This is due to increased lift for rearward-placed vgs which compensates rising drag. This indicates that possible gains in aerodynamic characteristics may be expected when laminar boundary layer is tripped in a region close to shockwave and these gains are more likely to occur as increase of lift-to drag ratio than as reduction of drag. Reduction of drag is difficult to achieve because elimination of laminar separation bubble is done at the cost of increase of friction in the newly-formed turbulent-flow region, however short, and particularly difficult when using mechanical turbulators which themselves are an obstacle for the flow and a source of drag.

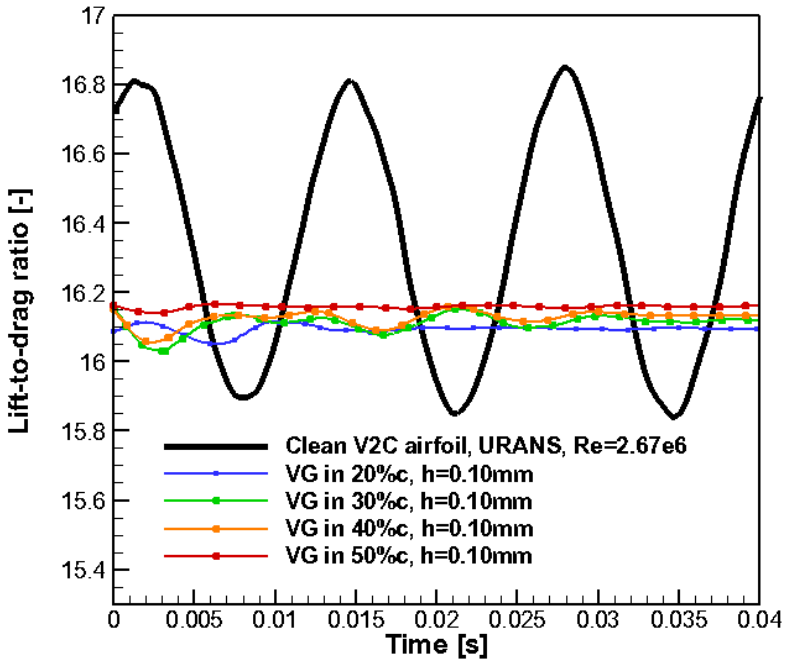


Figure 12. Effects of tripping laminar boundary layer with delta-shaped vortex generators located in 20%c, 30%c, 40%c and 50%c on lift-to-drag ratio.

A confirmation of the previously mentioned hypothesis is the case of tripping laminar boundary layer at an angle of attack of 7 degrees. The physical phenomena created by the vortex generators, as increased suction peak and elimination of separartion bubble at the foot of the shockwave, are very similar to case of flow at an angle of attack of 4 degrees.. An important difference in flow conditions is that at $\alpha=7°$ the shockwave is located closer to leading edge, at approximately 45% chord, in contrast to approximately 53%c at $\alpha=4°$. In these new flow conditions the combination of increased suction, forward shift of shockwave and lower negative surface slope lead to a very low increase of pressure drag due to increased suction on the negative-slope side of airfoil. In effect slight increase of lift-to-drag ratio over the values for clean airfoil was obtained for vortex generator located in 30% chord,, shown in Figure 14.

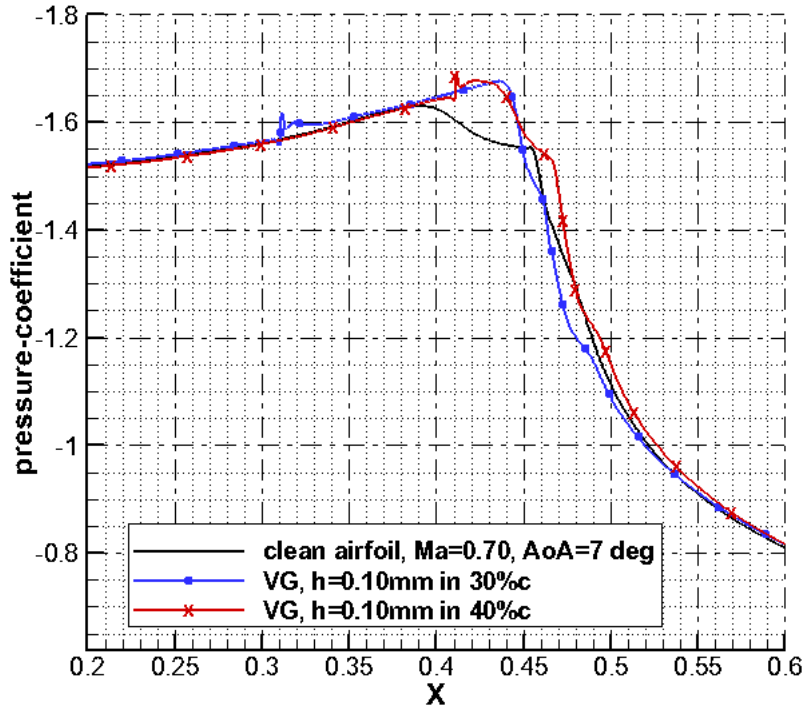


Figure 13. Comparison of distribution of pressure coefficient for clean airfoil case and for cases with simulated arreys of counter-rotating vortex generators 0.10mm-height in different positions of airfoil chord. α=7°, Ma=0.70, Re=2.67.

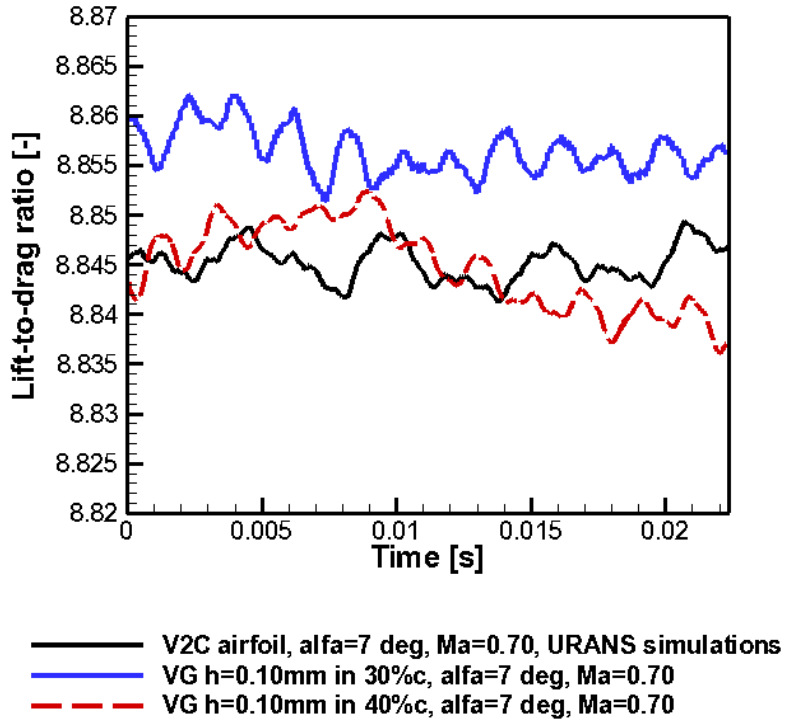


Figure 14.. Effects of tripping laminar boundary layer with delta-shaped vortex generators located in 30%c and 40%c on lift-to-drag ratio compared with results for clean V2C airfoil.

4.3 Transition produced by perpendicular micro-vanes

As it was found that effects of tripping laminar boundary layer include stabilisation of position of shockwave, this particular effect was studied further with another type of vortex generators, consisting of micro vanes, perpendicular to airfoil surface and to the flow, presented already in Figure 1. This choice was based on results of analyses showing that this turbulator was more effective at tripping laminar boundary layer then delta-shaped vgs when turbulator height was being increased [7]. The specific effects of tripping laminar boundary layer with this turbulators on distributions of intermittency in the boundary layer and tangential stress and pressure coefficient on the V2C airfoil are shown in Figure 15 and in Figure 16. In Figure 15 it can also be seen that for properly chosen height of micro-vanes intermittency rises to unity and stays constant until reaching shockwave, while increased positive tangential stress on airfoil surface is an indication of turbulised boundary layer and attached flow behind the turbulator. In Figure 16 it can be seen that in contrast to delta-shaped vortex generators which produce vortices of chordwise-oriented rotation axes, there is no increased suction in the region of pressure plateau on clean airfoil. Instead, the shockwave shifts to the front of airfoil more than for delta-shaped vortex generators and sharper rise of pressure over the shockwave, than for delta shaped vgs takes place. The sharp rise of pressure is, however, characteristic for shockwave appearing in turbulent flow region and was reported by other researchers analysing turbulent boundary layer – shockwave interactions [4]. In this respect the obtained solution is similar to that obtained for delta-shaped vgs. These features of the perpendicular-micro-vanes turbulator were taken advantage of in analysis of its effectiveness in damping buffet oscillations of shockwave. The analysis consisted in continuously increasing of free-stream Mach number from low pre-buffet value and passing through the buffet zone for clean airfoil. The results of this procedure are shown in Figure 17 where changes of moment coefficient are presented and in Figure 18 where changes of lift-to-drag ratio are presented for different values of vane height.

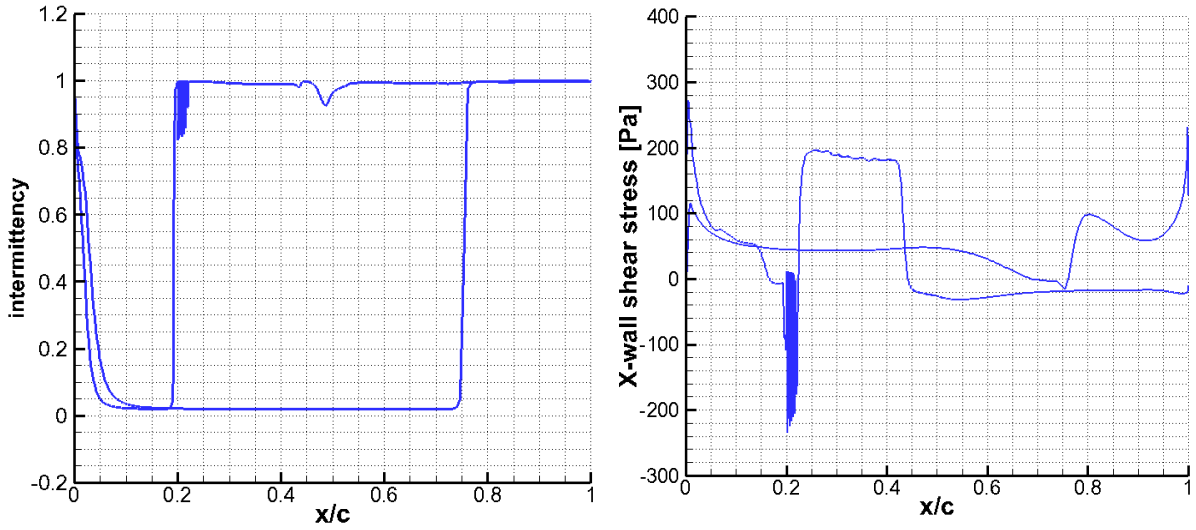


Figure 15. Intermittency at the height of 0.15mm above the airfoil surface for vanes of 0.25mm-height located in 20% chord and tangential stress in X direction on airfoil surface. Ma=0.68, α=6°.

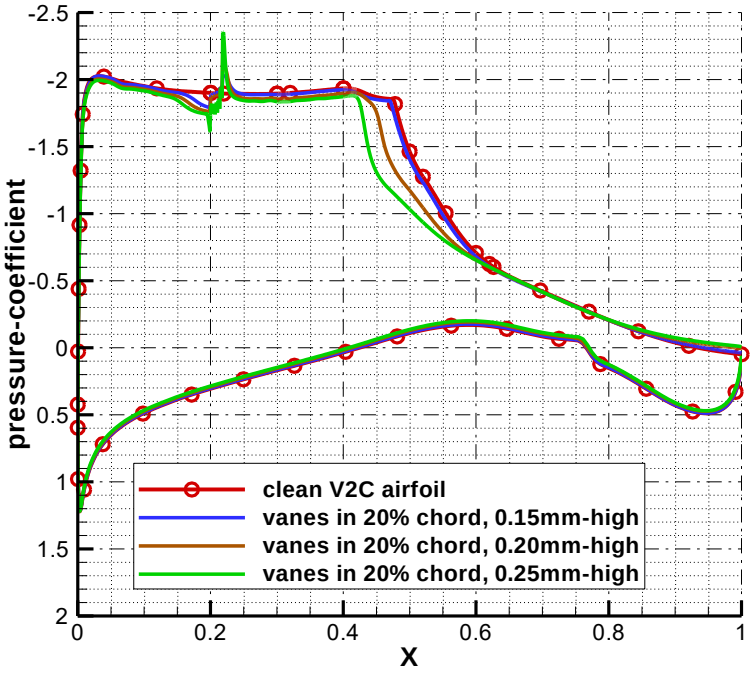


Figure 16. Comparison of distributions of pressure coefficient of clean airfoil and configurations with perpendicular vane turbulators of different height, placed in 20% airfoil chord. Ma=0.681, AoA=6°.

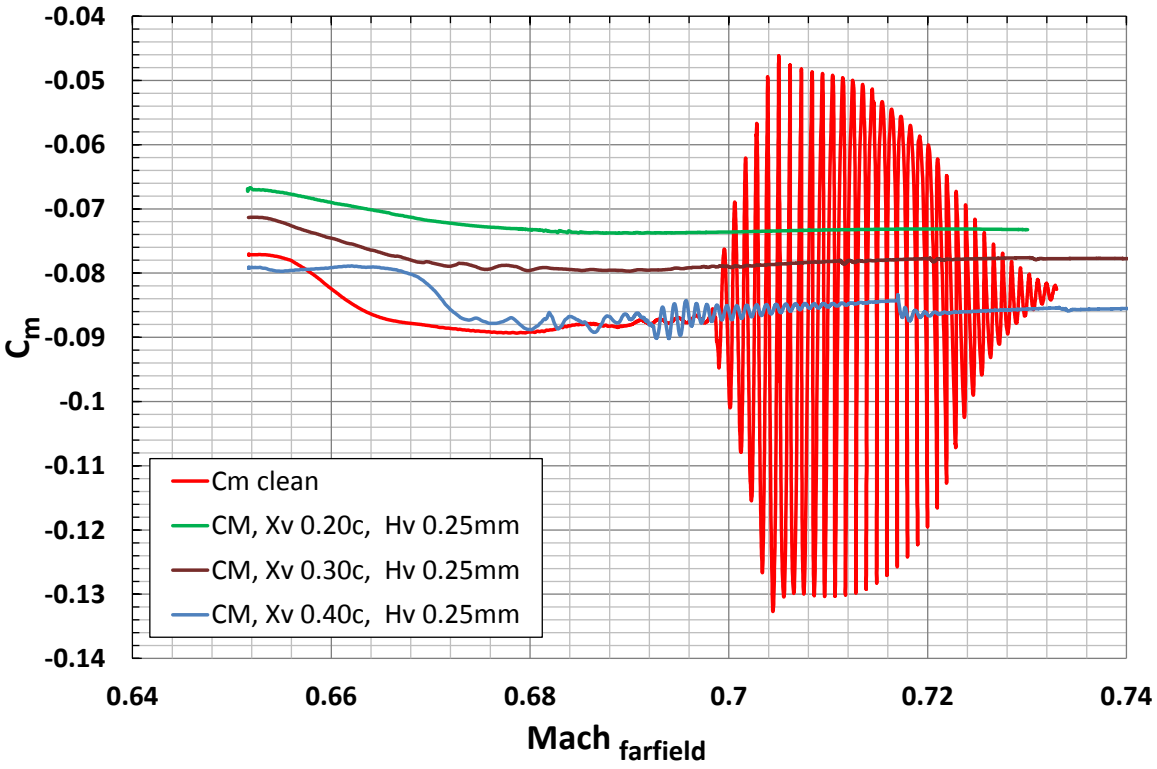


Figure 17. Effects of tripping of laminar boundary layer with perpendicular vane turbulators of different height, placed in 20% airfoil chord on pitching moment oscillations due to buffet phenomenon compared with results for clean V2C airfoil.

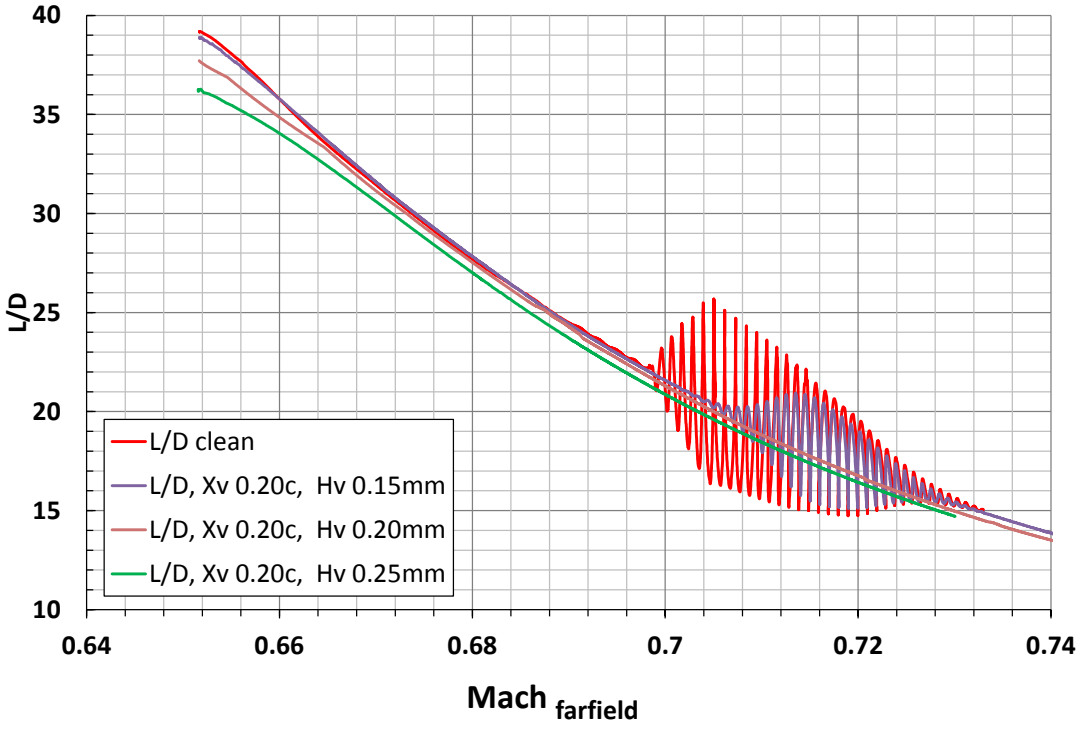


Figure 18. Effects of tripping of laminar boundary layer with perpendicular vane turbulators of different height, placed in 20% airfoil chord, on lift-to-drag ratio in transonic flow regime including buffet zone, compared with results for clean V2C airfoil.

5 CONCLUSIONS

- Improvement of aerodynamic characteristics of laminar airfoil in transonic flow with strong shockwave is more likely to occur as increase of lift-to-drag ratio, than as reduction of drag.
- Conditions when delta-shaped, counter-rotating vortex generators may improve aerodynamic characteristics of laminar airfoil in investigated conditions with strong shockwave include position of the vortex generator of height of 25% - 50% of boundary layer thickness, located approximately 5-10% of chord in front of the shockwave.
- Favourable effects of investigated delta-shaped vortex generators include, beside elimination of laminar separation at the shockwave foot, also increase of suction in front of shockwave which has positive effect on lift-to-drag ratio, compensating increased drag due to higher friction in turbulent boundary layer and due to presence of the vgs themselves.
- Both tested turbulators – the delta-shaped vortex generators and micro-vanes perpendicular to flow and to airfoil surface had positive effects on buffet phenomenon, damping oscillations of shockwave.
- The results obtained for the second type of turbulator – micro-vanes perpendicular to airfoil surface and flow have shown that for vane height equal 50% of thickness of laminar boundary layer buffet phenomenon was eliminated from Mach range where it existed on clean airfoil.

6 ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme **(FP7/2007-2013)** within the project TFAST (Transition Location Effect On Shock Wave Boundary Layer Interaction), under Grant Agreement No. 265455.
The project was also co-financed by the ministry of science of Poland from the funds dedicated to scientific research in 2012-2015, agreement no. **2641/7.PR/12/2013/2** and by Institute of Aviation.
Computational support was obtained from university of Warsaw Interdisciplinary Centre for Mathematical and Computational Modelling, in the computational grant no. G57-2.

REFERENCES

[1] ANSYS FLUENT Theory Guide, Release 14.5, Ansys.Inc, 2012.

[2] J. C. Lin, *Review of research on low-profile vortex generators to control boundary-layer separation*, Progress in Aerospace Sciences 38 (2002), 389–420.

[3] P.R Ashill., J.L Fulker, *A review of flow control research at DERA. Mechanics of Passive and Active Flow Control*, IUTAM Symposium, Goettingen, Germany, September 7–11, 1998, 43–56.

[4] A.L. Braslow, E.C. Knox, *Simplified Method for Determination of Critical Height of Distributed Roughness Particles for Boundary Layer Transition at Mach Numbers From 0 to 5*, National Advisory Comitee for Aeronautics, Technical Note 4363, Washington, September 1958.

[5] F. Grossi, *Physics and modeling of unsteady shock wave/boundary layer interactions over transonic airfoils by numerical simulation*, Ph.D. dissertation, Institut National Polytechnique de Toulouse, 2014.

[6] J.Sznajder, T. Kwiatkowski, *Effects Of Turbulence Induced By Micro Vortex Generators In Shockwave – Boundary Layer Interactions*, Journal of KONES Powertrain and Transport, Vol. 22, No. 2 2015

[7] T. Kwiatkowski, W. Stalewski, K. Surmacz, J. Sznajder, *Numerical simulation of laminar-turbulent transition and investigation of its influence on basic aerodynamic characteristics of laminar airfoil*, TFAST D5.2.3 Deliverable Report, Institute of Aviation, Warsaw, 2016.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ASSESSMENT OF THE FLUID-STRUCTURE INTERACTION CAPABILITIES FOR AERONAUTICAL APPLICATIONS OF THE OPEN-SOURCE SOLVER SU2.

Ruben Sanchez[1*], H. L. Kline[2], David Thomas[3], Anil Variyar[2], Marcello Righi[4], Thomas D. Economon[2], Juan J. Alonso[2], Rafael Palacios[1], Grigorios Dimitriadis[3], Vincent Terrapon[3]

[1]Department of Aeronautics, Imperial College London
London, SW7 2AZ, United Kingdom
e-mail: {r.sanchez-fernandez14, r.palacios}@imperial.ac.uk

[2] Department of Aeronautics & Astronautics, Stanford University
Stanford, CA, 94305, USA
e-mail: {hlkline, anilvar, economon, jjalonso}@stanford.edu

[3] Department of Mechanical and Aerospace Engineering, University of Liege
Liege, 4000, Belgium
e-mail: {dthomas, vincent.terrapon, gdimitriadis}@ulg.ac.be

[4] School of Engineering, Zurich University of Applied Sciences
Winterthur, 8401, Switzerland
e-mail: rigm@zhaw.ch

Keywords: Fluid-Structure Interaction, Computational Aeroelasticity, Reynolds-Averaged Navier Stokes, Geometrically-Nonlinear Solid Mechanics

Abstract. *We report on an international effort to develop an open-source computational environment for high-fidelity fluid-structure interaction analysis. In particular, we will focus on verification of the implementation for application in computational aeroelasticity. The capabilities of the SU2 code for aeroelastic analysis have been further enhanced both by developing natively embedded tools for the study of largely deformable solids, and by wrapping it using Python tools for an improved communication with external solvers. Both capabilities will be demonstrated on relevant test cases, including rigid-airfoil solutions with indicial functions, the Isogai Wing Section, test cases from the AIAA 2nd Aeroelastic Prediction Workshop, and the vortex-induced vibrations of a flexible cantilever in the wake of a square cylinder. Results show very good performance both in terms of accuracy and computational efficiency. The modularity and versatility of the baseline suite allows for a flexible framework for multidisciplinary computational analysis. The software libraries have been freely shared with the community to encourage further engagement in the improvement, validation and further development of this open-source project.*

1 INTRODUCTION

The interaction between fluids and structures on complex geometries is a challenging computational problem that has attracted substantial interest due to its relevance across disciplines, including aeronautical and civil engineering, biomechanics, and turbomachinery. In this work, we will focus on aeronautical applications, and particularly on computational aeroelasticity [1–3]. Due to the complexity of the constitutive equations and the strongly non-linear coupling effects between fluid and structure, the development of accurate and efficient solvers able to be run in highly-parallel environments becomes a determining factor, specially when addressing large problems that require refined discretisations.

With this in mind, the open-source software suite SU2 [4–6] has been used as the basic infrastructure for the development of a set of tools for Fluid-Structure Interaction (FSI). Due to its modularity and versatility, it has been possible to develop natively embedded tools for non-linear structural analysis in SU2, thus allowing for a self-contained, open-source suite capable to deal with FSI problems involving largely deformable solids [7]. At the same time, the suite has been wrapped using Python tools for improved communication with external solvers, which increases the flexibility for the user when aiming to solve problems using other in-house or commercial platforms. In this paper we will detail the modular implementations we have included in SU2 in the context of FSI, and we will assess the validity of these methods.

The paper is organized as follows. Section 2 briefly describes the governing equations for fluid and structural mechanics, and establishes the interface coupling conditions. Section 3 sketches the structure of the code for the fluid and structural solver, and the architecture that permits the coupling of the independent solvers with each other or with external tools via Python. Section 4 presents the first set of results obtained using the different approaches described. Finally, section 5 summarizes the current state of the project and outlines some of the planned future developments.

2 BACKGROUND

2.1 Fluid mechanics

2.1.1 Governing equations

We are concerned with compressible, turbulent fluid flows governed by the Reynolds-averaged Navier-Stokes (RANS) equations, which can be expressed in arbitrary Lagrangian-Eulerian (ALE) [8] differential form as

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{F}^c - \nabla \cdot \mathbf{F}^v - Q = 0 \qquad \text{in } \Omega \times [0, t] \tag{1}$$

where the conservative variables are given by $U = \{\rho, \rho\mathbf{v}, \rho E\}^\mathsf{T}$, and the convective fluxes, viscous fluxes, and generic source term are

$$\mathbf{F}^c = \begin{Bmatrix} \rho(\mathbf{v} - \dot{\mathbf{u}}_\Omega) \\ \rho\mathbf{v} \otimes (\mathbf{v} - \dot{\mathbf{u}}_\Omega) + \bar{\bar{I}}p \\ \rho E(\mathbf{v} - \dot{\mathbf{u}}_\Omega) + p\mathbf{v} \end{Bmatrix}, \quad \mathbf{F}^v = \begin{Bmatrix} \cdot \\ \bar{\bar{\tau}} \\ \bar{\bar{\tau}} \cdot \mathbf{v} + \mu^*_{tot} c_p \nabla T \end{Bmatrix}, \tag{2}$$

and $Q = \{q_\rho, \mathbf{q}_{\rho\mathbf{v}}, q_{\rho E}\}^\mathsf{T}$, where ρ is the fluid density, $\mathbf{v} = \{v_1, v_2, v_3\}^\mathsf{T} \in \mathbb{R}^3$ is the flow speed in a Cartesian system of reference, $\dot{\mathbf{u}}_\Omega$ is the velocity vector at a point for a domain in motion (mesh velocity after discretization), E is the total energy per unit mass, p is the static pressure,

c_p is the specific heat at constant pressure, T is the temperature, and the viscous stress tensor can be written in vector notation as

$$\bar{\bar{\tau}} = \mu_{tot}\left(\nabla\mathbf{v} + \nabla\mathbf{v}^{\mathsf{T}} - \frac{2}{3}\bar{\bar{I}}(\nabla\cdot\mathbf{v})\right). \tag{3}$$

In 3D, Eqn. 1 is a set of five coupled, nonlinear partial differential equations (PDEs) that are statements of mass (1), momentum (3), and energy (1) conservation in a fluid. Additional boundary and temporal conditions will be required and are problem dependent. Most commonly, no-slip conditions ($\mathbf{v} = \dot{\mathbf{u}}_\Omega$) are applied to solid surfaces, and far-field approximations that mimic the fluid behavior at infinity [9] are applied to outer boundaries.

Assuming a perfect gas with a ratio of specific heats γ and gas constant R, one can determine the pressure from $p = (\gamma-1)\rho\left[E - \frac{1}{2}(\mathbf{v}\cdot\mathbf{v})\right]$, the temperature is given by $T = p/(\rho R)$, and $c_p = \gamma R/(\gamma-1)$. In accord with the standard approach to turbulence modeling based upon the Boussinesq hypothesis [10], which states that the effect of turbulence can be represented as an increased viscosity, the total viscosity is divided into laminar and turbulent components, or μ_{dyn} and μ_{tur}, respectively. In order to close the system of equations, the dynamic viscosity μ_{dyn} is assumed to satisfy Sutherland's law [11] (a function of temperature alone), the turbulent viscosity μ_{tur} is computed via a turbulence model, and

$$\mu_{tot} = \mu_{dyn} + \mu_{tur}, \quad \mu^*_{tot} = \frac{\mu_{dyn}}{Pr_d} + \frac{\mu_{tur}}{Pr_t}, \tag{4}$$

where Pr_d and Pr_t are the dynamic and turbulent Prandtl numbers, respectively.

The turbulent viscosity μ_{tur} is obtained from a suitable turbulence model. The Shear Stress Transport (SST) model of Menter [12] and the Spalart-Allmaras (S-A) [13] model are two of the most common and widely used turbulence models. For treating purely laminar or inviscid problems, note that the laminar Navier-Stokes equations and the Euler equations can be recovered from Eqn. 1 by removing the contribution of the turbulence model to the viscosity and by eliminating all viscous terms, respectively, when appropriate. With the Euler equations, flow tangency boundary conditions are applied to solid surfaces in the place of a no-slip condition.

2.1.2 Numerical implementation

The governing fluid equations are spatially discretized on unstructured meshes via the Finite Volume Method (FVM) using a median-dual, vertex-based scheme with a standard edge-based structure. Instances of the state vector U are stored at the nodes of the primal mesh, and the dual mesh is constructed by connecting the primal cell centroids, face centroids, and edge midpoints surrounding a particular node.

Convective fluxes are discretized using either a centered scheme, such as the Jameson-Schmidt-Turkel (JST) [14] method, or an upwind scheme, such as the Roe [15] method. For upwind methods, second-order accuracy is easily achieved via reconstruction of variables on the cell interfaces by using a MUSCL approach with gradient limiters. The convection of the turbulence variables is discretized using an upwind scheme (typically first-order). Viscous fluxes are computed with the node-gradient-based approach due to Weiss et al. [16]. The Green-Gauss or weighted least-squares methods are available for approximating the spatial gradients of the flow variables. Source terms are approximated via piece-wise reconstruction in the finite-volume cells.

For unsteady flows, a dual time-stepping strategy [17, 18] has been implemented for obtaining a specified accuracy in time. In this method, the unsteady problem is transformed into a

series of steady problems at each physical time step that can then be solved using all of the well-known convergence acceleration techniques for steady problems. Each physical time step is relaxed in pseudo time using implicit integration.

During time-accurate calculations on dynamic meshes, the coordinates and velocities of the grid nodes must be updated at each physical time step using suitable methods for displacing the nodes of the volume mesh and computing the resulting grid node velocities. Two typical strategies involve rigid mesh transformations or dynamically deforming meshes. In the former, the grid node coordinates and velocities are often computed analytically based on a prescribed motion of the mesh as a rigid body. The dynamically deforming case, which is typically necessary for FSI problems, requires an additional technique to deform the fluid volume mesh to conform to specified changes in the positions of solid boundaries. After updating the position of all nodes in the fluid mesh, the local grid velocity at a node can be computed by storing the node coordinates at prior time instances and using a finite differencing approximation that is consistent with the chosen dual time-stepping scheme.

Finally, when computing unsteady flows on dynamic meshes with the ALE form of the equations, a Geometric Conservation Law (GCL) should be satisfied. First introduced by Thomas and Lombard [19], it has been shown mathematically and through numerical experiment [20–22] that satisfying the GCL can improve the accuracy and stability of the chosen scheme. A straightforward technique for the numerical implementation of the GCL [23, 24] has been included as part of the dual-time stepping approach.

2.2 Structural mechanics

2.2.1 Rigid body models

The following model can be considered for the constrained displacement of non-deformable bodies. The kinematics can be described by a finite set of degrees of freedom corresponding to a translation of body-attached point O and a rotation around this point. This is illustrated in Fig. 1 where both an absolute frame of reference and a relative frame of reference attached to the solid body (denoted by the ' symbol) are considered. Initially, these two referential frames are superposed. After a displacement, the absolute position $\mathbf{s}$ of any point P within the solid can thus be written as [25]:

$$\mathbf{s}_\mathrm{P} = \mathbf{s}_\mathrm{O} + \mathbf{r}' = \mathbf{s}_\mathrm{O} + \mathbf{R}\mathbf{r} \tag{5}$$

where $\mathbf{s}_\mathrm{O}$ is the translation of the reference point, $\mathbf{r}'$ is the relative position of P in the body-attached frame of reference, $\mathbf{r}$ is the position of P before the displacement and $\mathbf{R}$ is a rotation matrix.

The rotation matrix is based on the choice of the parametrization for the rotation and describes the rotation of the relative frame of reference O' in the absolute referential frame. In the case of a three-dimensional rotation, it can be separated into ordered rotations of angles θ, ϕ and ψ around the x, y and z-axis respectively, which gives the following rotation matrix:

$$\mathbf{R}(\theta,\phi,\psi) = \begin{bmatrix} \cos\phi\cos\psi & \sin\theta\sin\phi\cos\psi - \cos\theta\sin\psi & \cos\theta\sin\phi\cos\psi + \sin\theta\sin\psi \\ \cos\phi\sin\psi & \sin\theta\sin\phi\sin\psi + \cos\theta\cos\psi & \cos\theta\sin\phi\sin\psi - \sin\theta\cos\psi \\ -\sin\phi & \sin\theta\cos\phi & \cos\theta\cos\phi \end{bmatrix}$$

In the case where the rigid body displacement is dynamically constrained by some constant stiffness or damping, Lagrange's equation [26, 27] can be used to build the equations of motion

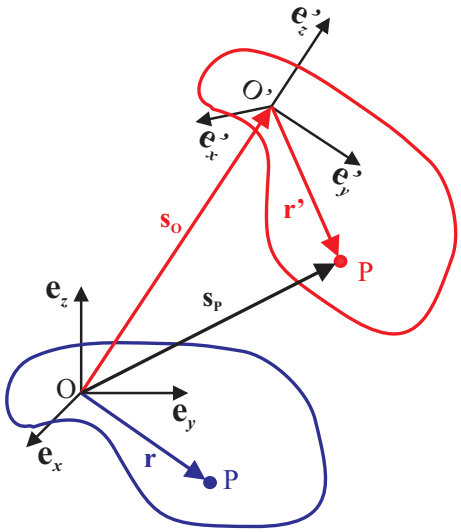


Figure 1: Vectorial description of a rigid body displacement, before (blue) and after (red) the displacement.

for an arbitrary number n (from one to six) of degrees of freedom u_j:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{u}_j}\right) - \left(\frac{\partial T}{\partial u_j}\right) + \left(\frac{\partial V}{\partial u_j}\right) + \left(\frac{\partial D}{\partial \dot{u}_j}\right) = F_j(t) \qquad j = 1, \cdots, n \tag{6}$$

where T is the kinetic energy related to the mass and the inertia, V the potential energy related to the stiffness, D is the dissipation function related to the damping and F_j is the generalized time-dependent force associated to the j^{th} degree of freedom. This can be expressed in matrix form as:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{F}(t) \tag{7}$$

where $\mathbf{u}$ is the vector containing the u_j.

2.2.2 Solid mechanics

In the context of geometrically non-linear, solid mechanics problems, the governing equation may be written [28] as

$$\rho_s \frac{\partial^2 \mathbf{u}}{\partial t^2} = \nabla(\mathbf{F} \cdot \mathbf{S}) + \rho_s \mathbf{f} \quad \text{in } \Omega \times [0, t] \tag{8}$$

where ρ_s is the structural density, $\mathbf{u}$ are the displacements of the solid, $\mathbf{F}$ the material deformation gradient, $\mathbf{f}$ the volume forces on the structural domain, and $\mathbf{S}$ the second Piola-Kirchhoff stress tensor, which are related to the Cauchy Stress tensor $\boldsymbol{\sigma}$ by means of a Piola transformation [29]. The mechanical response of the continuum is addressed by means of a Lagrangian (*spatial*) description, where the displacements, velocities and accelerations of the structure at the nodes are tracked by following material particles [29, 30].

The current implementation of a solid structural solver within SU2 accounts both for linear elastic problems, and also for geometrically non-linear problems with a hyperelastic, Neo-Hookean material model. In a linear setting, and assuming no structural damping, Eq. 8 may be discretized in space using the Finite Element Method (FEM), resulting in

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}_L\mathbf{u} = \mathbf{F}(t) \tag{9}$$

where $\mathbf{F}(t)$ is the vector of externally applied forces, $\mathbf{M}$ the mass matrix and $\mathbf{K}_L$ the linear stiffness matrix, which multiply, respectively, the acceleration and displacement terms of the solution [31]. In a non-linear setting, on the other hand, we linearise the principle of virtual work and reformulate the problem in an incremental, implicit Newton-Raphson way [29,31], as

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{u}}_{t+1} + \mathbf{K}_{NL}\Delta\mathbf{u}^{(k} &= -\mathbf{R}_{t+1} \\ \mathbf{u}_{t+1}^{(k} &= \mathbf{u}_{t+1}^{(k-1} + \Delta\mathbf{u}^{(k} \\ \mathbf{R}_{t+1} &= \mathbf{T}_{t+1}^{(k-1} - \mathbf{F}_{t+1} \end{aligned} \tag{10}$$

where the tangent matrix $\mathbf{K}_{NL}$ accounts for the constitutive and stress terms of the equations and $\mathbf{R}$ is the residual vector which is computed as the difference between internal equivalent nodal forces $\mathbf{T}$ and the external forces $\mathbf{F}$. In order to obtain time-accurate solutions on solid structural analysis, both the Newmark [32] and the Generalized-α [33] integration methods have been implemented in SU2.

2.3 Coupling

2.3.1 Coupling conditions

It is necessary to impose appropriate boundary conditions on the interface in order to fully define the coupled problem. First of all, continuity of displacements may be imposed over the Fluid-Structure interface Γ as

$$\mathbf{u}_{\Gamma_f} = \mathbf{u}_{\Gamma_s} = \mathbf{u}_{\Gamma}. \tag{11}$$

Second, assuming a viscous flow with no-slip boundary conditions, the tractions on the fluid interface are defined as $\mathbf{t}_f = -p\mathbf{n}_f + \bar{\bar{\boldsymbol{\tau}}}_f\mathbf{n}_f$, where p is the pressure of the fluid on the interface, $\bar{\bar{\boldsymbol{\tau}}}$ the viscous stress tensor and $\mathbf{n}_f$ the dimensional, fluid normal outwards from the surface. On the other hand, the tractions on the structural interface are defined as $\mathbf{t}_{s,n} = \bar{\bar{\boldsymbol{\sigma}}}_s\mathbf{n}_s$, where $\mathbf{n}_s$ is the dimensional, structural normal pointing away from the surface.

We can now define a Dirichlet-to-Neumann [34, 35] non-linear operator that maps the displacements on the fluid interface into the fluid tractions, $\mathbf{t}_{\Gamma_f} = \mathscr{F}(\mathbf{u}_{\Gamma_f})$ on Γ_f. Analogously for the structural side, $\mathbf{t}_{\Gamma_s} = \mathscr{S}(\mathbf{u}_{\Gamma_s})$ on Γ_s. Imposing equilibrium of tractions, and combining it with the continuity condition in Eqn. (11), we obtain a Steklov-Poincaré equation [34, 35]

$$\mathscr{F}(\mathbf{u}_{\boldsymbol{\Gamma}}) + \mathscr{S}(\mathbf{u}_{\boldsymbol{\Gamma}}) = 0. \tag{12}$$

This equation can be rewritten in the form of a fixed-point equation [35] using an inverse Neumann-to-Dirichlet operator on the structural side, $\mathbf{u}_{\Gamma_s} = \mathscr{S}^{-1}(\boldsymbol{\lambda}_{\Gamma_s})$, resulting in the interface condition

$$\mathscr{S}^{-1}(-\mathscr{F}(\mathbf{u}_{\boldsymbol{\Gamma}})) = \mathbf{u}_{\boldsymbol{\Gamma}}. \tag{13}$$

2.3.2 Time coupling

A partitioned approach has been adopted in this work, which permits the employment of an adequate solver for each subproblem [35] while maintaining the modularity of SU2. Two possible strategies are available for the integration in time of the coupled problem: an explicit, loosely-coupled method, and an implicit, strongly-coupled method. The choice of the scheme will depend on the particularities of the problem to be solved.

The explicit procedure runs sequentially one single solution of the the fluid and structure solvers and advances the solver in time without enforcing the coupling conditions at the end of

each time step. Due to its simple and modular implementation, this is widely used in computational aeroelasticity, where the coupled dynamics are often mostly determined by the vibrations characteristics of the structure [3, 36–38]. However, this scheme may be unstable or inaccurate in certain cases, due to its high dependency on the density ratio ρ_s/ρ_f and the compressibility of the flow [35, 39, 40].

An implicit scheme improves the quality of the solution for cases in which explicit algorithms are inadequate. Implicit schemes require both solvers to meet the coupling conditions defined in section 2.3.1 at the end of each time step. In particular, our implementation uses the Block Gauss-Seidel (BGS) method, which iteratively solves the fluid and structural problems within the time step until a tolerance criterion is met in the continuity condition (11).

However, low convergence and even divergence has been reported when density ratios are low, the flow is incompressible or the structural deformations are large [40–42]. One common strategy to improve the convergence of the scheme is the use of high-order displacement predictors combined with the employment of relaxation techniques [35,40,43,44]. The definition of a fixed relaxation parameter ω generally results on a large number of iterations [43]; however, the use of the Aitken's Δ^2 dynamic relaxation parameter [45], which has also been included in SU2, is a simple and efficient option [40, 43, 44] to improve the convergence of implicit schemes.

2.3.3 Spatial coupling

The computational study of a Fluid-Structure Interaction problem using a partitioned approach requires the transfer of the coupling conditions defined in 2.3.1 between the fluid and structural discretizations in an appropriate way. Due to differences in the physical behavior and the possibility of local effects on one, or both, of the domains, matching discretizations will often either over- or under-refine one of the domains. This motivates the implementation of interpolation between non-matching discretizations. It should be noted that the interpolation of fields on the non-matching interface may compromise the accuracy of the whole simulation [46], and several authors have addressed the problem of coupling two non-matching discretizations [1, 46–51]. The general criteria for choosing the appropriate interpolator relates to its efficiency, robustness, accuracy, and conservation of momentum and energy [1, 46, 49, 52].

It is a common approach in the literature to apply the principle of virtual work [1, 46, 49, 53] to define a conservative transfer scheme. Brown [52] and Farhat [1] have introduced projection methods that apply the conservation of virtual work to the problem of transferring load and deformations in FSI problems on non-matching meshes. By requiring the interpolation method to conserve virtual work across the displacements and forces on the two meshes, the resulting distributions are physically consistent and result in the same integrated forces. Following their work an expression for the virtual work performed by the forces applied to the structure $\mathbf{F}$, moved through a virtual nodal displacement $\delta\mathbf{u}_{\Gamma,s}$ is:

$$\delta W_E = \mathbf{F}^T \cdot \delta\mathbf{u}_{\Gamma,s} \tag{14}$$

The virtual work must equal the similar expression using the forces of the fluid solution and the virtual displacements of the fluid mesh:

$$\delta W_E = \int_{\Gamma,f} (-p\vec{n}_f + \bar{\bar{\boldsymbol{\tau}}}_f \vec{n}_f)^T \cdot \delta\mathbf{u}_{\Gamma,f} dS_f \tag{15}$$

where $\vec{n}_f$ is the unit normal on the fluid side of the problem and S_f the curvilinear coordinate

along the fluid interface. Equating Equations (14) and (15):

$$\mathbf{F}^T \cdot \delta\mathbf{u}_{\Gamma,s} = \int_{\Gamma,f} (-p\vec{n}_f + \bar{\bar{\boldsymbol{\tau}}}_f \vec{n}_f)^T \cdot \delta\mathbf{u}_{\Gamma,f} dS_f \tag{16}$$

In the discrete interface, the displacements on the fluid side may be mapped using the displacements on the structural side through the use of a matrix $\mathbf{H}$ which will need to be defined for each problem and method used, thus resulting in $\mathbf{u}_f = \mathbf{H}\mathbf{u}_s$. For the common case of a finer fluid mesh we have a set of distributed forces $\boldsymbol{\lambda}_f$ including both surface pressure and traction. According to the previously described principle of virtual work, the tractions at the structural interface may be mapped from the fluid domain as $\boldsymbol{\lambda}_s = \mathbf{H}^T\boldsymbol{\lambda}_f$. Therefore, it is possible to define a full space-coupling scheme that would meet the criteria of conservation of work by adequately defining a transformation matrix $\mathbf{H}$. Global conservation of forces can also be achieved by imposing that the rowsum of $\mathbf{H}$ is equal to one [49].

3 IMPLEMENTATION

The SU2 software suite was conceived as a common infrastructure for solving multi-physics PDE-based problems on unstructured meshes. The full suite is composed of compiled C++ executables and high-level Python scripts that perform a wide range of tasks related to PDE analysis, PDE-constrained optimization, or coupled multi-physics problems. Below, we will describe some of the key implementation details for the various FSI approaches for both C++ and Python. A more complete description of the suite can be found in Economon et al. [54].

3.1 Fluid solver

Much of the C++ class design in SU2 is shared by all of the core modules. In particular, this includes the geometry (i.e., grids), integration, configuration (input file), and output class structures. Only specific numerical methods for the convective, viscous, and source terms are re-implemented for different physical models where necessary. There is no fundamental limitation on the number of state variables or governing equations that can be solved simultaneously in a coupled or segregated way (other than the physical memory available on a given computer architecture), and the more complicated algorithms and numerical methods (including parallelization, multigrid, and linear solvers) have been implemented in such a way that they can be applied without special consideration during the implementation of a new physical model.

The schematic in Fig. 2 shows the structure of the fluid solver in SU2. A similar structure would arise for other single-physics solvers, as will be seen below for the structural solver. The main SU2 loop drives the iterations of the particular physics, and in this example, the iteration class for the flow problem is executed. Within that iteration, some classes that are shared among all solvers are leveraged, such as the geometry and integration classes, while others have been specialized to the flow problem, such as the solver, numerics, and variable classes. The solver, numerics, and variable classes contain the routines that are specific to a chosen physical model.

3.2 Structural solver

The structural solver has been designed based on the original fluid solver, and follows its main structure, as shown in Fig. 3. However, some modifications have been carried out in order to account for the specific characteristics of solid mechanics problems. In particular, two layers of abstraction have been included that deal, independently, with geometrical and material nonlinear effects. The Finite Element analysis is performed at the solver level and

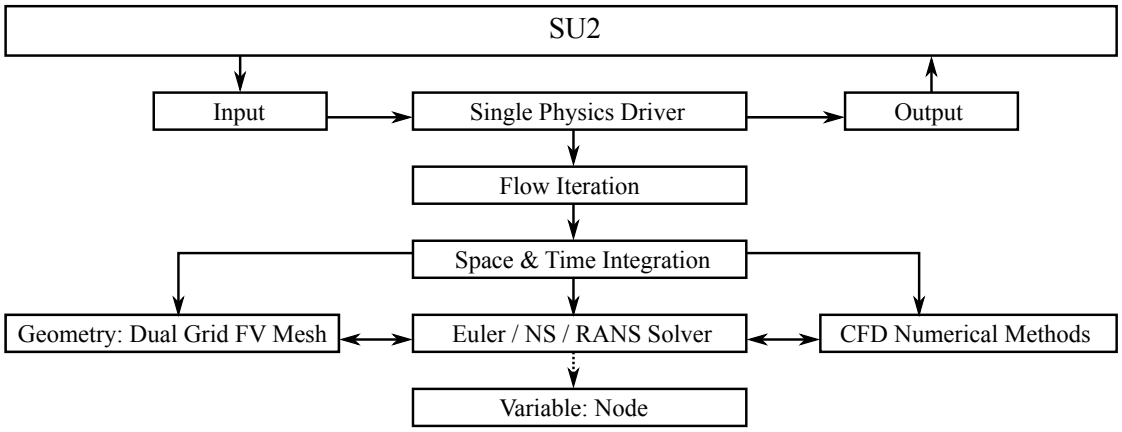


Figure 2: Structure of the fluid solver.

relies on a C++ class common for any kind of element. This structure has been designed with the philosophy of maintaining the flexibility of the code while easing and encouraging further contributions. Further details may be found in Sanchez *et al.* [7].

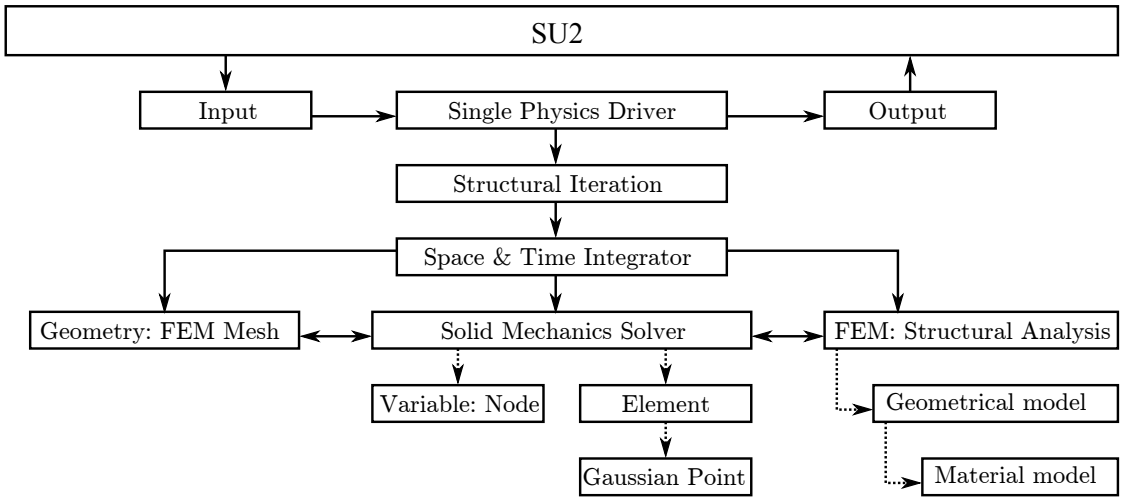


Figure 3: Structure of the Solid Mechanics solver.

3.3 Multizone solver

The solution process in SU2 has been redesigned to improve its ability to run several solvers with a single instance of the software. With this objective, a so-called *driver* structure has been introduced, in which the developer can make use of the different parts of the code as if they were black-boxes. When different solvers are used and need to interact with each other, it is necessary to introduce mechanisms to communicate information at their interface. This is a

complex task, specially when the solution process is run in multiple processors.

This situation is illustrated in Fig. 4 for an example in which a solid wall is immersed in a flow in a channel and run in two processors. A multi-core, multi-physics problem run on a distributed memory computer requires the MPI communication of information to be done not only within each independent solver, but also across the different solvers. Moreover, an optimal partition of the discretization of each zone of the physical domain, which reduces to a minimum the number of edges cut by the partitions, may result in non-compliant regions across zones. This fact forces the transfer mechanisms to be able to efficiently distribute the information from processor *i* in the *fluid* domain to processor *j* in the *structural* domain. This issue has been resolved by abstracting the physics of the information to be transferred from the MPI transfer routines, thus allowing for different and specialized teams to develop and improve each particular part of the problem independently. More specifically, and for Fluid-Structure Interaction applications, the resulting code structure is shown in Fig. 5.

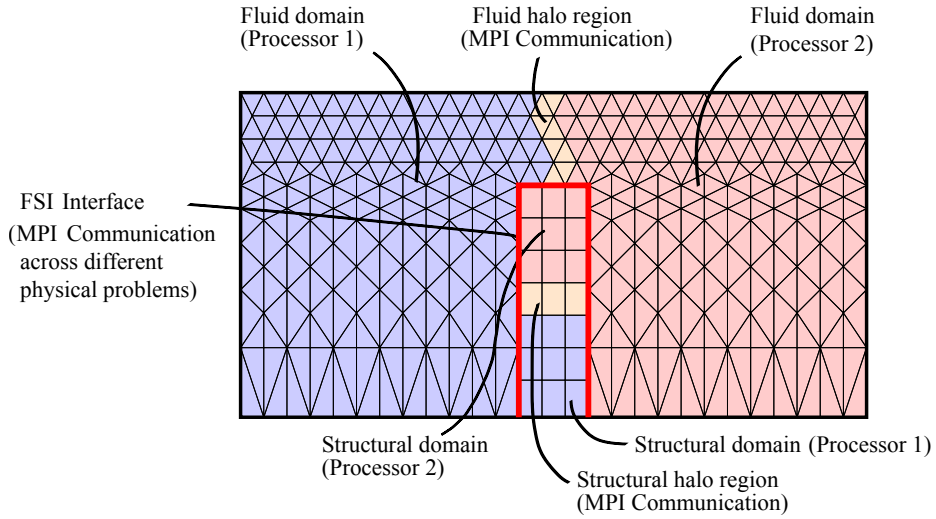


Figure 4: Sample FSI problem run in multiple cores.

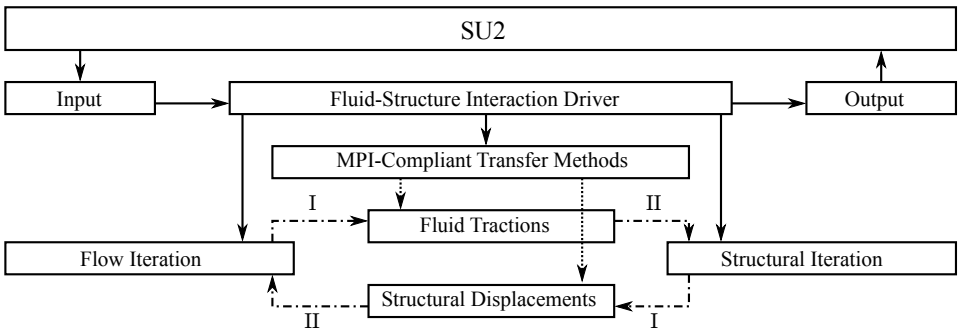


Figure 5: Transfer structure for Fluid-Structure Interaction problems.

3.3.1 Time-accurate solver

The self-contained FSI solver implemented within SU2 includes a Block Gauss-Seidel (BGS), partitioned, strongly-coupled, implicit solution method in order to advance the fluid and structural solvers in time, as described in section 2.3.1. This solver takes advantage of the modular implementation of SU2, in particular of the *driver* and *iteration* methods which allow to use each solver as a black-box, together with the transfer methods described in section 3.3. The code subiterates within each time step according to the diagram shown in Fig. 6, until the convergence criteria that has been established over the interface displacements is met.

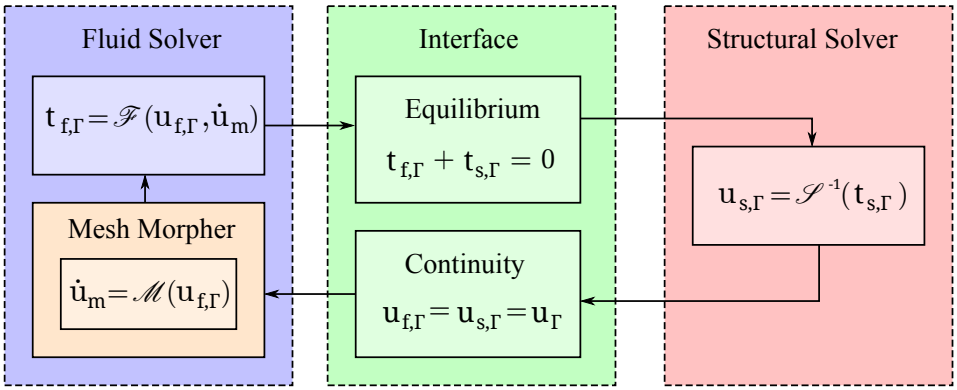


Figure 6: Block Gauss-Seidel subiterations for strong coupling.

3.3.2 Interpolation methods

It is rarely the case that the separate fluid and structure problems will require the same mesh distributions. For this reason the capability to interpolate between non-matching meshes has been implemented in SU2. The location of interpolation within the problem is shown in Fig. 7.

Interpolation is implemented as a class (CInterpolator) which locates the 'donor' nodes and calculates weighting coefficients. These values are evaluated once in a preprocessing step of the multi-zone problem, and stored such that the transfer structure will continue as described in section 3.3 with the only difference that the transferred information originates from a weighted sum of nodes rather than being restricted to a single node. The particular weighting and selection of nodes is determined by which interpolation method is chosen. Each interpolation method is implemented as a child class of CInterpolator, inheriting the data structures and methods needed by most interpolation routines.

The implemented interpolation methods include a Nearest-Neighbor approach which identifies the closest point and uses a weighting value of 1.0, and an Isoparametric approach which computes isoparametric coefficients for the nodes of the closest surface element. A 'Conservative' approach is also implemented, which re-uses the computation of isoparametric coefficients, but rather than re-computing the coefficients in either direction it computes the coefficients from the coarser mesh to the finer mesh and mirrors these values for the transfer in the opposite direction, such that the transfer matrix is the same in both directions. This is consistent with methods described in literature [52] and section 2.3.3.

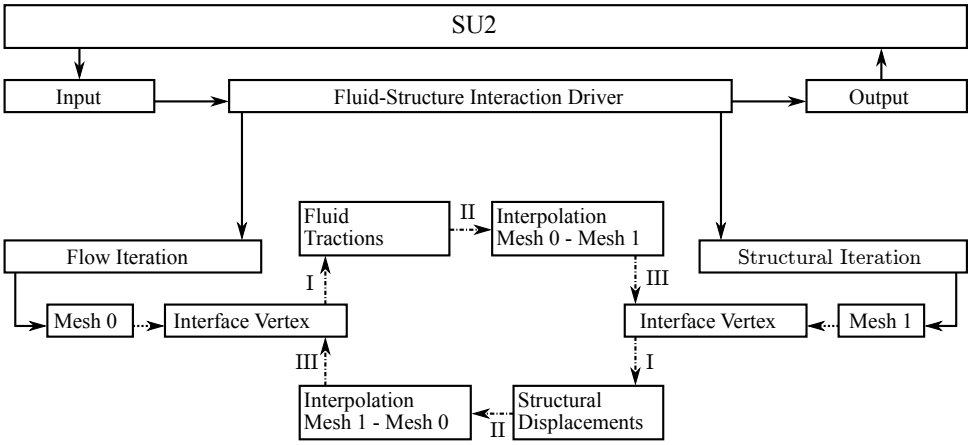


Figure 7: Interpolation structure for Fluid-Structure Interaction problems.

3.4 Python wrapper

SU2 can also be used for FSI computations when it is externally coupled with a third-party structural solver. This particular type of coupling is achieved through a Python wrapper designed to synchronize the solvers within one single Python environment. The implementation of the wrapper is detailed first, and the coupling mechanisms are illustrated after that.

3.4.1 Implementation of the wrapper

The flexibility of the Python language is used to couple SU2 with other third-party structural solvers in order to perform externally-coupled FSI simulations. Some specific functions implemented in SU2, such as those performing CFD iterations or mesh deformation, are rearranged in a new C++/API object representing the SU2 solver as the fluid part of the FSI algorithm. The SU2 code with the API object are then compiled as a dynamic library using any C++ compiler. An additional compilation using SWIG is required to translate the API into a wrapping Python module that is linked with the library and importable in any Python script where the FSI algorithm is built. In this way, the SU2 solver, through the Python-wrapped API, can be managed in an intuitive language but the critical and computationally intensive calculations are performed with the same C++ routines as in the basic source code. This wrapping methodology is illustrated in the left (black) part of Fig. 8.

3.4.2 Coupling with a third party solver

The coupling with external structural solvers is achieved within the Python script itself, providing the considered solver also has a Python-translated API that can be imported as a Python module representing the solid part of the computation. This is illustrated in the right (gray) part of Fig. 8. For modularity purposes, the methods of the API object of any structural solver to be coupled with the Python wrapper should be built on a template that covers the main steps of an FSI algorithm.

The communications between the fluid and solid modules can be directly performed through memory by explicitly exchanging memory addresses as input/output to/from the specific API functions. For example, when an update of the solid displacement is needed, the solid part

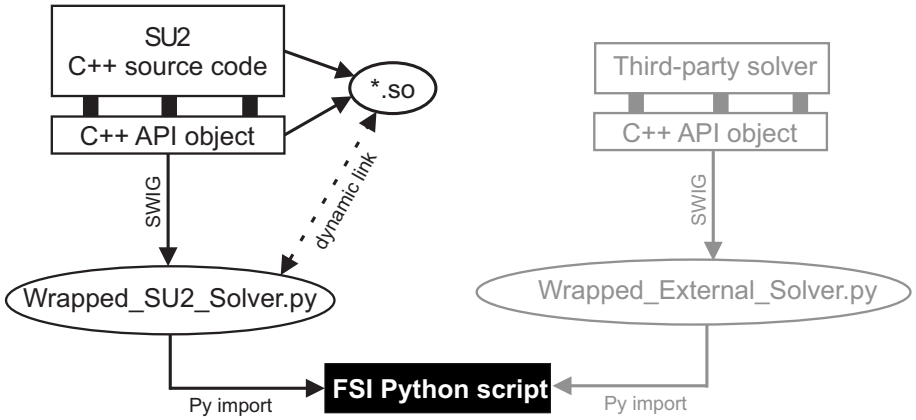


Figure 8: Python wrapping of SU2, from source code to FSI script (black). A symmetric architecture can then be applied to the structural solver (gray).

sends the memory address of the array where the new position of the fluid-solid interface is stored. These values can thus be interpolated from the solid to the fluid mesh and processed before deforming the fluid mesh. The same mechanism is used for communicating fluid loads acting on the solid. This leads to an intuitive and flexible way for synchronizing the solvers, since they are not considered as independent codes anymore but as Python objects within the FSI script. Another advantage of the wrapper is that all of the methodologies for FSI that were described in section 3.3 are also available in this approach. This implementation is also fully compatible with code parallelization using MPI thanks to appropriate MPI Python modules so that the MPI communication pattern within the solvers is kept. Third party structural solvers of various complexity can be coupled through the wrapper. In this paper, two different cases are presented. The first is one where the wrapper is used to couple SU2 to a spring-analogy solver and in the second, SU2 is coupled with an external finite element solver capable of handling 2d and 3d structural meshes for linear and non-linear structural analysis.

In the case of rigid body aeroelastic applications, the wrapper is used to couple an external in-house spring analogy solver. This solver is used to predict the dynamically constrained rigid body motion that was previously described in section 2.2.1 by integrating the equations of motion, Eq. 7, in time for a six-DOF rigid body system with the Generalized-α method. The solver itself takes the structural parameters, such as the mass, moment of inertia, stiffness or damping as inputs.

The surface tractions on each node of the interface are communicated to the spring analogy solver through the wrapper. The global aerodynamic loads, such as the lift and drag forces or the aerodynamic moments, are then computed in order to solve Eq. 7 and get the new positions of each node of the solid interface with Eq. 5. The interface displacement is then communicated from the spring analogy solver to SU2 as inputs for the dynamic mesh deformation.

For this particular case, only the fluid part is parallelized since the structural part can be almost instantaneously solved without any significant computational cost. The absence of mesh discretisation in the solid part enables the coupled simulation to get rid of the interpolation of fields at the interface. The amount of data needed to be exchanged is also small so that the communication between the fluid and the solid is performed on the master process, then the data are broadcasted/gathered to/from the other processes if needed.

For the second case, SU2 is coupled with the Toolkit for Analysis of Composite Structures (TACS) [55], a finite element based structural analysis solver. TACS has a python interface that

allows it to be coupled easily with SU2. The load transfer between the fluid and structure solvers is performed using an in-house python module developed for finite element based structural weight estimation for aircraft configurations [56]. The python based FSI framework comprising SU2, TACS and the load transfer module permits the solution of different FSI problems (steady and unsteady) in two and three dimensions. Different fluid structure coupling architectures like the Conventional Staggered or Block Gauss Seidel can be formulated without recompiling SU2 or TACS simply by rearranging the function calls to the fluid and structural modules in the Python driver scripts. A test case for the unsteady solution of an elastic beam behind a square cylinder (shown in in Section 4.2.3) is used to demonstrate the capabilities of the framework.

4 RESULTS

4.1 Fluid-Structure Interaction solver for rigid body applications

A typical aeroelastic reference model is the two-dimensional pitching-plunging airfoil [57] (2 DOF) with a chord c in a free-stream flow of velocity U_∞ (or Mach number M_∞). This model is illustrated in Fig. 9. The displacement h of the elastic axis is positive downwards and the pitch angle α is positive clockwise. The positions along the chord of the center of gravity x_{CG} and the elastic axis x_{f} are considered from the nose of the airfoil. The product between the distance $(x_{\mathrm{CG}} - x_{\mathrm{f}})$ and the airfoil mass m is the static unbalance S. The spring analogy controls the motion of the airfoil with a stiffness K_h (or K_α) and a damping C_h (or C_α) for the plunging (or pitching) mode.

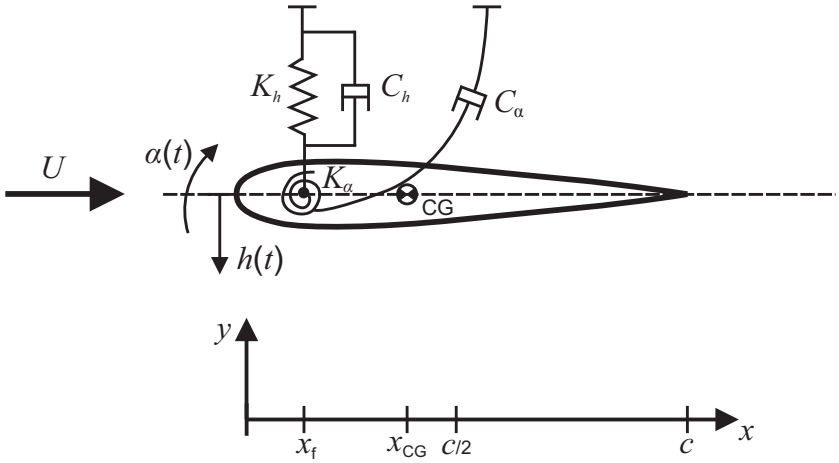


Figure 9: Schematic of a two degrees of freedom pitching-plunging airfoil aeroelastic model.

The equations of motion, Eq. 7, for this aeroelastic system can be written as:

$$\begin{aligned} m\ddot{h} + S\ddot{\alpha} + C_h\dot{h} + K_h h &= -L \\ S\ddot{h} + I_{\mathrm{f}}\ddot{\alpha} + C_\alpha\dot{\alpha} + K_\alpha\alpha &= M \end{aligned} \tag{17}$$

where I_{f} is the inertia of the airfoil around the elastic axis, L the lift and M the aerodynamic moment around the elastic axis. The lift is positive upwards and the aerodynamic moment around the elastic axis is considered positive clockwise, as the pitch angle. This model can also be described by a set of non-dimensional parameters, i.e., the normalized static unbalance χ and inertia r_α, the uncoupled natural frequency ratio $\overline{\omega}$ and the mass ratio μ:

$$\chi = \frac{S}{mb},\ r_\alpha = \frac{I_{\mathrm{f}}}{mb^2},\ \overline{\omega} = \frac{\omega_h}{\omega_\alpha},\ \mu = \frac{m}{\pi\rho_\infty b^2} \tag{18}$$

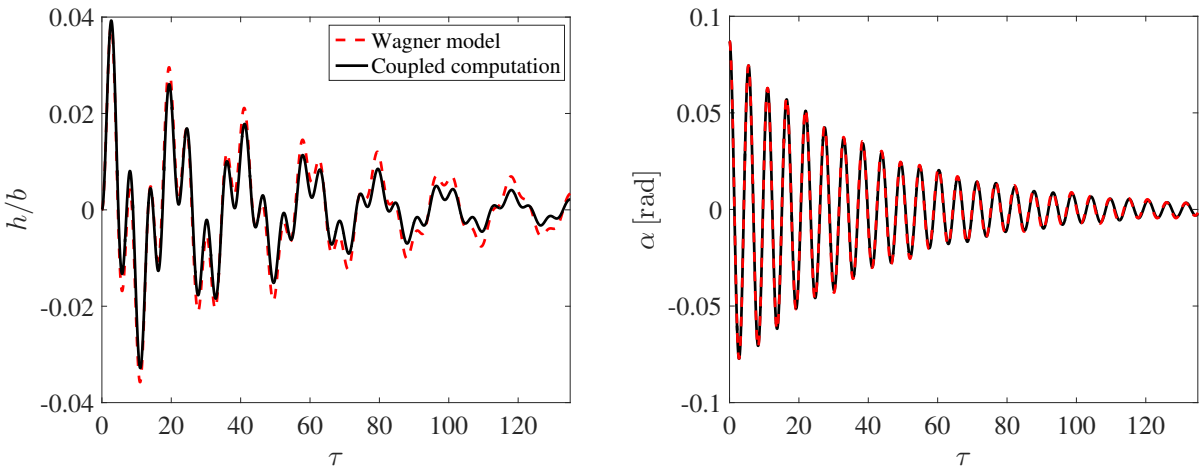
In these expressions b is the half-chord, $\omega_h = \sqrt{K_h/m}$ ($\omega_\alpha = \sqrt{K_\alpha/I_\mathrm{f}}$) is the uncoupled plunging (pitching) natural frequency and ρ_∞ is the free-stream density of the fluid.

4.1.1 Two-dimensional pitching-plunging NACA 0012 airfoil

The Python wrapper is used here to couple SU2 with the basic spring analogy integrator described in section 3.4.2 to solve the basic two-dimensional aeroelastic problem involving a pitching-plunging NACA 0012 airfoil in a free-stream flow. The unsteady RANS solver of SU2 is used with the $k-\omega$ SST turbulence model and the FSI simulation is based on a strongly-coupled scheme with a tolerance on the solid displacement of 10^{-6} (leading to about 4 FSI iterations).

The case is set with the non-dimensional parameters $\chi = 0.25$, $r_\alpha = 0.5$, $\overline{\omega} = 0.3185$ and $\mu = 100$. The elastic axis is placed at the quarter-chord point. Several free-stream Mach numbers are considered in order to capture both sub- and post-critical conditions. The chord-based Reynolds number is fixed at a value $\mathrm{Re}_c = 4 \cdot 10^6$ for a chord $c = 1$ m. The natural pitching frequency is set to $\omega_\alpha = 45$ rad/s. A pitch angle for the airfoil of $\alpha = 0.872$ rad (5°) is used as initial condition. These parameters are chosen in order to predict a subsonic flutter Mach number that allows the comparison with incompressible theoretical models (see below). The calculations are performed with 139 time steps per period of the pitch mode for accuracy purposes.

The sub-critical response of the system for $M_\infty = 0.1$ is compared with theoretical predictions using the thin airfoil theory [58] for aerodynamic load predictions and Wagner function [59] for dynamic fluid-solid coupling. Fig. 10 shows the normalized plunge displacement and pitch angle as a function of a non-dimensional time $\tau = \omega_\alpha t$. At this sub-critical Mach number the system is strongly damped. The coupled computation is in good agreement with the theoretical expectations, particularly for the pitch mode.



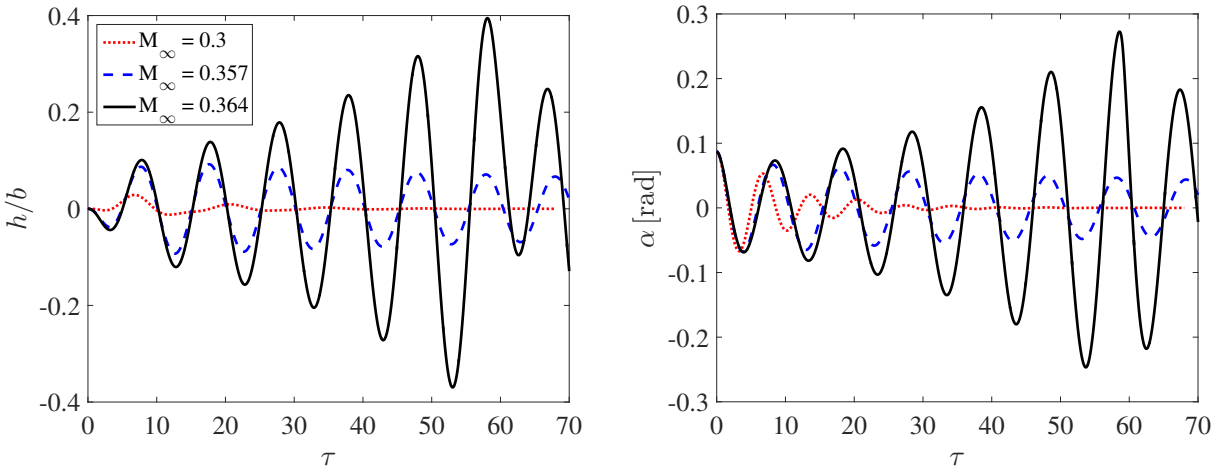
(a) Plunge displacement

(b) Pitch angle

Figure 10: Aeroelastic response as a function of time at sub-critical conditions ($M_\infty = 0.1$).

The Python wrapper is thus able to capture the response at flow conditions being quite far from critical conditions (flutter). Fig. 11 shows the computed responses of the plunging and

pitching modes at three different free-stream Mach numbers corresponding to sub-critical, critical and post-critical conditions. The flutter Mach number, $M_\infty = 0.357$, obtained from the computation differs only slightly from the theoretical prediction based on Wagner model, i.e., $M_\infty^{th} = 0.371$. As expected, the response at $M_\infty = 0.3$ is strongly damped. An increase of the Mach number reduces the damping ratio until critical conditions are reached, at which the amplitude of the response remains constant in time. Finally, at post-critical conditions (e.g., $M_\infty = 0.364$), the response becomes unstable.



(a) Plunge displacement

(b) Pitch angle

Figure 11: Aeroelastic response as a function of time at sub- and post-critical conditions.

Fig. 12 shows the computed dimensionless frequencies of the two airfoil modes as a funcion of the free-stream Mach number. The evolution of the frequencies is in good agreement with the theoretical model. In particular, the two frequencies converge with increasing M_∞ until they coincide at the flutter Mach number and beyond it.

At the post-critical conditions the amplitude of the response is strongly amplified during the first 6 cycles, but then reaches a limit-cycle oscillation (LCO). This is due to the nonlinear effects of boundary layer separation and stall that appear for large displacements and angles of attack. Fig. 13 shows the temporal response of the system over a longer time period. The red bullets correspond to instants when significant flow separation occurs alternatively on the extrados or intrados of the airfoil. A sudden drop of the amplitude is systematically observed after each separation. Subsequently, the amplitude increases again due to flutter instabilities before reaching another separation point. This mechanism allows the aeroelastic response to remain bounded in time and explains the development of a LCO mode.

A visualization of this flow dynamics is shown in Fig. 14 for times corresponding to the second set of red bullets in Fig. 13. The separation occurs when the two airfoil modes reach their local maximum values. On the suction side, the local Mach number becomes very large (close to 1.6) [Fig. 13 (a) and (b)]. Then a massive flow separation event is observed while the airfoil moves back towards its neutral position ($\alpha = h = 0$) [Fig. 13 (c) and (d)]. Finally, the flow progressively reattaches until the instantaneous displacement (both modes) of the airfoil is close to zero [Fig. 13 (e) and (f)]. A similar dynamics takes place for negative pitching angles. This highlights the capabilities of the wrapper coupling method to predict complex and nonlinear dynamics that can be observed in aeroelastic systems beyond the linear coupled mode

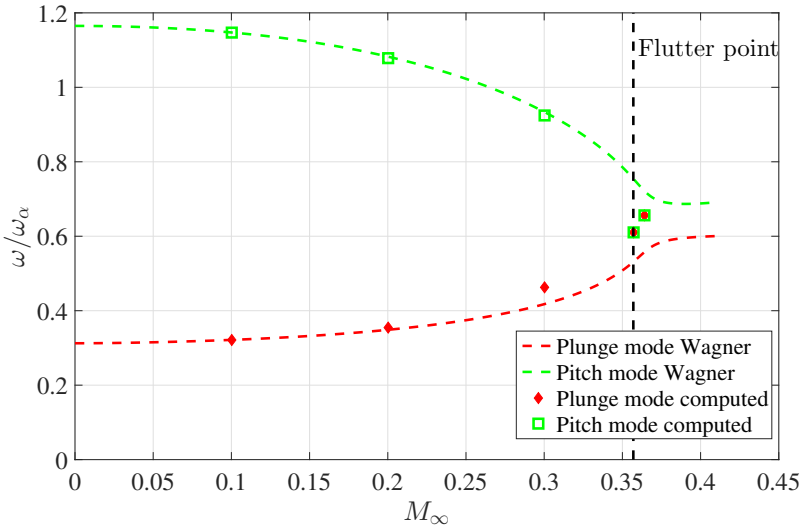


Figure 12: Dimensionless frequencies of the aeroelastic response as a function of the free-stream Mach number.

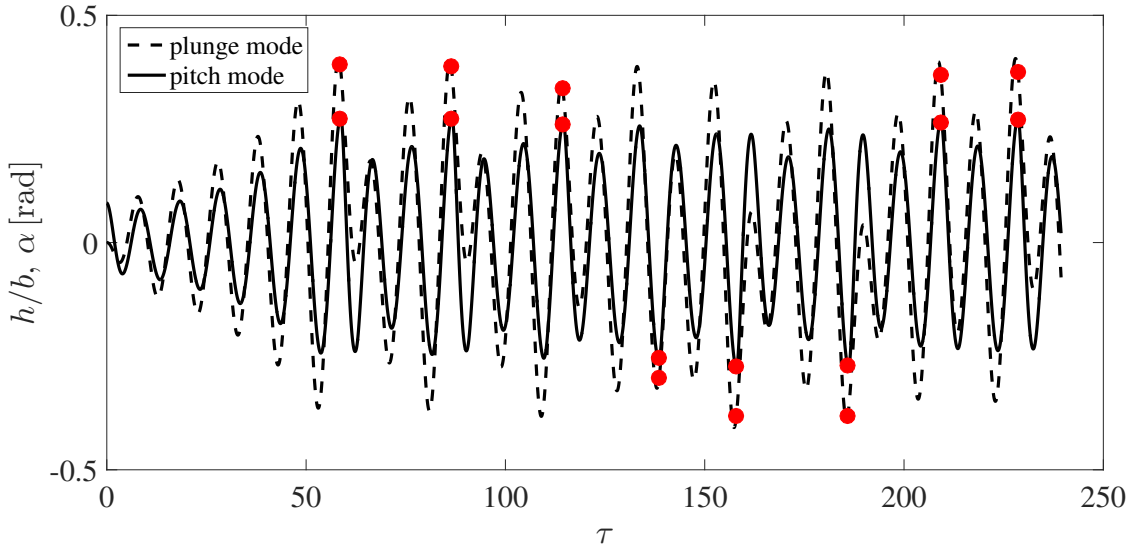


Figure 13: LCO response of the system at $M_\infty = 0.364$ under non-linear aerodynamic phenomena. The times when massive boundary layer separation occurs are marked by the red bullets.

flutter, keeping in mind the loss of accuracy for separated flows due to the RANS model.

4.1.2 Isogai wing section

The Python wrapper coupling SU2 and the spring-analogy integrator is also applied to the classical Isogai wing section aeroelastic case (case A) [60, 61]. This test case represents the dynamics of the outboard portion of a swept back wing in the transonic regime. The corresponding parameters are $\chi = 1.8$, $r_\alpha = 1.865$, $\bar{\omega} = 1$, $\mu = 60$. The elastic axis is placed in front of the airfoil at a distance $x_\mathrm{f} = -b$ from the leading edge and the natural pitching frequency is here

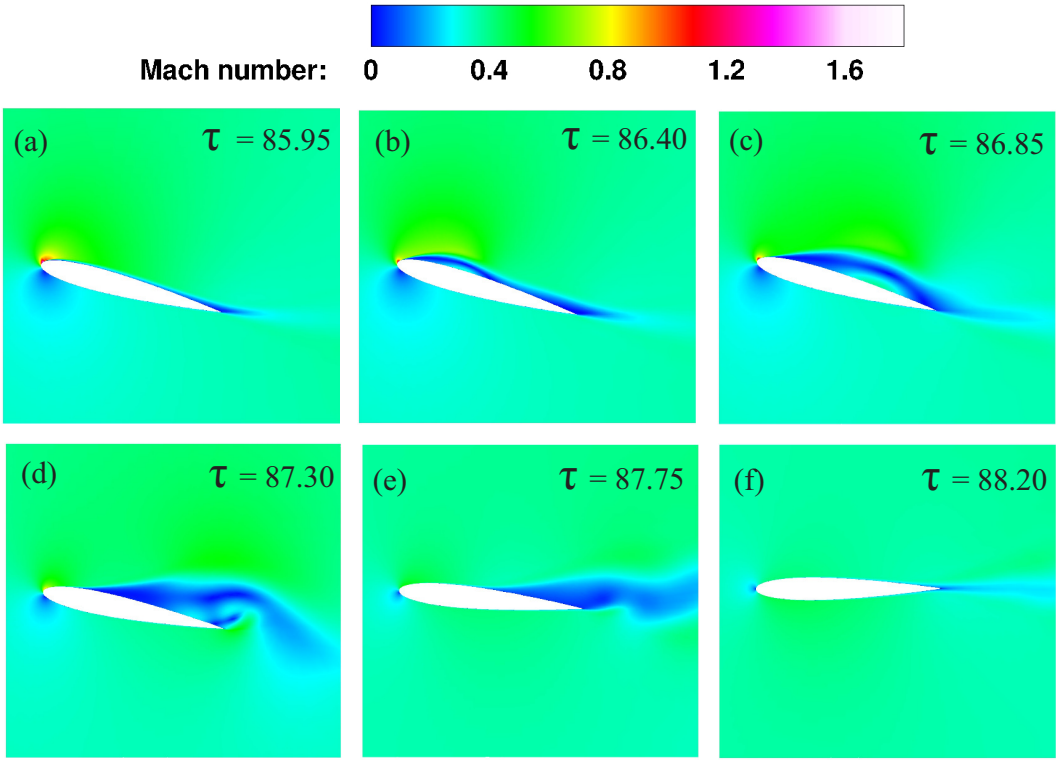


Figure 14: Visualization of the flow separation mechanism leading to a LCO response of the system beyond the flutter point.

set to $\omega_\alpha = 100$ rad/s.

The Euler equations are considered in the transonic regime ($M_\infty = 0.7 - 0.9$) with an initial pitch angle for the airfoil of $\alpha_0 = 0.0174$ rad (1^o) and the FSI simulation is again performed with a strong-coupling scheme with the same tolerance than the previous test case. For this case, the number of time steps per period of the pitch mode is reduced to 39.

Several FSI simulations at different transonic Mach numbers are performed with variable speed index

$$V^* = \frac{U_\infty}{b\omega_\alpha\sqrt{\mu}} \tag{19}$$

in order to predict the flutter point. This state is reached when the damping extracted from the dynamic response of the system is close to zero. The results are illustrated in Fig. 15 that shows a comparison of the computed flutter speed indices with those found in other references [62–65]. The best approximation curve (spline) is a representation of the limit between stable and unstable regions. It can be seen that the "transonic dip" and the typical "S-shape" flutter boundary for M_∞ between 0.7 and 0.9 are both well predicted.

4.1.3 Test cases from the 2nd AIAA Aeroelastic Prediction Workshop

The second AIAA Aeroelastic Prediction Workshop [66] was officially launched in January 2015 and took place in January 2016 (both events at Scitech) with the aim of assessing the accuracy of the predictions provided by the available numerical approaches. The planned test cases include the investigation of forced and unforced wing oscillations as well as the study of wing flutter with a pitch-and-plunge kinematics. All measurements were carried out by

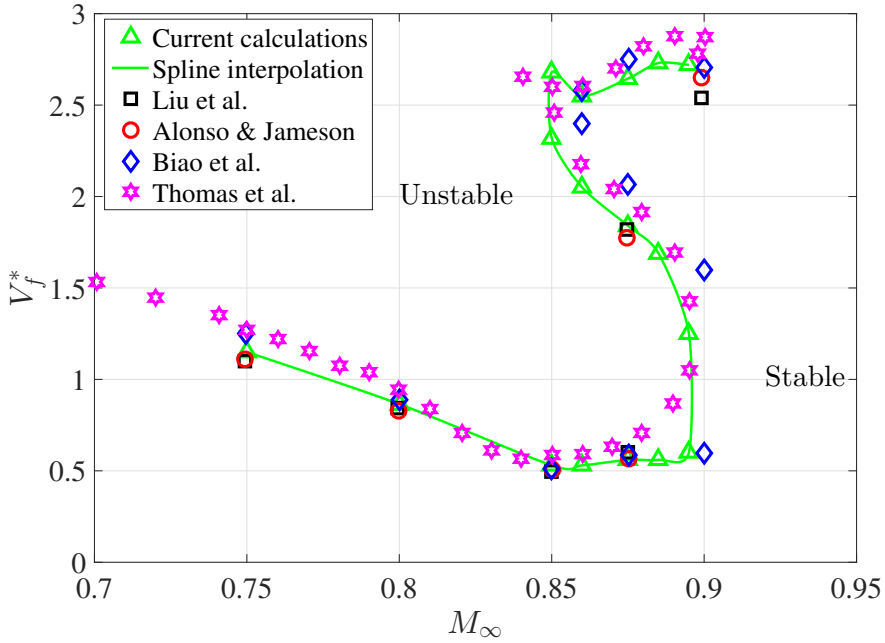


Figure 15: Flutter speed index as a function of the free-stream Mach number. Comparison between current computations and numerical results from the literature.

NASA in Langley's TDT wind tunnel. The wing model is the Benchmark Supercritical Wing (BSCW) [67–69]. All test cases are transonic and supercritical for the BSCW wing. Mach number ranges between 0.70 and 0.84, Reynolds number (based on chord) is around 5 million. The fluids used for the experiments are heavy gases (R12 and R134a depending on the test case). The angle of attack is chosen in order to position the shock wave differently in each test case. The boundary layer is tripped at 7% of the chord. The three dimensionality of the flow in the tip region – the BSCW wing has an aspect ratio of 4 – contributes to the complexity of the case.

Pre-requisites for a physically consistent simulation include the full resolution of the boundary layer, and an acceptable prediction of the steady-state shock-boundary layer interaction, hence a correct positioning of the shock on the upper and lower side of the wing.

The fluid solver in SU2 has been used to simulate test case 1, which includes forced pitch oscillations around the axis at 30% of the chord at frequency 10 Hz with an amplitude of $\pm 1^\circ$. The median angle of attack of 3° is sufficient to generate a small supersonic region around the upper side of the nose of the wing and a shock at approximately 10% of the chord. The distribution of the pressure, obtained from SU2, is shown in Fig. 16. A time-accurate simulation with rigid grid motion was conducted with the second order dual-time stepping, following a steady-state solution, also obtained with SU2. Both simulations were run with the central scheme (JST). The Spalart-Allmaras one-equation turbulence model [13] was used. For the steady-state simulation and for the subiterations in the time-accurate run, a CFL of 4.0 was used with the Euler implicit time stepping (iterations limited to 5).

Time-accurate pressure measurements - static data and first harmonic - are available in a number of points at the section at 60% of the span. The results obtained with SU2 are compared with experimental data, whenever available, and with those obtained with Edge [70], a well-known CFD package developed by the Swedish Defence Research Centre (FOI) for aeroelastic

problems.

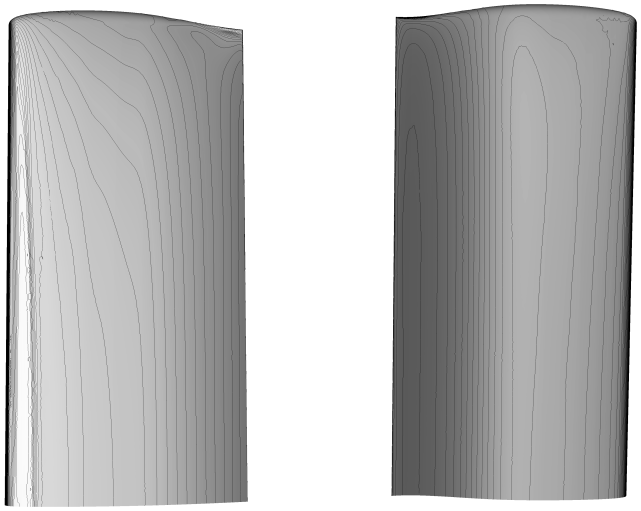


Figure 16: Steady state solution, pressure coefficient distribution over upper and lower wing sides

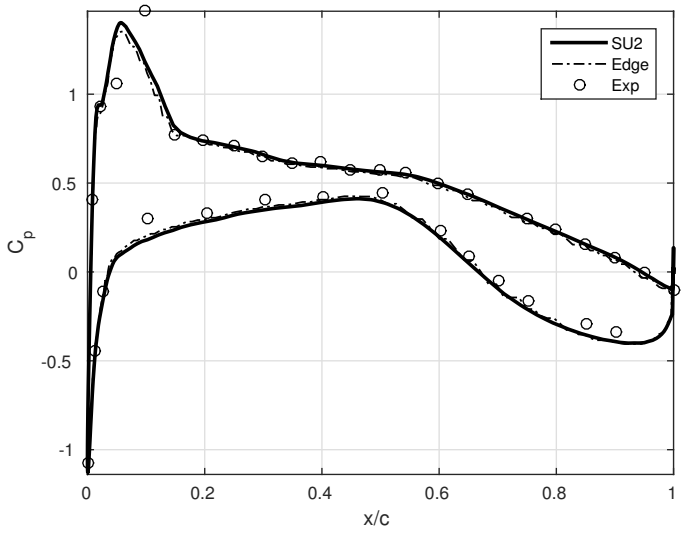


Figure 17: Pressure coefficient distribution averaged over ten periods

The agreement with the experimental data in terms of average pressure coefficient, presented in Fig. 16, may be considered as good despite a difference of a few percentage points of chord in shock position. The agreement between SU2 and Edge is excellent. The comparison of the first harmonic (10 Hz), which is presented in Fig. 18 (magnitude of oscillations) and Fig. 19 (phase delay with respect to pitch signal) shows again a very good agreement between SU2 and Edge and larger deviations, but still acceptable to engineering standards, between CFD and experiments. The most noticeable differences concern the recovery area, downstream of

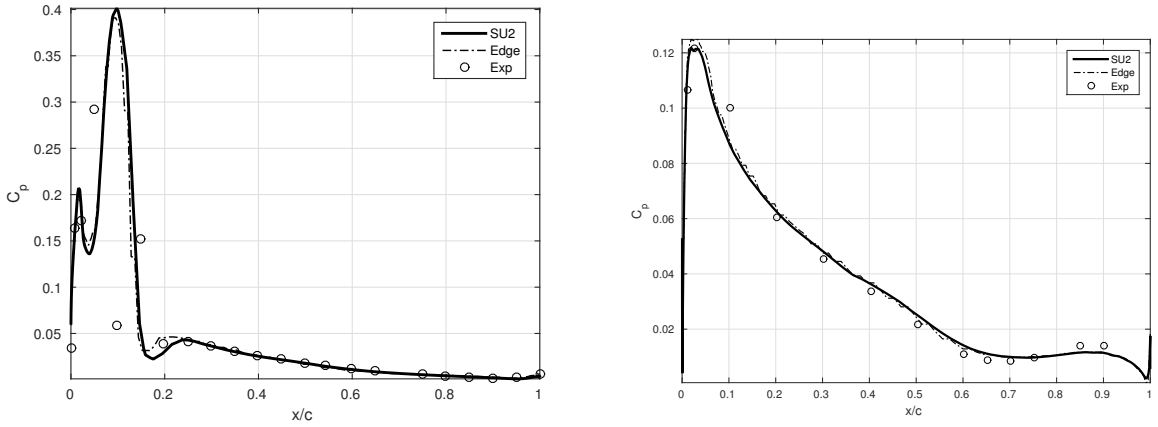


Figure 18: First harmonic of response, magnitude, pressure coefficient, upper (lhs) and lower (rhs) wing side

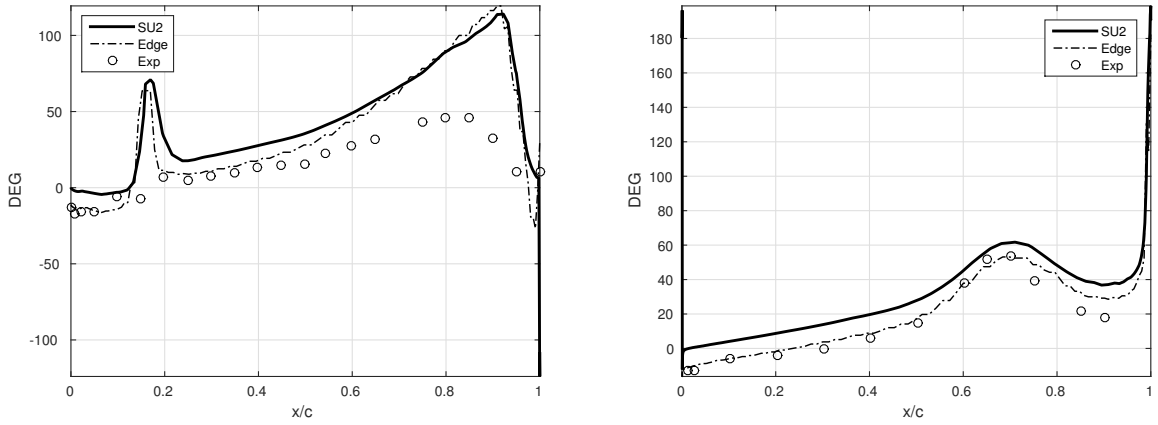


Figure 19: First harmonic of response, phase delay in degrees, pressure coefficient, upper (lhs) and lower (rhs) wing side

the shock, where the predicted phase angle is overestimated, and the shock area, where the experimental data show much lower value in correspondence of sensor 5. The most likely cause - according to the researchers and organizers of the Workshop - is some local deformation effect which is not being taken into consideration in CFD simulation. Two effects are expected to play a major role in the evolution of static pressure on the upper wing side as a function of the oscillating pitch angle; on the one hand, the motion of the stagnation point causes a first pressure peak to grow in phase with pitch angle, on the other hand, the shock wave moves up and downstream also in phase with the pitch angle. The two effects generate a "double peak", which is clearly visible in Fig. 18, in both CFD and experimental data.

The flow being strongly non-linear, the response of the aerodynamics contains several non-negligible harmonics; the Fourier expansion of the pressure coefficients has a non negligible amplitude at the reference frequency $f = 10\,Hz$ and also at $2 \times f$, $3 \times f$ and $4 \times f$. This is presented in Fig. 20. Again, the agreement between SU2 and Edge is very good. No experimental data is available for frequencies higher than 10 Hz.

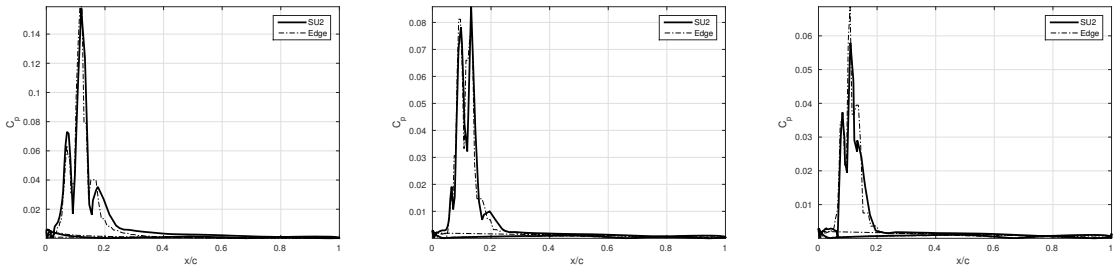


Figure 20: Second, third and fourth harmonics of response, magnitude, pressure coefficient

4.2 Fluid-Structure Interaction solver with deformable solids

4.2.1 Benchmark problem for native FSI solver in SU2

In order to test our native implementation of the FSI solver for deformable solids within SU2, we focus on a well-known benchmark test case first proposed by Wall and Ramm [71] (see, for example, [35, 44, 72–75]). They investigated the dynamics of a flexible cantilever attached to the downwind side of a square cylinder in a low-speed flow, as described in Fig. 21.

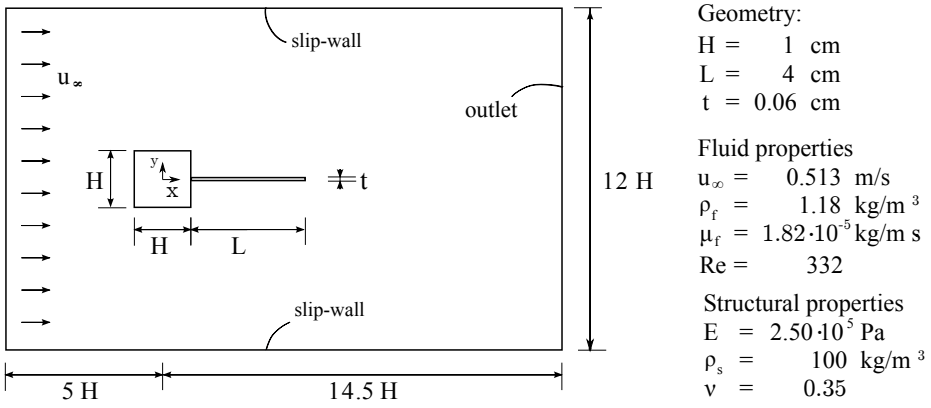


Figure 21: Benchmark for FSI solver with deformable solids.

The physics that lie behind this problem may be briefly described as follows. The square cylinder, rigid and static, generates vortical structures in the low-Reynolds number flow. These vortices generate an alternating pressure in the wake, that induces motions of the flexible cantilever attached to the downwind side of the square, as shown in Fig. 22.

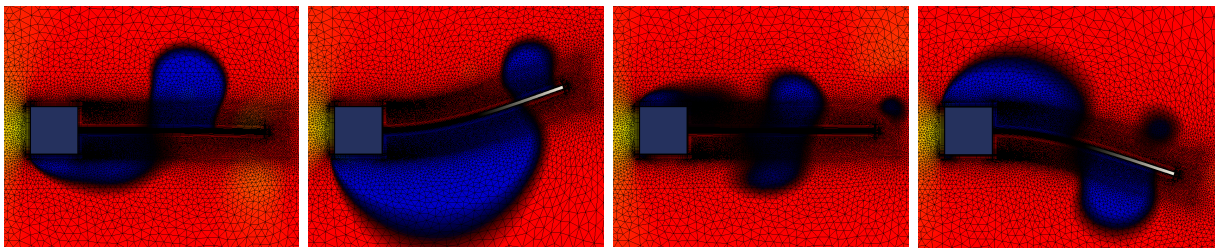


Figure 22: Pressure contours and structural displacements for the benchmark test cases for a full cycle of oscillation. From left to right, T=0, T=π/2, T= π, T=3π/2.

The two parameters that have been used in the literature to assess the validity of different FSI implementations are the frequency and amplitude of the vertical displacement at the tip of the cantilever. With regards to the frequency, published values range from 2.78 to 3.25 Hz, this is, in the surroundings of the first natural frequency of the cantilever, equal to 3.03 Hz. The maximum tip displacement has been found to be within the range 1-1.4 cm, thus a 25-35% of the cantilever length, which results in geometrically-nonlinear effects on the structure.

We have tested this problem using SU2 and several different time and space discretisations. In these tests, we have obtained values of frequency ranging from 3.05 to 3.15 Hz and maximum tip displacements in the range of 1.05 - 1.15 cm, showing a very good agreement with the literature. In a previous work, see [7], we have also shown the ability of the solver to capture some modulations in the amplitude due to the complex physics involved, and the different behaviour of the coupled problem when modifying some relevant parameters, such as the structural density and first natural frequency.

In order to reduce the computational time, we have also tested in this work the ability of the Aitken's dynamic parameter and the first order displacement predictor to reduce the number of BGS subiterations. In Fig. 23, we compare different configurations against a BGS strongly-coupled strategy in which we use a fixed relaxation parameter ω_{fixed} = 0.5. This approach converges slowly to the solution, in about 8-10 iterations per time step.

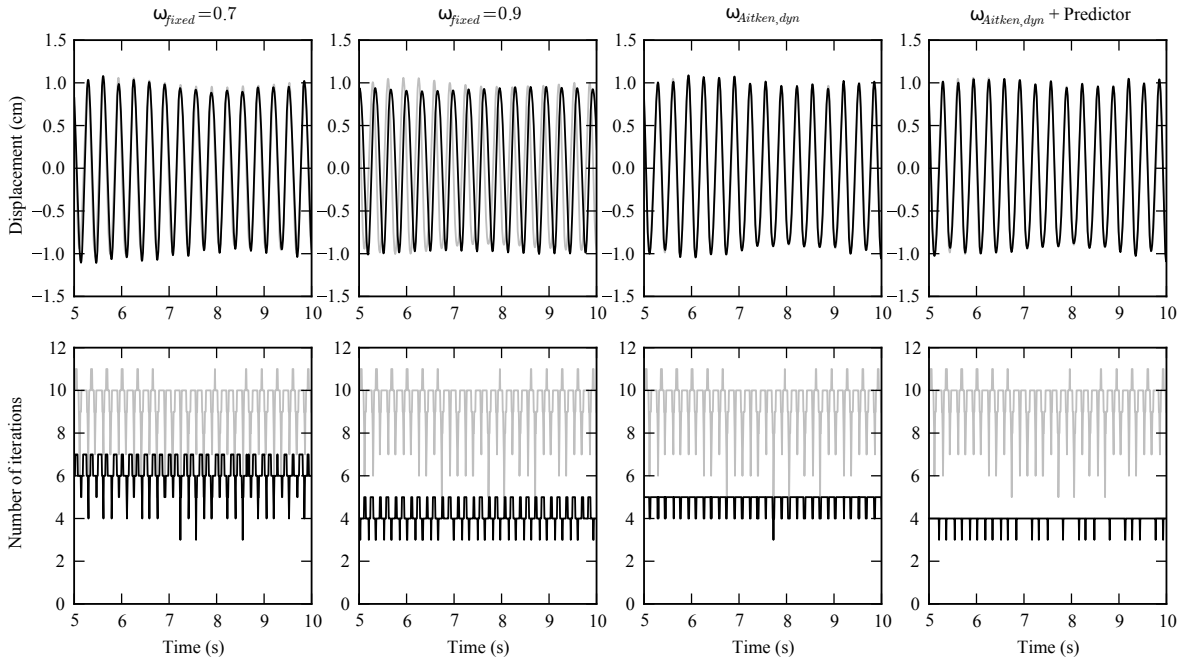


Figure 23: Time histories of the vertical tip displacements for different values of the relaxation parameter for the BGS FSI solver. The base solution (in grey) is obtained with ω_{fixed} = 0.5, and no predictor.

By increasing the fixed relaxation parameter to ω_{fixed} = 0.7, we reduce the computational time in a 33%. However, as it may be seen in Figure 23, a higher relaxation parameter introduces some deviations in the solution, that we believe are closely related to the value of ω being too large in the first subiteration, ω^0. This effect becomes more clear when we increase the fixed relaxation parameter to ω_{fixed} = 0.9, when both the frequency and the amplitude are affected

and the amplitude modulation is almost damped out.

On the other hand, the use of the Aitken's dynamic relaxation parameter ($\omega_{Aitken,dyn}$) clearly improves the convergence of the scheme. In particular, the number of BGS subiterations is reduced to 4-5 per time step, resulting in the computational time being a 38% shorter. By using a first order predictor on the first BGS subiteration, we can further reduce the number of iterations to 3-4 and the computational time by a 56% with respect to the baseline case, while obtaining effectively the same solution. It is important to note that, in these two cases, we have limited the value of the relaxation parameter in the first BGS iteration to $\omega^0 = 0.5$, then allowing it to adapt dynamically in the remaining subiterations.

4.2.2 Interpolation framework for non-matching discretizations

To test the implementation of the interpolation routines, the case of a solid wall immersed in a flow in a channel case is used. The geometry is shown in Fig. 24, and the flow conditions were set as Mach 0.15 viscous unsteady flow with adiabatic no-slip walls on the flexible vertical wall within the channel, and slip wall conditions on the upper and lower surfaces. The Reynolds number is 300 per meter. The ratio of densities between the structure and the fluid is $1 : 0.0106$.

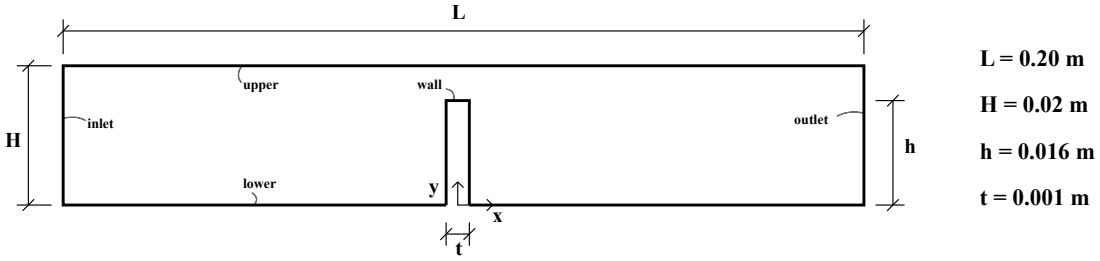


Figure 24: Test case used for interpolation routines.

Several meshes were generated with varying refinement. The results shown are from a 'fine' mesh with 100 elements on the upwind surface of the vertical wall, and a 'coarser' mesh with 90 elements on that same wall. Fig. 25 illustrates the results of simulating the coupled dynamics with either matching 'fine' discretizations, matching 'coarser' discretizations, or a non-matching discretization with the 'coarser' mesh used for the structural dynamics. This plot illustrates the effect of using these methods on the accuracy of the solution. Figure 26 illustrates the deformed geometry at time step 25. We can see from these plots that the nearest-neighbor interpolation results in discontinuous values. Using isoparametric weighting coefficients results in a smoother deformation, however when examining the accuracy of this deformation in Figure 25, we can see that a smoother interpolation has not resulted in more accurate results as compared to the nearest neighbor approach. Both of these methods use a transformation matrix $\mathbf{H}$ that has been calculated separately for the two transfer directions. The combination of this smoother interpolation and the use of a single $\mathbf{H}$ results in the most accurate interpolated results.

4.2.3 Python-wrapped FSI with interface to external solvers

In order to test/demonstrate the flexibility of the python interface of SU2, the FSI problem of a flexible cantilever attached to the downwind side of a square cylinder (described in Section 4.2.1) is solved by coupling the flow and structural solvers using the python interface. The

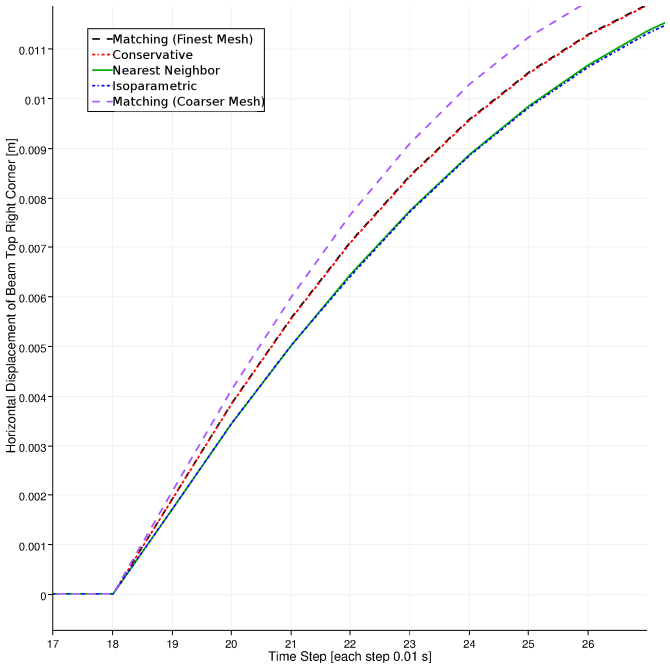
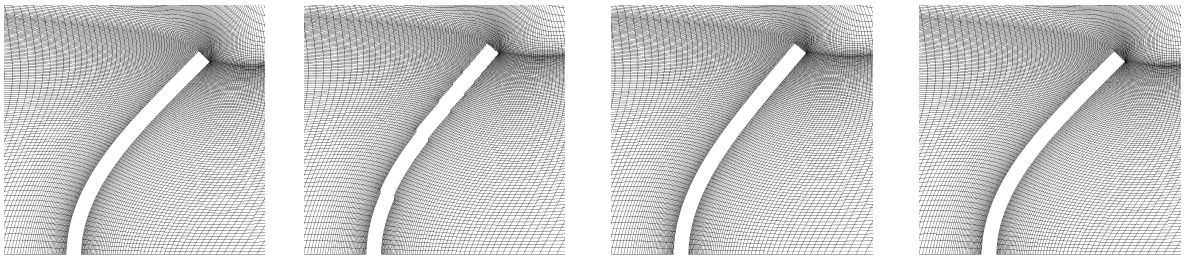


Figure 25: Time history of horizontal displacement of the upper right-hand corner of the deflecting beam. The geometry has been held static until time step 18.



(a) Matching discretization.

(b) Unmatched discretization with Nearest-Neighbor interpolation. Discontinuities can be seen in the deformed shape.

(c) Unmatched discretization with Isoparametric weights used in interpolation. The deformation is smooth, but does not agree with matched discretization results.

(d) Unmatched discretization with 'Conservative' method as in Section 2.3.3. The deformation agrees with the matched discretization, and is smooth.

Figure 26: Deformed geometry after 25 time steps

geometry of the problem is the same as Section 4.2.1 but the conditions are slightly different corresponding to those specified in Dettmer et al. [73]. These are stated again in Fig. 27. The results shown in Figs. 28, 29 were computed using a conventional staggered (CS) coupling approach. We see that the maximum tip displacement of the beam, not including the transient region, is 1.03 and the average frequency is 2.93 Hz showing a good agreement with [73] and the results in Section 4.2.1. Simulations with the Block Gauss Seidel (BGS) coupling approach give similar results but these have not been shown here.

Geometry:		Fluid properties:		Structural properties:	
H	= 1	U_∞	= 51.3	E	= $2.5 \cdot 10^6$
L	= 4	ρ_f	= $1.18 \cdot 10^{-3}$	ρ_s	= 0.1
t	= 0.06	μ_f	= $1.82 \cdot 10^{-4}$	v	= 0.35
		Re	= 333		

Figure 27: Geometry, flow and structural properties used for the 2d simulation using the python interface

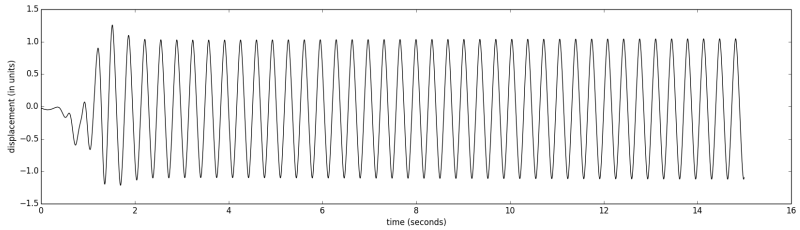


Figure 28: Time history of vertical tip displacements obtained using the python based coupling

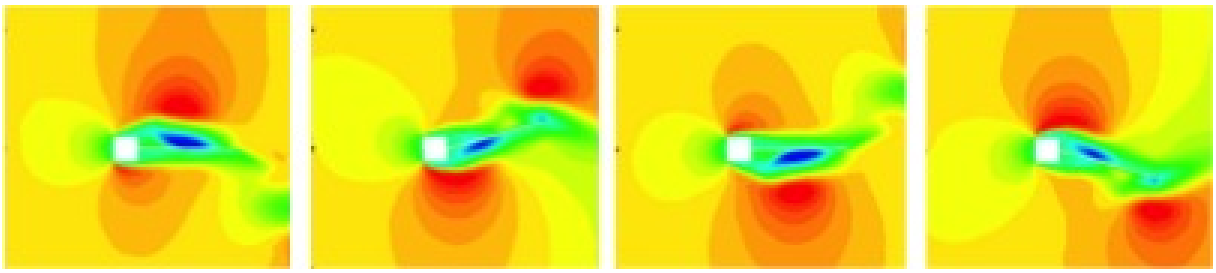


Figure 29: Convective flux contours and structural displacements for the benchmark test cases (using the python based framework) for a full cycle of oscillation. From left to right, T=0, T=$\pi/2$, T= π, T=$3\pi/2$.

5 CONCLUSION

We have introduced a new open-source Fluid-Structure Interaction solution framework tailored to high-fidelity analysis in computational aeroelasticity. It includes a natively embedded FSI solver and a code wrapper methodology that simplify the communication of the code with other external commercial and in-house solvers. We have carried out some relevant tests of the methodology and the different capabilities that have been implemented. We have compared the results obtained using this software with some relevant benchmark test cases, obtaining results that agree well with the open literature.

This infrastructure has been implemented by an international group of researchers with different backgrounds and expertise, and is publicly available in a github repository that can be accessed through the SU2 project website, http://su2.stanford.edu/. The main goal of this collaborative work is to progressively incorporate new capabilities into the platform while maintaining its modularity, thus encouraging for further developments of state-of-the-art techniques

within the different disciplines that conform the study of the interaction between fluids and structures.

ACKNOWLEDGEMENTS

H. Kline would like to acknowledge the support of a NASA Space Technology Research Fellowship.

REFERENCES

[1] C. Farhat, M. Lesoinne, and P. LeTallec. Load and motion transfer algorithms for fluid/structure interaction problems with non-matching discrete interfaces: Momentum and energy conservation, optimal discretization and application to aeroelasticity. *Computer Methods in Applied Mechanics and Engineering*, 157(1-2):95–114, 1998.

[2] A. Beckert. Coupling fluid (CFD) and structural (FE) models using finite interpolation elements. *Aerospace Science and Technology*, 4(1):13–22, 2000.

[3] C. Farhat, K.G. van der Zee, and P. Geuzaine. Provably second-order time-accurate loosely-coupled solution algorithms for transient nonlinear computational aeroelasticity. *Computer Methods in Applied Mechanics and Engineering*, 195(17-18):1973–2001, 2006.

[4] F. Palacios, M.R. Colonno, A.C. Aranake, A. Campos, S.R. Copeland, T.D. Economon, A.K. Lonkar, T.W. Lukaczyk, T.W.R. Taylor, and J.J. Alonso. Stanford University Unstructured (SU2): An open-source integrated computational environment for multi-physics simulation and design. In *AIAA 51st Aerospace Sciences Meeting, 7-10 January*, Grapevine, TX, 2013.

[5] F. Palacios, T.D. Economon, A.C. Aranake, S.R. Copeland, A.K. Lonkar, T.W. Lukaczyk, D.E. Manosalvas, K.R. Naik, A. Santiago Padrn, B. Tracey, A. Variyar, and J.J. Alonso. Stanford university unstructured (SU2): Open-source analysis and design technology for turbulent flows. In *AIAA 52nd Aerospace Sciences Meeting, SciTech, 13-17 January*, National Harbor, MD, 2014.

[6] T. Economon, F. Palacios, J. Alonso, G. Bansal, D. Mudigere, A. Deshpande, A. Heinecke, and A. Smelyanskiy. Towards High-Performance Optimizations of the Unstructured Open-Source SU2 suite. In *AIAA 53rd Aerospace Sciences Meeting, Scitech, 5-9 January*, Kissimmee , FL, 2015.

[7] R. Sanchez, R. Palacios, T.D. Economon, H.L. Kline, J.J. Alonso, and F. Palacios. Towards a Fluid-Structure Interaction solver for problems with large deformations within the open-source SU2 suite. In *57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA SciTech, 4-8 Jan*, 2016.

[8] J. Donea, A. Huerta, J.-Ph. Ponthot, and A. Rodriguez-Ferran. *Arbitrary Lagrangian-Eulerian Methods in Encyclopedia of Computational Mechanics*. John Wiley and Sons, 2004.

[9] C. Hirsch. *Numerical Computation of Internal and External Flows*. Wiley, New York, 1984.

[10] D.C. Wilcox. *Turbulence Modeling for CFD*. 2nd Ed., DCW Industries, Inc., 1998.

[11] F. M. White. *Viscous Fluid Flow*. McGraw Hill Inc., New York, 1974.

[12] F.R. Menter. Zonal two equation $k - \omega$, turbulence models for aerodynamic flows. *AIAA Paper 1993-2906*, 1993.

[13] P. Spalart and S. Allmaras. A one-equation turbulence model for aerodynamic flows. Number AIAA Paper 1992-0439, 1992.

[14] A. Jameson, W. Schmidt, and E. Turkel. Numerical solution of the Euler equations by finite volume methods using Runge-Kutta time stepping schemes. Number AIAA Paper 1981-1259, 1981.

[15] P. L. Roe. Approximate riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43(2):357 – 372, 1981.

[16] J. M. Weiss, J. P. Maruszewski, and A. S. Wayne. Implicit solution of the Navier-Stokes equation on unstructured meshes. *AIAA Paper 1997-2103*, 1997.

[17] A. Jameson. Time dependent calculations using multigrid, with applications to unsteady flows past airfoils and wings. Number AIAA Paper 1991-1596, 1991.

[18] A. Jameson and S. Schenectady. An assessment of dual-time stepping, time spectral and artificial compressibility based numerical algorithms for unsteady flow with applications to flapping wings. Number AIAA Paper 2009-4273, 2009.

[19] P. D. Thomas and C. K. Lombard. Geometric conservation law and its application to flow computations on moving grids. *AIAA Journal*, 17(10):1030–1037, October 1979.

[20] J. T. Batina. Unsteady Euler airfoil solutions using unstructured dynamic meshes. *AIAA Journal*, 28(8):1381–1388, August 1990.

[21] M. Lesoinne and C. Farhat. Geometric conservation laws for flow problems with moving boundaries and deformable meshes, and their impact on aeroelastic computations. *Computer Methods in Applied Mechanics and Engineering*, 134(1–2):71 – 90, 1996.

[22] C. Farhat, P. Geuzaine, and C. Grandmont. The discrete geometric conservation law and the nonlinear stability of ALE schemes for the solution of flow problems. *Journal of Computational Physics*, 174(2):669–694, December 2001.

[23] S. A. Morton, R. B. Melville, and M. R. Visbal. Accuracy and coupling issues of aeroelastic Navier-Stokes solutions on deforming meshes. *Journal of Aircraft*, 35(5):798–805, 1998.

[24] R. T. Biedron and J. L. Thomas. Recent enhancements to the FUN3D flow solver for moving-mesh applications. *AIAA Paper 2009-1360*, 2009.

[25] M. Geradin and A. Cardnna. *Flexible Multibody Dynamics, A Finite Element Approach*, chapter Kinematics of Finite Motion. John Wiley & Sons, LTD, 2001.

[26] M. Geradin and D.J. Rixen. *Mechanical Vibrations : Theory and Application to Structural Dynamics*, chapter Analytical Dynamics of Discrete Systems. Wiley, 2015.

[27] F. Amirouche. *Fundamentals of Multibody Dynamics, Theory and Applications*, chapter Hamilton-Lagrange and Gibbs-Appel Equations. Birkhauser, 2006.

[28] M. Hojjat, E. Stavropoulou, T. Gallinger, U. Israel, R. Wchner, and Bletzinger K.-U. *Fluid Structure Interaction II*, volume 73, chapter Fluid-Structure Interaction in the Context of Shape Optimization and Computational Wind Engineering, pages 351–381. Springer Berlin Heidelberg, 2010.

[29] J. Bonet and R. D. Wood. *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press, 1997.

[30] J. Donea, A. Huerta, J.-Ph. Ponthot, and A. Rodríguez-Ferran. *Encyclopedia of Computational Mechanics*, chapter Arbitrary Lagrangian-Eulerian Methods. John Wiley & Sons, Ltd, 2004.

[31] K.-J. Bathe. *Finite Element Procedures in Engineering Analysis*. Civil engineering and engineering mechanics series. Prentice-Hall, 1982.

[32] N. M. Newmark. A method of computation for structural dynamics. *Journal of the Engineering Mechanics Division*, 85(3):67–94, 1959.

[33] J. Chung and G.M. Hulbert. Time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method. *Journal of Applied Mechanics, Transactions ASME*, 60(2):371–375, 1993.

[34] S. Deparis, M. Discacciati, G. Fourestey, and A. Quarteroni. Fluid-structure algorithms based on Steklov-Poincaré operators. *Computer Methods in Applied Mechanics and Engineering*, 195(41-43):5797–5812, 2006.

[35] C. Kassiotis, A. Ibrahimbegovic, R. Niekamp, and H.G. Matthies. Nonlinear fluid-structure interaction problem. part I: Implicit partitioned algorithm, nonlinear stability proof and validation examples. *Computational Mechanics*, 47(3):305–323, 2011.

[36] S. Piperno and C. Farhat. Partitioned procedures for the transient solution of coupled aeroelastic problems part II: energy transfer analysis and three-dimensional applications. *Computer Methods in Applied Mechanics and Engineering*, 190(2425):3147–3170, 2001. Advances in Computational Methods for Fluid-Structure Interaction.

[37] C. Farhat, M. Lesoinne, P. Stern, and S. Lantri. High performance solution of three-dimensional nonlinear aeroelastic problems via parallel partitioned algorithms: Methodology and preliminary results. *Advances in Engineering Software*, 28(1):43–61, 1997.

[38] C. Farhat and M. Lesoinne. Two efficient staggered algorithms for the serial and parallel solution of three-dimensional nonlinear transient aeroelastic problems. *Computer Methods in Applied Mechanics and Engineering*, 182(34):499–515, 2000.

[39] W.G. Dettmer and D. Perić. A new staggered scheme for fluid-structure interaction. *International Journal for Numerical Methods in Engineering*, 93(1):1–22, 2013.

[40] M. Von Scheven and E. Ramm. Strong coupling schemes for interaction of thin-walled structures and incompressible flows. *International Journal for Numerical Methods in Engineering*, 87(1-5):214–231, 2011.

[41] J. Degroote, K.-J. Bathe, and J. Vierendeels. Performance of a new partitioned procedure versus a monolithic procedure in fluid-structure interaction. *Computers and Structures*, 87(11-12):793–801, 2009.

[42] P. Le Tallec and J. Mouro. Fluid structure interaction with large structural displacements. *Computer Methods in Applied Mechanics and Engineering*, 190(24-25):3039–3067, 2001.

[43] U. Küttler and W.A. Wall. Fixed-point fluid-structure interaction solvers with dynamic relaxation. *Computational Mechanics*, 43(1):61–72, 2008.

[44] C. Habchi, S. Russeil, D. Bougeard, J.-L. Harion, T. Lemenand, A. Ghanem, D.D. Valle, and H. Peerhossaini. Partitioned solver for strongly coupled fluid-structure interaction. *Computers and Fluids*, 71:306–319, 2013.

[45] B. M. Irons and R. C. Tuck. A version of the Aitken accelerator for computer iteration. *International Journal for Numerical Methods in Engineering*, 1:275–277, 1969.

[46] M.J. Smith, D.H. Hodges, and C.E.S. Cesnik. Evaluation of computational algorithms suitable for fluid-structure interactions. *Journal of Aircraft*, 37(2):282–294, 2000.

[47] J.R. Cebral and R. Löhner. Conservative load projection and tracking for fluid-structure problems. *AIAA Journal*, 35(4):687–692, 1997.

[48] G.P. Guruswamy. A review of numerical fluids/structures interface methods for computations using high-fidelity equations. *Computers and Structures*, 80(1):31–41, 2002.

[49] A. de Boer, A.H. van Zuijlen, and H. Bijl. Review of coupling methods for non-matching meshes. *Computer Methods in Applied Mechanics and Engineering*, 196(8):1515–1525, 2007.

[50] X. Jiao and M.T. Heath. Overlaying surface meshes, part I: Algorithms. *International Journal of Computational Geometry and Applications*, 14(6):379–402, 2004.

[51] R.K. Jaiman, X. Jiao, P.H. Geubelle, and E. Loth. Assessment of conservative load transfer for fluid-solid interface with non-matching meshes. *International Journal for Numerical Methods in Engineering*, 64(15):2014–2038, 2005.

[52] S A Brown. Displacement Extrapolations for CFD+CSM Aeroelastic Analysis. *AIAA Paper*, pages 291–300, 1997.

[53] A. Beckert and H. Wendland. Multivariate interpolation for fluid-structure-interaction problems using radial basis functions. *Aerospace Science and Technology*, 5(2):125–134, 2001.

[54] T. D. Economon, F. Palacios, S. R. Copeland, T. W. Lukaczyk, and J. J. Alonso. SU2: An Open-Source Suite for Multi-Physics Simulation and Design. *AIAA Journal*, (accepted), 2015.

[55] G. J. Kennedy and J.R.R.A. Martins. A parallel aerostructural optimization framework for aircraft design studies. *Structural and Multidisciplinary Optimization, Volume 50, Issue 6*, 2014.

[56] A. Variyar, T.D. Economon, and J. J. Alonso. Multifidelity conceptual design and optimization of strut-braced wing aircraft using physics based methods. *54th AIAA Aerospace Sciences Meeting*, 2016.

[57] R.L. Bisplinghoff, H. Ashley, and R. L. Halfman. *Aeroelasticity*. Dover Publications, 1996.

[58] Anderson J.D. *Fundamentals of Aerodynamics*, chapter Incompressible Flow over Airfoils. Mc Graw Hill, Inc., 2011.

[59] Y. C. Fung. *An Introduction to the theory of aeroelasticity*, chapter Fundamentals of flutter analysis. Dover Publications, 2002.

[60] K. Isogai. On the transonic-dip mechanism of flutter of a sweptback wing. *AIAA Journal*, 17(7):793–795, 1979.

[61] K. Isogai. Transonic-dip mechanism of flutter of a sweptback wing: part ii. *AIAA Journal*, 19(9):1240–1242, 1981.

[62] F. Liu, J. Cai, and Y. Zhu. Calculation of wing flutter by a coupled fluid-structure method. *Journal of Aircraft*, 38(2):334–342, 2001.

[63] J.J. Alonso and A. Jameson. Fully-implicit time marching aeroelastic solution. In *AIAA 32nd Aerospace Sciences Meeting and Exhibit, 10-13 January*, 1994.

[64] Z. Biao, Q. Zhide, and G. Chao. Transonic flutter analysis of an airfoil with approximate boundary method. In *26th international congress of the aeronautical sciences*, 2008.

[65] J.T. Thomas, K.C. Hall, and E.H. Dowell. Reduced-order aeroelastic modeling using proper-orthogonal decomposition. *CEAS/AIAA/ICASE/NASA Langley International Forum on Aeroelasticity and Structural Dynamics*, 1999.

[66] J. Heeg, P. Chwalowski, D. M. Schuster, D. Raveh, A. Jirasek, and M. Dalenbring. Plans and example results for the 2nd AIAA Aeroelastic Prediction Workshop. *AIAA Paper*, 2015.

[67] B. E. Dansberry, M. H. Durham, R. M. Bennett, D. L. Turnock, E. A. Silva, and J. A. Rivera Jr. *Physical properties of the benchmark models program supercritical wing*, volume 4457. Citeseer, 1993.

[68] D. J. Piatak and C. S. Cleckner. Oscillating turntable for the measurement of unsteady aerodynamic phenomena. *Journal of aircraft*, 40(1):181–188, 2003.

[69] R. M. Bennett, C. V. Eckstrom, J. A Rivera Jr, B. E. Dansberry, M. G. Farmer, and M. H. Durham. The benchmark aeroelastic models program: Description and highlights of initial results. 1991.

[70] P. Eliasson and P. Weinerfelt. Recent Applications of the Flow Solver Edge. *Proceedings of the 7th Asian CFD Conference*, 2007.

[71] W.A. Wall and E. Ramm. Fluid-Structure interaction based upon a stabilized (ALE) finite element method. In E. Oñate S.R. Idelsohn and E.N. Dvorkin (Eds.), editors, *Computational Mechanics. New Trends and Applications*. CIMNE, Barcelona, Spain, 1998.

[72] H. G. Matthies and J. Steindorf. Partitioned strong coupling algorithms for fluidstructure interaction. *Computers & Structures*, 81(811):805–812, 2003.

[73] W. Dettmer and D. Perić. A computational framework for fluid-structure interaction: Finite element formulation and applications. *Computer Methods in Applied Mechanics and Engineering*, 195(41-43):5754–5779, 2006.

[74] C. Wood, A.J. Gil, O. Hassan, and J. Bonet. A partitioned coupling approach for dynamic fluid-structure interaction with applications to biological membranes. *International Journal for Numerical Methods in Fluids*, 57(5):555–581, 2008.

[75] B. Froehle and P.-O. Persson. A high-order discontinuous Galerkin method for fluid-structure interaction with efficient implicit-explicit time stepping. *Journal of Computational Physics*, 272:97–104, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

CONVERGENCE ANALYSIS OF COUPLING ITERATIONS FOR THE UNSTEADY TRANSMISSION PROBLEM WITH MIXED DISCRETIZATIONS

Azahar Monge[1] **and Philipp Birken**[1]

[1]Centre for Mathematical Sciences
Lund University
Box 118, 22100, Lund, Sweden
e-mail: azahar.monge@na.lu.se, web page: http://www.maths.lu.se/staff/azahar-monge

Keywords: Thermal Fluid Structure Interaction, Coupled Problems, Transmission Problem, Fixed Point Iteration, Dirichlet-Neumann Iteration

Abstract. *We analyze the convergence rate of the Dirichlet-Neumann iteration for the fully discretized one dimensional unsteady transmission problem. Specifically, we consider the coupling of two linear heat equations on two identical non overlapping intervals. The Laplacian is discretized using finite differences on one interval and finite elements on the other and the implicit Euler method is used for the time discretization. Following previous analysis where finite elements where used on both subdomains, we provide an exact formula for the spectral radius of the iteration matrix for this specific mixed discretizations. We then show that these tend to the ratio of heat conductivities in the semidiscrete spatial limit, but to a factor of the ratio of the products of density and specific heat capacity in the semidiscrete temporal one. In the previous finite element analysis, the same result was obtained in the semidiscrete spatial limit but the factor in the temporal limit was lower. This explains the fast convergence previously observed for cases with strong jumps in the material coefficients. Numerical results confirm the analysis.*

1 INTRODUCTION

Thermal fluid structure interaction occurs when a heat flux from a fluid leads to temperature changes in a structure or vice versa. Examples for this are cooling of gas-turbine blades, cooling of rocket thrust chambers [11, 12], thermal anti-icing systems of airplanes [4], supersonic reentry of vehicles from space [10, 14] or gas quenching [8, 19].
Unsteady thermal fluid structure interaction is modelled using two partial differential equations describing a fluid and a structure on different domains. The equations are coupled at an interface to model the heat transfer between fluid and structure. The standard algorithm to find solutions of the coupled problem is the Dirichlet-Neumann iteration, where the PDEs are solved separately using Dirichlet-, respectively Neumann boundary conditions with data given from the solution of the other problem. The convergence rate of the method has been analyzed in any standard book on domain decomposition method, e.g. [18, 20]. There, the iteration matrix is derived in terms of the spatial discretization matrices and the convergence rate is the spectral radius of that. However, this does not provide a quantitative answer, since the spectral radius is unknown.
We consider the transmission problem because it is a basic building block in fluid structure interaction. For this case, a one dimensional stability analysis was presented by Giles [7]. There, an explicit time integration method was chosen with respect to the interface unknowns. Henshaw and Chand provided in [9] a method to analyze stability and convergence speed of the Dirichlet-Neumann iteration in 2D based on applying the continuous Fourier transform to the semi-discretized equations. Their result depends on ratios of thermal conductivities and diffusivities of the materials. This is similar to the situation in [1, 5] where the performance of the coupling for incompressible fluids is affected by the added mass effect. However, in the fully discrete case we observe that the iteration converges much faster for some choices of materials [3], and that the speed of the iteration does not depend on the thermal diffusivities in some cases. In [16] the discrete case was analyzed for finite element discretizations. In this paper, we present a similar one dimensional analysis in the case of specific mixed discretizations.
Thus, we consider a complete discretization of the coupled problem using finite differences on one domain and finite elements on the other (in space) and the implicit Euler method in time. Then, we compute the spectral radius of the iteration matrix exactly in terms of the eigendecomposition of the resulting matrices for the one dimensional case. The asymptotics of the convergence rates when approaching the continuous case in either time or space are also computed resulting in the ratio of the thermal conductivities in space and a factor of the ratio of the heat capacities in time. These results are consistent with our numerical experiments.
An outline of the paper now follows. In Section 2, we define the problem to be solved in terms of the partial differential equations, boundary conditions and interface conditions. We also give a description of the discretization. In Section 3, we explain the Dirichlet-Neumann model. Our analysis for the discrete case of the model problem using Dirichlet-Neumann interface conditions for mixed discretizations is presented in Section 4. In Section 5, we present numerical results that show the theoretical stability analysis.

2 MODEL PROBLEM

The unsteady transmission problem is as follows, where we consider a domain $\Omega \subset \mathbb{R}^d$ which is cut into two subdomains $\Omega = \Omega_1 \cup \Omega_2$ with transmission conditions at the interface $\Gamma = \Omega_1 \cap \Omega_2$:

$$\begin{aligned}
\alpha_m \frac{\partial u_m(\mathbf{x},t)}{\partial t} + \lambda_m \Delta u_m(\mathbf{x},t) = f(x), \;\; t \in [t_0, t_f], \;\; \mathbf{x} \in \Omega_m \subset \mathbb{R}^d, \; m = 1,2, \\
u_m(\mathbf{x},y) = 0, \;\; t \in [t_0, t_f] \;\; \mathbf{x} \in \partial\Omega_m \backslash \Gamma, \\
u_1(\mathbf{x},y) = u_2(\mathbf{x},y), \;\; \mathbf{x} \in \Gamma, \\
\lambda_1 \partial_x u_1(\mathbf{x},y) \cdot \mathbf{n} = \lambda_2 \partial_x u_2(\mathbf{x},y) \cdot \mathbf{n}, \;\; \mathbf{x} \in \Gamma, \\
u_m(\mathbf{x},0) = u_m^0(\mathbf{x}), \;\; \mathbf{x} \in \Omega_m.
\end{aligned} \tag{1}$$

where $\mathbf{n}$ denotes the normal vector and we consider $d = 1, 2$.
The constants λ_1 and λ_2 describe the thermal conductivities of the materials on Ω_1 and Ω_2 respectively. $\alpha_m = \rho_m C_m$ where ρ_m respresents the density and C_m the heat capacity of the material placed in Ω_m, $m = 1, 2$. D_1 and D_2 represent the thermal diffusivities of the materials and they are defined by

$$D_m = \frac{\lambda_m}{\alpha_m}. \tag{2}$$

We consider a constant mesh width of $\Delta x = 1/(N+1)$ with N being the number of interior space discretization points in both Ω_1 and Ω_2. Moreover, we discretize this problem using a finite difference method (FDM) on Ω_1 and a finite element method (FEM) on Ω_2. The implicit Euler method is used for the time discretization.

2.1 Space Discretization

First of all, we focus on the FDM formulation on Ω_1 of problem (1). For this, let $\mathbf{u}_I^{(1)}$ correspond to the unknowns on Ω_1 and $\mathbf{u}_\Gamma$ correspond to the unknows at the interface Γ. From now on we assume that $f_1 = 0$ in order to simplify the analysis.
Then, applying second order central differences

$$\Delta u_1(x_i,t) \approx \frac{1}{\Delta x^2}(u_1(x_{i+1},t) - 2u_1(x_i,t) + u_1(x_{i-1},t)), \;\; \text{for} \;\; i = 1, .., N, \tag{3}$$

to approximate the second order spatial derivative of (1) for $m = 1$, we can write the resulting discrete system as:

$$\alpha_1 \dot{\mathbf{u}}_I^{(1)} + \mathbf{A}_1 \mathbf{u}_I^{(1)} + \mathbf{A}_{I\Gamma}^{(1)} \mathbf{u}_\Gamma = \mathbf{0}. \tag{4}$$

$\mathbf{A}_1$ corresponds to the discretization of the Laplacian operator on Ω_1 and the required data from the interface is inserted in the equation by the matrix $\mathbf{A}_{I\Gamma}^{(1)}$.
On the other hand, we also need to discretize the problem on Ω_2 using FEM formulation. For this, we choose an approximation U_2 of u_2 in a finite-dimensional subspace S^N of H^1 having the form

$$U_2(\mathbf{x},t) = \sum_{j=1}^{N-1} c_j(t)\phi_j(x). \tag{5}$$

where N is the number of interior spatial discretization points on Ω_2, c_j are the coefficient functions and ϕ_j the test functions. Then, the Galerkin finite element problem is to determine $U_2 \in S^N$ such that

$$\begin{aligned} \alpha_2 \int_{\Omega_2} \frac{\partial U_2}{\partial t}\phi_j dV + \lambda_2 \int_{\Omega_2} \Delta U_2 \phi_j dV = \int_{\Omega_2} f_2 \phi_j dV \;\; t \in [t_0, t_f], \\ \int_{\Omega_2} U_2 \phi_j dV = \int_{\Omega_2} U_2^0 \phi_j dV, \;\; t = 0, \;\; j = 1, 2, ..., N-1. \end{aligned} \tag{6}$$

As before, let $\mathbf{u}_I^{(2)}$ correspond to the unknowns on Ω_2 and assume that $f_2 = 0$ in order to simplify the analysis. Then, applying Green's formula to (6) in order to remove the Laplacian operator and letting j run over the interior nodes on Ω_2, we can write the resulting discrete system as:

$$\mathbf{M}_2 \dot{\mathbf{u}}_I^{(2)} + \mathbf{M}_{I\Gamma}^{(2)} \dot{\mathbf{u}}_\Gamma + \mathbf{A}_2 \mathbf{u}_I^{(2)} + \mathbf{A}_{I\Gamma}^{(2)} \mathbf{u}_\Gamma = \mathbf{0}. \tag{7}$$

$\mathbf{A}_2$ and $\mathbf{M}_2$ are the stiffness and the mass matrix for the interior nodes on Ω_2 and they are defined as

$$\begin{aligned} (\mathbf{A}_2)_{ij} = \lambda_2 \int_{\Omega_2} \nabla \phi_i \nabla \phi_j dV, \;\; i, j = 1, 2, ..., N-1, \\ (\mathbf{M}_2)_{ij} = \alpha_2 \int_{\Omega_2} \phi_i \phi_j dV, \;\; i, j = 1, 2, ..., N-1. \end{aligned}$$

The required data from the interface is inserted in the equations by the matrices $\mathbf{A}_{I\Gamma}^{(2)}$ and $\mathbf{M}_{I\Gamma}^{(2)}$. However, the system (4)-(7) is not enough to describe (1). Extending the approach (4)-(7) for the unsteady transmission problem, we will look for an approximation of the normal derivatives on Γ. For this, we consider the interface discretization point $\mathbf{u}_\Gamma$ to be discretized with FDM with respect to the first equation in (1) for $m = 1$ and with FEM with respect to the first equation in (1) for $m = 2$.

On one hand, we use centered finite differences to discretize $\mathbf{u}_\Gamma$ for both first and second order derivatives because this choice forces second order accuracy at the interface point [13, pp. 31]. For this, we need to introduce another unknown $u_2(x_1, t)$ to get

$$\begin{aligned} \lambda_1 \frac{\partial u_1}{\partial n_1} \approx \frac{\lambda_1}{2\Delta x}(u_1(x_N, t) - u_2(x_1, t)), \\ \alpha_1 \dot{\mathbf{u}_\Gamma} + \frac{\lambda_1}{\Delta x^2}(u_1(x_N, t) - 2\mathbf{u}_\Gamma + u_2(x_1, t)) = 0. \end{aligned} \tag{8}$$

Removing $u_2(x_1, t)$ from (8) we get for the following approximation for the normal derivative:

$$\mu_{\text{FDM}} = \mathbf{M}_{\Gamma\Gamma}^{(1)} \dot{\mathbf{u}}_\Gamma + \mathbf{A}_{\Gamma\Gamma}^{(1)} \mathbf{u}_\Gamma + \mathbf{A}_{\Gamma I}^{(1)} \mathbf{u}_I^{(1)}. \tag{9}$$

On the other hand, given the local exact solution u_2, its normal derivative can be written as a linear functional by using Green's formula [20]. Thus,

$$\lambda_2 \int_\Gamma \frac{\partial u_2}{\partial n_2}\phi_j dS = \lambda_2 \int_{\Omega_2} (\Delta u_2 \phi_j + \nabla u_2 \cdot \nabla \phi_j) dV \tag{10}$$
$$= \int_{\Omega_2} (-\alpha_2 \frac{\partial u_2}{\partial t}\phi_j + \lambda_2 \nabla u_2 \cdot \nabla \phi_j) dV.$$

And now, letting j run over the nodes on Γ we obtain the following expressions for the normal derivative:

$$\mu_{\text{FEM}} = \mathbf{M}^{(2)}_{\Gamma\Gamma}\dot{\mathbf{u}}_\Gamma + \mathbf{M}^{(2)}_{\Gamma I}\dot{\mathbf{u}}^{(2)}_I + \mathbf{A}^{(2)}_{\Gamma\Gamma}\mathbf{u}_\Gamma + \mathbf{A}^{(2)}_{\Gamma I}\mathbf{u}^{(2)}_I. \tag{11}$$

Consequently, the equation

$$\mu_{\text{FDM}} + \mu_{\text{FEM}} = 0 \tag{12}$$

completes the system (4)-(7).
Now, we reformulate the coupled equations (4), (7) and (12) into an ODE for the vector of unknowns $\mathbf{u} = (\mathbf{u}^{(1)}_I, \mathbf{u}^{(2)}_I, \mathbf{u}_\Gamma)^T$

$$\tilde{\mathbf{M}}\dot{\mathbf{u}} + \tilde{\mathbf{A}}\mathbf{u} = 0 \tag{13}$$

where

$$\tilde{\mathbf{M}} = \begin{pmatrix} \alpha_1\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \mathbf{M}^{(2)}_{I\Gamma} \\ \mathbf{0} & \mathbf{M}^{(2)}_{\Gamma I} & \mathbf{M}^{(1)}_{\Gamma\Gamma} + \mathbf{M}^{(2)}_{\Gamma\Gamma} \end{pmatrix}, \quad \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{A}^{(1)}_{I\Gamma} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{A}^{(2)}_{I\Gamma} \\ \mathbf{A}^{(1)}_{\Gamma I} & \mathbf{A}^{(2)}_{\Gamma I} & \mathbf{A}^{(1)}_{\Gamma\Gamma} + \mathbf{A}^{(2)}_{\Gamma\Gamma} \end{pmatrix}.$$

2.2 Time Discretization

Applying the implicit Euler method with time step Δt to the system (13), we get for the vector of unknowns $\mathbf{u}^{n+1} = (\mathbf{u}^{(1),n+1}_I, \mathbf{u}^{(2),n+1}_I, \mathbf{u}^{n+1}_\Gamma)^T$

$$\mathbf{A}\mathbf{u}^{n+1} = \mathbf{u}^n \tag{14}$$

where $\mathbf{u}^n = (\alpha_1\mathbf{u}^{(1),n}_I,\ \mathbf{M}_2\mathbf{u}^{(2),n}_I + \mathbf{M}^{(2)}_{I\Gamma}\mathbf{u}^n_\Gamma,\ \mathbf{M}^{(2)}_{\Gamma I}\mathbf{u}^{(2),n}_I + \mathbf{M}_{\Gamma\Gamma}\mathbf{u}^n_\Gamma)^T$ and

$$\mathbf{A} = \tilde{\mathbf{M}} - \Delta t\tilde{\mathbf{A}} = \begin{pmatrix} \alpha_1\mathbf{I} - \Delta t\mathbf{A}_1 & \mathbf{0} & -\Delta t\mathbf{A}^{(1)}_{I\Gamma} \\ \mathbf{0} & \mathbf{M}_2 - \Delta t\mathbf{A}_2 & \mathbf{M}^{(2)}_{I\Gamma} - \Delta t\mathbf{A}^{(2)}_{I\Gamma} \\ -\Delta t\mathbf{A}^{(1)}_{\Gamma I} & \mathbf{M}^{(2)}_{\Gamma I} - \Delta t\mathbf{A}^{(2)}_{\Gamma I} & \mathbf{M}_{\Gamma\Gamma} - \Delta t\mathbf{A}_{\Gamma\Gamma} \end{pmatrix},$$

with $\mathbf{M}_{\Gamma\Gamma} = \mathbf{M}^{(1)}_{\Gamma\Gamma} + \mathbf{M}^{(2)}_{\Gamma\Gamma}$ and $\mathbf{A}_{\Gamma\Gamma} = \mathbf{A}^{(1)}_{\Gamma\Gamma} + \mathbf{A}^{(2)}_{\Gamma\Gamma}$.

3 FIXED POINT ITERATION

We now employ a standard Dirichlet-Neumann iteration to solve the discrete system (14). This corresponds to alternatively solving the discretized equations of the transmission problem (1) on Ω_1 with Dirichlet data on Γ and the discretization of (1) on Ω_2 with Neumann data on Γ. Due to the Dirichlet boundary condition, we can compute the solution of (1) on Ω_1 explicitly. However, a system of two equations is needed to find the solution on Ω_2 and Γ simultaneously due to the Neumann boundary condition.
Applying this to (14), one gets for the k-th iteration the two equation systems

$$(\alpha_1 \mathbf{I} - \Delta t \mathbf{A}_1)\mathbf{u}_I^{(1),n+1,k+1} = \Delta t \mathbf{A}_{I\Gamma}^{(1)} \mathbf{u}_\Gamma^{n+1,k} + \alpha_1 \mathbf{u}_I^{(1),n}, \tag{15}$$

$$\hat{\mathbf{A}}\hat{\mathbf{u}}^{k+1} = \hat{\mathbf{u}}^k, \tag{16}$$

to be solved in succession. Here,

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{M}_2 - \Delta t \mathbf{A}_2 & \mathbf{M}_{I\Gamma}^{(2)} - \Delta t \mathbf{A}_{I\Gamma}^{(2)} \\ \mathbf{M}_{\Gamma I}^{(2)} - \Delta t \mathbf{A}_{\Gamma I}^{(2)} & \mathbf{M}_{\Gamma\Gamma}^{(2)} - \Delta t \mathbf{A}_{\Gamma\Gamma}^{(2)} \end{pmatrix}, \quad \hat{\mathbf{u}}^{k+1} = \begin{pmatrix} \mathbf{u}_I^{(2),n+1,k+1} \\ \mathbf{u}_\Gamma^{n+1,k+1} \end{pmatrix}$$

and

$$\hat{\mathbf{u}}^k = \begin{pmatrix} \mathbf{M}_2 \mathbf{u}_I^{(2),n} + \mathbf{M}_{I\Gamma}^{(2)} \mathbf{u}_\Gamma^n \\ \Delta t \mathbf{A}_{\Gamma I}^{(1)} \mathbf{u}_I^{(1),n+1,k+1} - (\mathbf{M}_{\Gamma\Gamma}^{(1)} - \Delta t \mathbf{A}_{\Gamma\Gamma}^{(1)}) \mathbf{u}_\Gamma^{n+1,k} + \mathbf{M}_{\Gamma I}^{(2)} \mathbf{u}_I^{(2),n} + \mathbf{M}_{\Gamma\Gamma} \mathbf{u}_\Gamma^n \end{pmatrix}$$

with some initial condition, here $\mathbf{u}_\Gamma^{n+1,0} = \mathbf{u}_\Gamma^n$. The iteration is terminated according to the standard criterion $\|\mathbf{u}_\Gamma^{k+1} - \mathbf{u}_\Gamma^k\| \leq \tau$ where τ is a user defined tolerance [2].
We now rewrite (15)-(16) as an iteration for $\mathbf{u}_\Gamma^{n+1}$. To this end, we isolate the term $\mathbf{u}_I^{(1),n+1,k+1}$ from (15) and $\mathbf{u}_I^{(2),n+1,k+1}$ from the first equation in (16):

$$\mathbf{u}_I^{(1),n+1,k+1} = (\alpha_1 \mathbf{I} - \Delta t \mathbf{A}_1)^{-1} (\Delta t \mathbf{A}_{I\Gamma}^{(1)} \mathbf{u}_\Gamma^{n+1,k} + \alpha_1 \mathbf{u}_I^{(1),n}), \tag{17}$$

$$\mathbf{u}_I^{(2),n+1,k+1} = (\mathbf{M}_2 - \Delta t \mathbf{A}_2)^{-1} (-(\mathbf{M}_{I\Gamma}^{(2)} - \Delta t \mathbf{A}_{I\Gamma}^{(2)}) \mathbf{u}_\Gamma^{n+1,k+1} + \mathbf{M}_2 \mathbf{u}_I^{(2),n} + \mathbf{M}_{I\Gamma}^{(2)} \mathbf{u}_\Gamma^n). \tag{18}$$

Inserting (17) and (18) into the second equation of (16) one obtains the iteration $\mathbf{u}_\Gamma^{n+1,k+1} = \Sigma \mathbf{u}_\Gamma^{n+1,k} + \psi$, with iteration matrix

$$\Sigma = -{\mathbf{S}^{(2)}}^{-1} \mathbf{S}^{(1)}, \tag{19}$$

where

$$\mathbf{S}^{(1)} = (\mathbf{M}_{\Gamma\Gamma}^{(1)} - \Delta t \mathbf{A}_{\Gamma\Gamma}^{(1)}) - \Delta t^2 \mathbf{A}_{\Gamma I}^{(1)} (\alpha_1 \mathbf{I} - \Delta t \mathbf{A}_1)^{-1} \mathbf{A}_{I\Gamma}^{(1)}, \tag{20}$$

$$\mathbf{S}^{(2)} = (\mathbf{M}_{\Gamma\Gamma}^{(2)} - \Delta t \mathbf{A}_{\Gamma\Gamma}^{(2)}) - (\mathbf{M}_{\Gamma I}^{(2)} - \Delta t \mathbf{A}_{\Gamma I}^{(2)})(\mathbf{M}_2 - \Delta t \mathbf{A}_2)^{-1} (\mathbf{M}_{I\Gamma}^{(2)} - \Delta t \mathbf{A}_{I\Gamma}^{(2)}), \tag{21}$$

and ψ are other terms non dependent on $\mathbf{u}_\Gamma^{n+1,k}$.
Thus, the Dirichlet-Neumann iteration is a linear iteration and the rate of convergence is described by the spectral radius of the iteration matrix.

4 ANALYSIS

In this section, we study the iteration matrix Σ for an specific FDM-FEM discretizations. We will give an exact formula which computes the convergence rates. The behavior of the rates when approaching both the continuous case in time and space are also given.
Specifically, we consider $\Omega_1 = [0,1]$, $\Omega_2 = [1,2]$ and the standard piecewise-linear polynomials

$$\phi_k(x) := \left\{ \begin{array}{cc} \frac{x-x_{k-1}}{x_k-x_{k-1}}, & \text{if } x_{k-1} < x \leq x_k \\ \frac{x_{k+1}-x}{x_{k+1}-x_k}, & \text{if } x_k < x \leq x_{k+1} \\ 0, & \text{otherwise} \end{array} \right\} \tag{22}$$

as test functions on Ω_2.
If we consider $\mathbf{e}_j = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^T \in \mathbb{R}^N$ where the only nonzero entry is located at the j-th position, the discretization matrices are given by

$$\mathbf{A}_m = \frac{\lambda_m}{\Delta x^2} \begin{pmatrix} -2 & 1 & & 0 \\ 1 & -2 & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & -2 \end{pmatrix}, \quad \mathbf{M}_2 = \frac{\alpha_2}{6} \begin{pmatrix} 4 & 1 & & 0 \\ 1 & 4 & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 4 \end{pmatrix},$$

$$\mathbf{M}^{(1)}_{\Gamma\Gamma} = \frac{\alpha_1}{2}, \ \mathbf{M}^{(2)}_{\Gamma\Gamma} = \frac{2\alpha_2}{6}, \ \mathbf{A}^{(m)}_{\Gamma\Gamma} = -\frac{\lambda_m}{\Delta x^2}, \ m = 1,2.$$
$$\mathbf{A}^{(1)}_{I\Gamma} = \frac{\lambda_1}{\Delta x^2}\mathbf{e}_N, \ \mathbf{A}^{(2)}_{I\Gamma} = \frac{\lambda_2}{\Delta x^2}\mathbf{e}_1, \ \mathbf{M}^{(2)}_{I\Gamma} = \frac{\alpha_2}{6}\mathbf{e}_1,$$
$$\mathbf{A}^{(1)}_{\Gamma I} = \frac{\lambda_1}{\Delta x^2}\mathbf{e}_N^T, \ \mathbf{A}^{(2)}_{\Gamma I} = \frac{\lambda_2}{\Delta x^2}\mathbf{e}_1^T, \ \mathbf{M}^{(2)}_{\Gamma I} = \frac{\alpha_2}{6}\mathbf{e}_1^T.$$

where $\Delta x = 1/(N+1)$ and $\mathbf{A}_m$, $\mathbf{M}_2 \in \mathbb{R}^{N\times N}$, $\mathbf{A}^{(m)}_{I\Gamma}$, $\mathbf{M}^{(2)}_{I\Gamma} \in \mathbb{R}^{N\times 1}$ and $\mathbf{A}^{(m)}_{\Gamma I}$, $\mathbf{M}^{(2)}_{\Gamma I} \in \mathbb{R}^{1\times N}$ for $m = 1,2$.
Note that the iteration matrix Σ is just a real number in this case and thus its spectral radius is its modulus. The goal now is to compute $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$. Inserting the corresponding matrices specified in (19) we have

$$\begin{aligned} \mathbf{S}^{(1)} &= \left(\frac{\alpha_1}{2} + \Delta t \frac{\lambda_1}{\Delta x^2}\right) - \Delta t^2 \frac{\lambda_1^2}{\Delta x^4} \mathbf{e}_N^T (\alpha_1 \mathbf{I} - \Delta t \mathbf{A}_1)^{-1} \mathbf{e}_N \\ &= \left(\frac{\alpha_1}{2} + \Delta t \frac{\lambda_1}{\Delta x^2}\right) - \Delta t^2 \frac{\lambda_1^2}{\Delta x^4} \alpha^1_{NN}, \end{aligned} \tag{23}$$

$$\begin{aligned} \mathbf{S}^{(2)} &= \left(\frac{\alpha_2}{3} + \Delta t \frac{\lambda_2}{\Delta x^2}\right) - \left(\frac{\alpha_2}{6} - \Delta t \frac{\lambda_2}{\Delta x^2}\right)^2 \mathbf{e}_1^T (\mathbf{M}_2 - \Delta t \mathbf{A}_2)^{-1} \mathbf{e}_1 \\ &= \left(\frac{\alpha_2}{3} + \Delta t \frac{\lambda_2}{\Delta x^2}\right) - \left(\frac{\alpha_2}{6} - \Delta t \frac{\lambda_2}{\Delta x^2}\right)^2 \alpha^2_{11}, \end{aligned} \tag{24}$$

where α_{ij}^1 represent the entries of the matrix $(\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1)^{-1}$ and α_{ij}^2 the entries of the matrix $(\mathbf{M}_m - \Delta t\mathbf{A}_m)^{-1}$ for $i, j = 1, ..., N$. Observe that the matrices $(\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1)$ and $(\mathbf{M}_2 - \Delta t\mathbf{A}_2)$ are tridiagonal Toeplitz matrices but their inverses are full matrices. The computation of the exact inverses is based on a recursive formula which runs over the entries [6] and consequently, it is not trivial how to find α_{NN}^1 and α_{11}^2 this way.
Due to these difficulties, we propose to rewrite the matrices $(\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1)^{-1}$ and $(\mathbf{M}_m - \Delta t\mathbf{A}_m)^{-1}$ in terms of their eigendecomposition:

$$(\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1)^{-1} = \left[\text{tridiag}\left(-\frac{\lambda_1\Delta t}{\Delta x^2}, \frac{\alpha_1\Delta x^2 + 2\lambda_1\Delta t}{\Delta x^2}, -\frac{\lambda_1\Delta t}{\Delta x^2}\right)\right]^{-1} = \mathbf{V}\Lambda_1^{-1}\mathbf{V}, \tag{25}$$

$$(\mathbf{M}_2 - \Delta t\mathbf{A}_2)^{-1} = \left[\text{tridiag}\left(\frac{\alpha_2\Delta x^2 - 6\lambda_2\Delta t}{6\Delta x^2}, \frac{2\alpha_2\Delta x^2 + 6\lambda_2\Delta t}{3\Delta x^2}, \frac{\alpha_2\Delta x^2 - 6\lambda_2\Delta t}{6\Delta x^2}\right)\right]^{-1} = \mathbf{V}\Lambda_2^{-1}\mathbf{V}, \tag{26}$$

where the matrix $\mathbf{V}$ has the eigenvectors of any symmetric tridiagonal matrix as a columns and the matrices Λ_1, Λ_2 are a diagonal matrices having the eigenvalues of $\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1$ and $\mathbf{M}_2 - \Delta t\mathbf{A}_2$ as entries. These are known and given e.g. in [15, pp. 514-516]

$$\begin{aligned} v_{ij} &= \frac{1}{\sum_{k=1}^N \sin^2\left(\frac{k\pi}{N+1}\right)} \sin\left(\frac{ij\pi}{N+1}\right) \quad \text{for } i, j = 1, ..., N, \\ \lambda_j^1 &= \frac{1}{\Delta x^2}\left(\alpha_1\Delta x^2 + 2\lambda_1\Delta t - 2\lambda_1\Delta t\cos\left(\frac{j\pi}{N+1}\right)\right), \\ \lambda_j^2 &= \frac{1}{3\Delta x^2}\left(2\alpha_2\Delta x^2 + 6\lambda_2\Delta t + (\alpha_2\Delta x^2 - 6\lambda_2\Delta t)\cos\left(\frac{j\pi}{N+1}\right)\right) \\ &\qquad \text{for } j = 1, ..., N \;\; m = 1, 2. \end{aligned} \tag{27}$$

The entries α_{NN}^1 and α_{11}^2 of the matrices $(\alpha_1\mathbf{I} - \Delta t\mathbf{A}_1)^{-1}$ and $(\mathbf{M}_2 - \Delta t\mathbf{A}_2)^{-1}$, respectively, are now computed through their eigendecomposition resulting in

$$\alpha_{NN}^1 = \frac{\sum_{i=1}^N \frac{1}{\lambda_i^1}\sin^2\left(\frac{i\pi N}{N+1}\right)}{\sum_{i=1}^N \sin^2\left(\frac{i\pi}{N+1}\right)} = \frac{s_1}{\sum_{i=1}^N \sin^2(i\pi\Delta x)}, \tag{28}$$

$$\alpha_{11}^2 = \frac{\sum_{i=1}^N \frac{1}{\lambda_i^2}\sin^2\left(\frac{i\pi}{N+1}\right)}{\sum_{i=1}^N \sin^2\left(\frac{i\pi}{N+1}\right)} = \frac{s_2}{\sum_{i=1}^N \sin^2(i\pi\Delta x)}, \tag{29}$$

with

$$s_1 = \sum_{i=1}^N \frac{\Delta x^2 \sin^2(i\pi\Delta x)}{\alpha_1\Delta x^2 + 2\lambda_1\Delta t(1 - \cos(i\pi\Delta x))}, \tag{30}$$

$$s_2 = \sum_{i=1}^N \frac{3\Delta x^2 \sin^2(i\pi\Delta x)}{2\alpha_2\Delta x^2 + 6\lambda_2\Delta t + (\alpha_2\Delta x^2 - 6\lambda_2\Delta t)\cos(i\pi\Delta x)}. \tag{31}$$

Now, inserting (28) and (29) into (23) and (24) we get for $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$,

$$\mathbf{S}^{(1)} = \frac{\alpha_1\Delta x^2 + 2\lambda_1\Delta t}{2\Delta x^2} - \frac{\lambda_1^2\Delta t^2}{\Delta x^4}\frac{s_1}{\sum_{i=1}^{N}\sin^2(i\pi\Delta x)}, \tag{32}$$

$$\mathbf{S}^{(2)} = \frac{\alpha_2\Delta x^2 + 3\lambda_2\Delta t}{3\Delta x^2} - \frac{(\alpha_2\Delta x^2 - 6\lambda_2\Delta t)^2}{36\Delta x^4}\frac{s_2}{\sum_{i=1}^{N}\sin^2(i\pi\Delta x)}. \tag{33}$$

With this we obtain an explicit formula for the spectral radius of the iteration matrix Σ as a function of Δx and Δt:

$$\begin{aligned}\rho(\Sigma) = |\Sigma| = |{\mathbf{S}^{(2)}}^{-1}\mathbf{S}^{(1)}|\\ = \left(\frac{\alpha_2\Delta x^2 + 3\lambda_2\Delta t}{3\Delta x^2} - \frac{(\alpha_2\Delta x^2 - 6\lambda_2\Delta t)^2}{36\Delta x^4}\frac{s_1}{\sum_{i=1}^{N}\sin^2(i\pi\Delta x)}\right)^{-1}\\ \cdot\left(\frac{\alpha_1\Delta x^2 + 2\lambda_1\Delta t}{2\Delta x^2} - \frac{\lambda_1^2\Delta t^2}{\Delta x^4}\frac{s_2}{\sum_{i=1}^{N}\sin^2(i\pi\Delta x)}\right).\end{aligned} \tag{34}$$

To simplify this, the finite sum $\sum_{i=1}^{N}\sin^2(i\pi\Delta x)$ can be computed. We first rewrite the sum of squared sinus into a sum of cosinus using the identity $\sin^2(x/2) = (1-\cos(x))/2$. Then, the resulting sum can be converted into a geometric sum using Euler's formula:

$$\begin{aligned}\sum_{j=1}^{N}\sin^2(j\pi\Delta x) &= \frac{1-\Delta x}{2\Delta x} - \frac{1}{2}\sum_{j=1}^{N}\cos(2j\pi\Delta x)\\ &= \frac{1-\Delta x}{2\Delta x} - \frac{1}{2}Re\left(\sum_{j=1}^{N}e^{2ij\pi\Delta x}\right)\\ &= \frac{1-\Delta x}{2\Delta x} - \frac{1}{2}Re\left(\frac{e^{2i\pi\Delta x}\left(1-e^{2iN\pi\Delta x}\right)}{1-e^{2i\pi\Delta x}}\right)\\ &= \frac{2\Delta x\cos^2(\pi\Delta x) - 2\Delta x + 1}{2\Delta x}\end{aligned} \tag{35}$$

Inserting (35) into (34) we get after some manipulations

$$|\Sigma| = \frac{9\Delta x(\alpha_1\Delta x^2 + 2\lambda_1\Delta t)(2\Delta x\cos^2(\pi\Delta x) - 2\Delta x + 1) - 36\lambda_1^2\Delta t^2 s_1}{3\Delta x(2\alpha_2\Delta x^2 + 6\lambda_2\Delta t)(2\Delta x\cos^2(\pi\Delta x) - 2\Delta x + 1) - (\alpha_2\Delta x^2 - 6\lambda_2\Delta t)^2 s_2}. \tag{36}$$

This is a computable formula that gives exactly the convergence rates of the Dirichlet-Neumann iteration for given Δx, Δt, α_m and λ_m, $m = 1, 2$.
We are now interested in the asymptotics of (36). In particular, we want to know the behaviour of (36) when Δt or Δx tend to 0. However, the denominators of (30) and (31) become zero when Δx tends to 0. To solve this problem, we reformulate (36) in terms of $c = \Delta t/\Delta x^2$. To get that, we multiply (36) by $1/\Delta x^2$ both in the numerator and denominator getting

$$|\Sigma| = \frac{9\Delta x(\alpha_1 + 2\lambda_1 c)(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - 36\Delta x \lambda_1^2 c^2 s_1'}{3(2\alpha_2 + 6\lambda_2 c)(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - \Delta x(\alpha_2 - 6\lambda_2 c)^2 s_2'} \tag{37}$$

where

$$s_1' = \sum_{i=1}^{N} \frac{\sin^2(i\pi\Delta x)}{\alpha_1 + 2\lambda_1 c(1 - \cos(i\pi\Delta x))}, \tag{38}$$

$$s_2' = \sum_{i=1}^{N} \frac{3\sin^2(i\pi\Delta x)}{2\alpha_2 + 6\lambda_2 c + (\alpha_2 - 6\lambda_2 c)\cos(i\pi\Delta x)}. \tag{39}$$

for $m = 1, 2$.
Computing the limits of (37) when $c \to 0$ and $c \to \infty$ we get

$$\begin{aligned} \lim_{c\to 0} |\Sigma| &= \frac{9\alpha_1(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1)}{6\alpha_2(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - \Delta x \alpha_2^2 \sum_{i=1}^{N} \frac{3\sin^2(i\pi\Delta x)}{\alpha_2(2+\cos(i\pi\Delta x))}} \\ &= \frac{3\alpha_1(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1)}{\alpha_2\left(2(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - \Delta x \sum_{i=1}^{N} \frac{\sin^2(i\pi\Delta x)}{2+\cos(i\pi\Delta x)}\right)} =: \gamma, \end{aligned} \tag{40}$$

$$\begin{aligned} \lim_{c\to\infty} |\Sigma| &= \lim_{c\to\infty} \frac{18\lambda_1 c(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - 36(\lambda_1 c)^2 \sum_{i=1}^{N} \frac{\sin^2(i\pi\Delta x)}{2\lambda_1 c(1-\cos(i\pi\Delta x))}}{18\lambda_2 c(2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) - 36(\lambda_2 c)^2 \sum_{i=1}^{N} \frac{3\sin^2(i\pi\Delta x)}{6\lambda_2 c(1-\cos(i\pi\Delta x))}} \\ &= \frac{\lambda_1\left((2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) + \Delta x - 1 - \Delta x \sum_{i=1}^{N} \cos(i\pi\Delta x)\right)}{\lambda_2\left((2\Delta x \cos^2(\pi\Delta x) - 2\Delta x + 1) + \Delta x - 1 - \Delta x \sum_{i=1}^{N} \cos(i\pi\Delta x)\right)} = \frac{\lambda_1}{\lambda_2} =: \delta. \end{aligned} \tag{41}$$

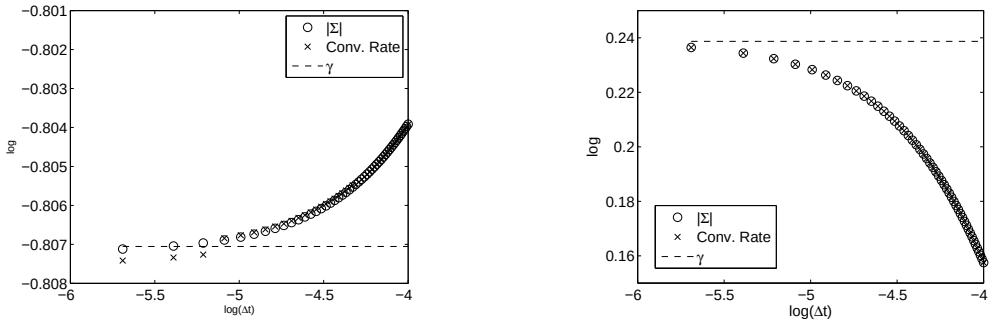
The result obtained in (41) matches with [16] and strong jumps in the thermal conductivities of the materials placed in Ω_1 and Ω_2 will imply fast convergence. This is the case when modelling thermal fluid structure interaction, where usually a compressible fluid with low thermal conductivity and density is coupled with a structure having higher thermal conductivity and density. However, the result in (40) does not only depend on the ratio of α_m, $m = 1, 2$. It also depends on Δx, and therefore, in the case of mixed discretizations more factors affect the rates. We will compare (40) with the results obtained in [16] in more details in the next section.

5 NUMERICAL RESULTS

In this section we present a set of numerical experiments designed to show how (36) computes the convergence rates. We also show that the theoretical asymptotics deduced both in (40) and (41) match with the numerical experiments.
Figure 1 shows the cases 1 and 2 specified in table 1. The circles correspond to $\rho(\Sigma)$, the crosses to the experimental convergence rates and the dashed line to the corresponding asymptotic (γ when $\Delta t \to 0$ and δ when $\Delta x \to 0$). Notice that the method does not converge for some cases. Nevertheless $|\Sigma|$ describes the convergence rate and γ and δ their asymptotics.

Case	D_1	D_2	λ_1	λ_2	δ	γ	γ_{FEM}	$1.74 \cdot \gamma_{\text{FEM}}$
1	1	0.3	0.3	1	0.3	0.26	0.15	0.16
2	0.3	1	0.3	1	0.3	1.04	0.6	1.04
3	1	0.5	0.3	1	0.3	0.16	0.09	0.16
4	0.5	1	0.3	1	0.3	1.73	1	1.74

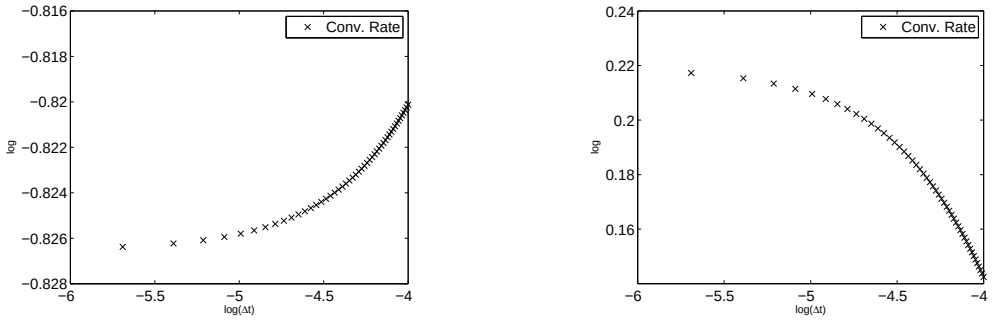
Table 1: The first four columns contain the input parameters for the different one dimensional test cases. γ and δ are the resulting limits when approaching the continuous case in time and space of the discrete estimator. γ_{FEM} is the semidiscrete limit in time specified in [16] and the last column corresponds to the relation between γ_{FEM} and γ.



(a) Case 1. (b) Case 2.

Figure 1: Cases 1 and 2 from table 1. The circles correspond to $|\Sigma|$, the crosses to the experimental convergence rates and the dashed line to γ. The curves are restricted to the discrete values $\Delta t = 1e-4/50, 2 \cdot 1e-4/50, ..., 50 \cdot 1e-4/50$ and $\Delta x = 1/20$. One observes how γ describes the behaviour of the convergence rates when we enforce the condition $\Delta t/\Delta x^2 << 1$.

Even though a 2D analysis has not been done, one can expect that a similar analysis will hold due to the results shown in figure 2. To illustrate this, figure 2 shows the experimental convergence rates for cases 1 and 2 in 2D. One observes that the rates in the 2D case have a similar behavior than in the 1D case (figure 1).

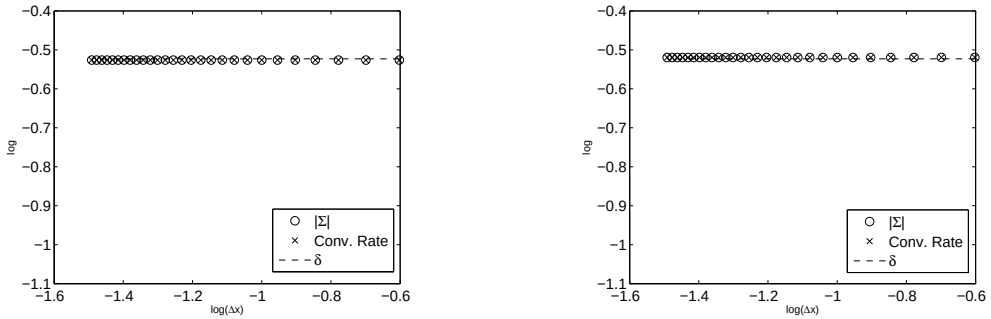


(a) Case 1. (b) Case 2.

Figure 2: The experimental convergence rates in 2D for cases 1 and 2 in table 1. The curves are restricted to the discrete values $\Delta t = 1e-4/50, 2 \cdot 1e-4/50, ..., 50 \cdot 1e-4/50$ and $\Delta x = 1/20$. A similar behaviour to the 1D case is observed here.

Now we want to test if the convergence rates when approaching the continuous case in space have dependency on the thermal diffusivities D_1 and D_2 as predicted in [9] or they do not as

in our analysis. For that, figure 3 shows the cases 1 and 2 from table 1. Here, the thermal conductivities λ_1 and λ_2 are the same in both plots but the thermal diffusivities are switched (meaning that D_1 in case 1 corresponds to D_2 in case 2 and D_2 in case 1 corresponds to D_1 in case 2). We can observe that the asymptotics of the convergence rates do not vary in both plots. This result is consistent with [17] where a similar behaviour was observed for the 2D version of the coupled unsteady transmission problem discretized with finite differences and also with [16] where a 1D and 2D convergence analysis for the unsteady transmission problem discretized with finite elements was performed. Finally, observe that the convergence rate does not vary a lot when we decrease Δx. For a fairly large choice of Δx (for instance $\Delta x = 1/10$), the convergence rates are already quite close to δ.



(a) Case 1. (b) Case 2.

Figure 3: Cases 1 and 2 from table 1. The curves are restricted to the discrete values $\Delta x = 1/3, 1/4, ..., 1/30$ and $\Delta t = 100$. One observes how δ describes the behaviour of the convergence rates when we enforce the condition $\Delta t/\Delta x^2 >> 1$.

Before ending this section, we want to present a comparison between the semidiscrete limit in time presented in this paper (γ) and the one computed in [16] when both subdomains Ω_1 and Ω_2 are discretized with finite elements (γ_{FEM}) which was defined as

$$\gamma_{\text{FEM}} := \frac{\alpha_1}{\alpha_2}. \tag{42}$$

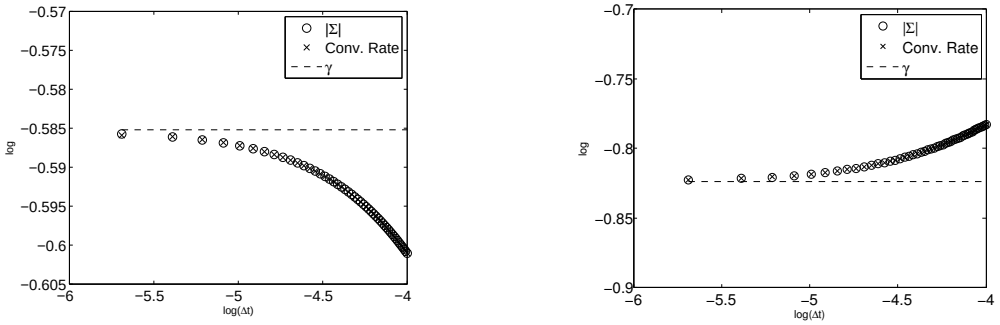
Figure 4 and 5 show the comparison between γ and γ_{FEM} for the cases 3 and 4 specified in table 1. One can observe that $\gamma > \gamma_{\text{FEM}}$, and therefore, the method will be slower using mixed discretizations (FDM - FEM) than using pure FEM discretization when approaching the continuous case in time. More in general, from the numerical experiments is possible to observe that

$$\gamma \approx \frac{c_1\alpha_1}{c_2\alpha_2} = \frac{c_1}{c_2}\gamma_{\text{FEM}} \tag{43}$$

with constants $c_1 = 9$ and $c_2 = 5.18$. This phenomena can also been checked in table 1 where this approximation has been included in the last column.

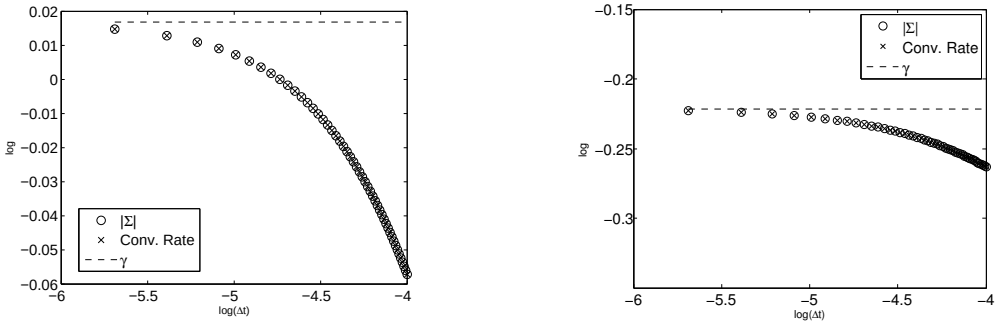
6 CONCLUSIONS AND FURTHER WORK

We have described the Dirichlet-Neumann iteration for the coupling of two heat equations on two identical domains. In particular, the coupled PDE were discretized into a system of algebraic equations using finite differences on Ω_1 and finite elements on Ω_2. Afterwards, a fixed



(a) Mixed discretizations (b) FEM discretization

Figure 4: Case 3 from table 1. Comparison between γ (left) and γ_{FEM} (right). The curves are restricted to the discrete values $\Delta t = 1e-4/50, 2\cdot 1e-4/50, ..., 50\cdot 1e-4/50$ and $\Delta x = 1/20$. Observe that $\gamma > \gamma_{\text{FEM}}$ and in particular, $\gamma \approx 1.74\gamma_{\text{FEM}}$.



(a) Mixed discretizations (b) FEM discretization

Figure 5: Case 4 from table 1. Comparison between γ (left) and γ_{FEM} (right). The curves are restricted to the discrete values $\Delta t = 1e-4/50, 2\cdot 1e-4/50, ..., 50\cdot 1e-4/50$ and $\Delta x = 1/20$. Observe that $\gamma > \gamma_{\text{FEM}}$ and in particular, $\gamma \approx 1.74\gamma_{\text{FEM}}$.

point iteration was performed and the iteration matrix was found. An exact formula describing the convergence rates is derived. Finally, the limits of the convergence rates when approaching the continuous case either in space ($\delta := \lambda_1/\lambda_2$) or time ($\gamma$) are computed. In the numerical results, we have presented four different test cases which show how the computed asymptotics predict the behaviour of the convergence rates and a comparison with the results of the pure finite element discretization of the problem in [16].

From the first four test cases we conclude that Δx does not strongly affect the convergence rates. However, they are affected by Δt. Moreover, our results show that when approaching the continuous case in space ($\Delta x \to 0$) the convergence rates do not depend on the thermal diffusivities D_1 and D_2 as predicted in [9] for the semidiscrete case. We found the analysis in [9] to be correct and are not sure where the discrepancy comes from, this is subject of further investigation.

From the comparison between γ and γ_{FEM} we can observe that $\gamma > \gamma_{\text{FEM}}$ and more specifically, $\gamma \approx 9\alpha_1/5.18\alpha_2 = 1.74\cdot\gamma_{\text{FEM}}$. This explains why the convergence of the method when approaching the continuous case in time is slower when using the mixed discretizations FDM - FEM with respect to the pure FEM discretization.

There are a variety of future directions for this work. A 2D analysis of this mixed discretization can be done. Another goal would be to analyze the convergence rates of the method using finite

volumes on one domain and finite elements on the other. This is very interesting in the context of fluid structure interaction where the fluid is usually discretized using finite volumes and the structure with finite elements. Another future direction will be to study the convergence speed of an actual non linear application.

REFERENCES

[1] Badia, S., Quaini, A. and Quarteroni, A., *Modular vs. non-modular preconditioners for fluid-structure systems with large added-mass effect*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 4216-4232.

[2] Birken, P., *Termination criteria for inexact fixed point methods*, Numer. Linear Algebra Appl., 22(4) (2015), pp. 702-716.

[3] Birken, P., Gleim, T., Kuhl, D. and Meister, A., *Fast Solvers for Unsteady Thermal Fluid Structure Interaction*, Int. J. Numer. Meth. Fluids, 79(1) (2015), pp. 16-29.

[4] Buchlin, J.M., *Convective Heat Transfer and Infrared Thermography*, J. Appl. Fluid Mech., 3 (2010), pp. 55-62.

[5] Causin, P., Gerbeau, J.F. and Nobile, F., *Added-mass effect in the design of partitioned algorithms for fluid-structure problems*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 4506-4527.

[6] Fonseca, C.M. and Petronilho, J., *Explicit inverses of some tridiagonal matrices*, Linear Algebra and its Applications, 325(1-3) (2001), pp. 7-21.

[7] Giles, M.B., *Stability Analysis of Numerical Interface Conditions in Fluid-Structure Thermal Analysis*, Int. J. Num. Meth. in Fluids, 25 (1997), pp. 421-436.

[8] Heck, U., Fritsching, U. and Bauckhage K., *Fluid flow and heat transfer in gas jet quenching of a cylinder*, International Journal of Numerical Methods for Heat and Fluid Flow, 11 (2001), pp. 36-49.

[9] Henshaw, W.D. and Chand, K.K., *A composite grid solver for conjugate heat transfer in fluid-structure systems*, Journal for Computational Physics, 228 (2009), pp. 2708-3741.

[10] Hinderks, M. and Radespiel, R., *Investigation of Hypersonic Gap Flow of a Reentry Nosecap with Consideration of Fluid Structure Interaction*, AIAA Paper, 6 (2006).

[11] Kowollik, D.S.C., Horst, P. and Haupt, M.C., *Fluid-structure interaction analysis applied to thermal barrier coated cooled rocket thrust chambers with subsequent local investigation of delamination phenomena*, Progress in Propulsion Physics, 4 (2013), pp. 617-636.

[12] Kowollik, D., Tini, V., Reese, S. and Haupt, M., *3D fluid-structure interaction analysis of a typical liquid rocket engine cycle based on a novel viscoplastic damage model*, Int. J. Numer. Meth. Engng, 94 (2013), pp. 1165-1190.

[13] LeVeque, R.J., *Finite Difference Methods for Ordinary and Partial Differential Equations*, Siam, 2007.

[14] Mehta, R.C., *Numerical Computation of Heat Transfer on Reentry Capsules at Mach 5*, AIAA-Paper, 178 (2005).

[15] Meyer, C.D., *Matrix Analysis and Applied Linear Algebra*, 2000.

[16] Monge, A. and Birken, P., *The Dirichlet-Neumann Iteration for Unsteady Thermal Fluid Structure Interaction: Convergence Analysis for Finite Element Discretizations*, SMAI Journal of Computational Mathematics, submitted.

[17] Monge, A. and Birken, P., *Convergence speed of coupling iterations for the unsteady transmission problem*, in B. Schrefler, E. Oate and M. Papadrakakis (Eds), VI International Conference on Computational Methods for Coupled Problems in Science and Engineering, COUPLED PROBLEMS 2015, pp. 452-463.

[18] Quarteroni, A. and Valli, A., *Domain Decomposition Methods for Partial Differential Equations*, Oxford Science Publications, 1999.

[19] Stratton, P., Shedletsky, I. and Lee, M., *Gas Quenching with Helium*, Solid State Phenomena, 118 (2006), pp. 221-226.

[20] Toselli, A. and Widlund, O., *Domain Decomposition Methods - Algorithms and Theory*, Springer, 2004.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

USING IMPROVED RADIAL BASIS FUNCTIONS METHODS FOR FLUID-STRUCTURE COUPLING AND MESH DEFORMATION

Giorgos A. Strofylas[1], Georgios I. Mazanakis[1], Sotirios S. Sarakinos[1], Georgios N. Lygidakis[1] and Ioannis K. Nikolos[2]

Technical University of Crete, School of Production Engineering and Management
University Campus, Chania, GR-73100, Greece
[1] e-mail: {gstrofylas, gmazanakis, ssarakinos, glygidakis}@isc.tuc.gr
[2] e-mail: jnikolo@dpem.tuc.gr

Keywords: Fluid-Structure Interaction, Radial Basis Functions, Information Transfer, Mesh Deformation, Non-matching meshes, Partitioned Coupling.

Abstract. *In this work the development of a partitioned FSI coupling procedure is reported, aiming to facilitate interaction between an open-source CSD (Computational Structural Dynamics) and an in-house academic CFD (Computational Fluid Dynamics) code. Attention is mainly directed towards the efficient and accurate transfer of predicted displacements, velocities (by CSD) and loads (by CFD). More precisely, spatial coupling is achieved using Radial Basis Functions (RBFs) interpolation, which enables point-based interaction, needing therefore no information for connectivities and, consequently, allowing for the utilization of different type or even intersecting structural and flow grids. Although RBFs method seems to be particularly attractive for both data transfer and mesh deformation, it suffers from a significant drawback; it calls for relatively excessive memory and computation time requirements (in its initial formulation). In case of data transfer the Partition of Unity (PoU) approach is adopted as a remedy of the aforementioned deficiency, which regards the decomposition of the examined problem into several smaller ones, to be solved independently and hence more efficiently. In mesh deformation though, improvement of computational performance is succeeded with a surface point reduction technique, based on the agglomeration of the adjacent boundary nodes, i.e., on the fusion of the RBFs centers. Despite the notable reduction of RBFs base points, the proposed method preserves sufficiently the quality of the initial grid. The proposed algorithm is evaluated against a benchmark (for FSI solvers) test case, considering the analysis of the wind action over a standard tall building model. The obtained numerical results confirm its potential for such simulations, highlighting additionally the radically improved computational performance of data transfer and grid deformation procedures.*

1 INTRODUCTION

During the past decades the analysis of Fluid-Structure-Interaction (FSI) with numerical methods gained significant popularity, mainly due to the broad range of its applications, e.g., in aerodynamic shape optimization, aeroelastic analysis, etc. Although several FSI solvers, commercial, industrial or academic, have already been developed, the need for more accurate and more efficient methodologies is still on the run, making FSI a continuously evolving scientific field. One of the most widely applied techniques to deal with this kind of problems is the partitioned coupling approach. According to this method the fluid and structural physical systems are solved independently, while interaction between them is obtained via the exchange of appropriate variables on boundary surfaces [1, 2]; as a result, it allows for the integration of separately developed CSD (Computational Structural Dynamics) and CFD (Computational Fluid Dynamics) applications. Nevertheless, the aforementioned solvers' combination is raising several issues in terms of information transfer, considering the main task of accurate as well as efficient FSI simulations.

For the numerical solution of FSI problems, the Dirichlet-Neumann domain decomposition coupling approach is usually followed [3]. It begins with the flow prediction, while at next the obtained pressure field at the flow side of the CFD/CSD interface is transferred to the corresponding structural one. Subsequently, a CSD analysis is performed, deriving the deflections of the structure and a new flow grid is generated incorporating the new interface location. Information transfer across the fluid-structure interface has to be performed with respect to a kinematic condition (Dirichlet), assuring the continuity of the velocity field, and a dynamic condition (Neumann) satisfying the requirement for conservation of normal stresses [4]. Hence, the most important criteria to be fulfilled over the interface, concern the consistency between the structural and the flow solver and the instabilities that may occur due to the partitioning process as well as the strategy followed for the data transfer. Similarly to stresses' conservation, an energy one has to be succeeded at the common interface [1] to ensure the stability of the whole numerical solution, i.e., the virtual work of the flow forces and this of the structural ones must be equal. The aforementioned conservation requirements across the fluid-structure interface depend strongly on both the employed temporal and spatial coupling schemes [5].

This study focus is mainly towards the spatial coupling and information transfer between the CFD and CSD solvers. The computational nodes of the structural and flow grids do not coincide usually at their interfaces, since their underlying resolutions can significantly differ. This non-matching interface may lead to gaps and intersections between them, necessitating thus for an extra interpolation step to enable information transfer. Numerous methods can be found in the literature to deal with this problem, varying from the simplistic neighbor interpolation method [6] to more sophisticated approaches, such as the weighting ones [7, 8]. More elaborate attempts include the use of projections and integrations of elements from one mesh to the other [9], while popular among researchers are the Constant Volume Tetrahedra [10, 11] and the Boundary Element (BEM) [12] methods. In this work information transfer is achieved via an interpolation scheme based on Radial Basis Functions (RBF) [13, 14], the appealing properties of which have been extensively analyzed by Beckert and Wendland in [15], e.g., the ability to inherently conserve total force and moment. RBFs methodology has another significant advantage; it does not require any connectivity information, as it relies only on arbitrary point data. As a result, it can cope both with structured and unstructured grids.

Apart from the data transfer, in FSI problems a new volume flow mesh has to be generated, depending on the boundary displacements and, consequently, on the new interface location; this procedure is typically implemented several times during an FSI simulation, necessitating thus

to be versatile, efficient, and accurate, as well as to sufficiently preserve the initial mesh topology [16]. A considerably appealing method for mesh deformation is based on the extension of RBFs interpolation, used to transfer data across the interface, in the context of mesh motion [16]. Nevertheless, the method, in its initial formulation, calls for relatively excessive memory and computation time requirements. For the alleviation of the aforementioned shortcoming the Partition of Unity (PoU) [17] approach is adopted in data transfer, which regards the decomposition of the interfacing boundary into several smaller regions, where local interpolations are performed independently and hence more efficiently. Although this method ensures energy conservation and offers physical distribution of the forces on the interfacing area [17, 18], it is revealed to be inappropriate for mesh deformation, as the dependence matrix takes into account all the flow mesh nodes, entailing significantly increased computational cost . In order to mitigate this inadequacy, a relatively recently developed by the authors methodology is followed, considering agglomeration of the surface nodes, and as such reduction of the dimensions of the pre-mentioned matrix [19]; it is based on the fusion strategy, developed for a corresponding CFD agglomeration multigrid method [20]. The method achieves significant acceleration of the deformation procedure, without downgrading the quality of the derived meshes [19].

In this work a partitioned FSI coupling procedure was developed to facilitate interaction between an open-source CSD [21] and an in-house academic CFD [20] code. Attention is mainly directed toward the efficient and accurate transfer of predicted displacements (by CSD) and forces (by CFD). Therefore, spatial coupling is obtained by a RBFs interpolation scheme, employing the aforementioned PoU approach [18] to accelerate the exchange procedure. For mesh deformation a corresponding RBFs interpolation process is applied; its efficiency is significantly improved with a surface point reduction technique, based on the agglomeration of adjacent boundary nodes. The proposed algorithm is evaluated against a benchmark (for FSI solvers) test case, considering the analysis of the wind action over a standard tall building model [22, 23]. The obtained numerical results confirm its potential for such simulations, highlighting additionally the radically improved computational performance of data transfer and grid deformation procedures.

The rest of the paper is organized as follows: In Section 2 the coupling procedure, concerning exchange of displacements and forces, is described in detail, while in Section 3 the incorporated mesh deformation process is outlined. Section 4 includes a brief description of the flow and structural models. In Section 5 the validation results of the proposed methodology over a benchmark test case are presented, while Section 6 contains some conclusions, based on the aforementioned results.

2 COUPLING PROCEDURE

The spatial coupling procedure can be mainly divided in two stages, namely, (a) the transfer of the displacement vectors from the structural to the flow grid and (b) the transfer of nodal forces from the flow mesh back to the structural one. The coupling scheme, for both aforementioned steps, can be expressed in a unified fashion by defining a coupling matrix H, which associates the interface boundary nodes of the two grids (flow and structural). Then, interpolation of the displacements can be performed according to:

$$d_s = H d_f \,, \tag{1}$$

where, d_s and d_f denote the displacements at the structural and flow interface side, respectively. For forces' association, the transpose matrix H^{T} is used as

$$f_s = H^{\mathrm{T}} f_f \,, \tag{2}$$

where f_s and f_f are the forces acting on the structural and flow sides of the interface, respectively. With the adoption of this scheme the virtual work can be effectively conserved over the interface [24, 25]; nevertheless, the conservation of total force and moment depends mainly on the choice of the coupling matrix H, and hence on the corresponding method applied to obtain it.

2.1 Radial Basis Functions (RBFs) interpolation

In this work, the RBFs technique is implemented to construct the interpolation matrix, which is used for exchanging information between the structural and the flow solver. According to the RBFs theory [15], the interpolation function is defined as a weighted sum of the evaluations of the basis functions:

$$s(\boldsymbol{x}) = \sum_{i=1}^{n_c} a_i \Phi(\|\boldsymbol{x} - \boldsymbol{x_{c_i}}\|) + p(\boldsymbol{x}) \,, \tag{3}$$

where a_i are the weight coefficients, $p(x)$ a polynomial, and $\boldsymbol{x_{c_i}} = [x_{c_i}, y_{c_i}, z_i]$ the RBFs centers, e.g., the coordinates of the boundary mesh nodes; n_c is the number of those centers and Φ is the basis function. The use of the linear polynomial term ensures the exact recovery of translations and rotations [15]. The polynomial term in x direction can be expressed as follows

$$p^x(\boldsymbol{x}_{c_i}) = \gamma_0^x + \gamma_1^x x + \gamma_2^x y + \gamma_3^x z \,, \tag{4}$$

while the coefficients a_i and γ_i are computed by interpolation conditions, fulfilling Equation (3) requirements at the centers of the basis functions, along with an additional condition, described as

$$s(\boldsymbol{x}_{c_i}) = \boldsymbol{U}_{c_i} = \left[u_{c_i}^x, u_{c_i}^y, u_{c_i}^z\right]^T \,, \tag{5}$$

$$\sum_{i=1}^{n_c} a_i q(\boldsymbol{x_{c_i}}) = 0 \,, \tag{6}$$

where $\boldsymbol{U}_{c_i}$ are the known boundary displacements of the structural side of the interface.

Several examples of different basis functions can be found in the literature [13, 15, 26], which can be broadly classified to functions with global and local support. In the first case the functions influence the whole domain, while in the latter ones the variable to be interpolated is scaled with a support radius R ($\xi = x/R$), indicating the region of influence of the function from each center [26]. Moreover, it should be noted that smoother basis functions result in more realistic deformations, which in turn provoke dense matrix systems entailing numerical instability. In this work Wendland's C^2 basis function with local support radius was adopted, due to its robustness and accuracy [16].

To construct the coupling matrix H, the structural mesh displacements of the boundary surfaces are used; combining Equations (5) and (6), the coefficient vectors $\boldsymbol{a^{x,y,z}}$ and $\boldsymbol{b^{x,y,z}}$ can be evaluated by solving the following linear system in each spatial direction (expressed in matrix notation in the x direction)

$$\begin{bmatrix} \boldsymbol{U_x} \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{M_{cc}} & \boldsymbol{P_c} \\ \boldsymbol{P_c^T} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{a^x} \\ \boldsymbol{b^x} \end{bmatrix} = \boldsymbol{C}_{cc} \begin{bmatrix} \boldsymbol{a^x} \\ \boldsymbol{b^x} \end{bmatrix} , \tag{7}$$

where

$$\boldsymbol{a^x} = \begin{bmatrix} a_{c_1}^x \\ a_{c_2}^x \\ \vdots \\ a_{c_n}^x \end{bmatrix}, \; \boldsymbol{b^x} = \begin{bmatrix} \gamma_0^x \\ \gamma_1^x \\ \gamma_2^x \\ \gamma_3^x \end{bmatrix}, \tag{8}$$

and

$$\boldsymbol{P_c} = \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{c_n} & y_{c_n} & z_{c_n} \end{bmatrix}, \; \boldsymbol{M_{cc}} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1c_n} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2c_n} \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{c_n1} & \Phi_{c_n2} & \cdots & \Phi_{c_nc_n} \end{bmatrix}, \tag{9}$$

with $\Phi_{ij} = \Phi\left(\left\|x_i - x_j\right\|\right)$.

The required approximation of displacements at the flow side of the interface are calculated by applying the following equation for each spatial direction:

$$s(x_a) = \boldsymbol{A}_{ac}\boldsymbol{C}_{cc}^{-1}x_c = \boldsymbol{H}x_c, \tag{10}$$

where

$$\boldsymbol{A}_{ac} = \begin{bmatrix} 1 & x_{\alpha_1} & y_{\alpha_1} & z_{\alpha_1} & \Phi_{\alpha_1 c_1} & \Phi_{\alpha_1 c_2} & \cdots & \Phi_{\alpha_1 c_n} \\ 1 & x_{\alpha_2} & y_{\alpha_1} & z_{\alpha_1} & \Phi_{\alpha_2 c_1} & \Phi_{\alpha_2 c_2} & \cdots & \Phi_{\alpha_2 c_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{\alpha_{N_\alpha}} & y_{\alpha_{N_\alpha}} & z_{\alpha_{N_\alpha}} \Phi_{\alpha_{N_\alpha} c_1} & \Phi_{\alpha_{N_\alpha} c_2} & \cdots & \Phi_{\alpha_{N_\alpha} c_n} \end{bmatrix}, \tag{11}$$

with a_i being the i-th flow mesh node of the boundary and N_α the total number of boundary nodes of the same grid. If the following term is additionally set

$$\widetilde{M}_{cc} = (P_c M_{cc}^{-1} P_c^T)^{-1}, \tag{12}$$

then the coupling matrix H can be expressed as:

$$\boldsymbol{H} = \boldsymbol{A}_{ac} \begin{bmatrix} \widetilde{M}_{cc} P_c M_{cc}^{-1} \\ M_{cc}^{-1} - M_{cc}^{-1} P_c^T \widetilde{M}_{cc} P_c M_{cc}^{-1} \end{bmatrix}. \tag{13}$$

Following the coupling process described above, it is evident that the computation of the coupling matrix is obtained by solving the linear systems described in Equation (7), while the interpolation of displacements and forces from the one mesh to the other involve only matrix-vector multiplications. However, the method suffers from excessive computational and memory requirements, especially when large-scale problems are encountered, since the coupling matrix $\boldsymbol{H}$ is of size N_s x N_α, where N_s denotes the number of the structural mesh surface nodes and N_α the number of the flow mesh surface nodes. Computational efficiency is further degraded due to the necessity to invert the (n_c+4)x(n_c+4) matrix $\boldsymbol{C_{cc}}$. A thorough literature review of methods that deal with the aforementioned deficiency can be found in [19]; the boundary node coarsening [19, 27, 28] technique seems to be particularly attractive for the mesh deformation process (as it will be analyzed in the next section). However, it is considered inappropriate for the data transfer one, as it might result in loss of information and, as such, jeopardize the requirement of force conservation over the interface.

Another approach, also focusing on the efficiency improvement of the method, is the Partition of Unity (PoU) scheme, according to which the initial problem is divided into several smaller ones by grouping base points into sets to be processed separately; it resembles the domain decomposition approach, employed for the parallelization of complex numerical methods. It has been successfully implemented against FSI problems in [17, 18] with very promising results. Besides reducing computational cost, it avoids the undesirable influence of support radius into the interpolation process.

2.2 Partition Of Unity (PoU)

The basic concept of the PoU method regards the degradation of the initial problem into several smaller ones to be solved independently, while the global solution arises by a weighted

combination of the local ones. Given this condition, the term "unity" refers to the summation of the weighting coefficients of the entire domain that must be equal to one. The process begins by dividing the set Ω of the boundary nodes, including the nodes of the interface of both computational meshes (structural and flow) into a K number of mutually overlapping regions, in a way to ensure that each Ω_j patch includes only a small number of points. For each of these patches a local displacement is computed by solving a local interpolation problem with the RBFs method. Subsequently, the global interpolation is obtained by summing all local interpolations, weighted by their respective coefficients, as

$$s(\boldsymbol{x}) = \sum_{j=1}^{K} w_j(\boldsymbol{x}) s_j(\boldsymbol{x}), \quad \sum_{j=1}^{K} w_j(\boldsymbol{x}) = 1, \tag{14}$$

where $w_j(\boldsymbol{x})$ represents a non-negative weighting function, which comes into effect only for its associated patches Ω_j, while it is zeroed for the rest. Using the transpose of the global matrix, formed by the local interpolants, assures that the conservation criteria are also met in the PoU approach [17].

A variation of the pre-described method, proposed by Rendall and Allen [18], concerns the application of the PoU method on each boundary flow node separately, instead of decomposing the set of nodes into overlapping patches. Following this strategy, the complex step of decomposing the structural grid into overlapping regions is evaded, and thus the possibility of facing discontinuities between the patches due to the partitioning process is diminished. In order to calculate the local interpolants, an association of each flow mesh node with a pre-defined number of its closest structural boundary ones has to be defined. Therefore, the use of sophisticated data structures appears to be mandatory in order to link the boundary nodes of the two grids (flow and structural) in an efficient manner. Octree data structures are a common tactic for addressing these kind of problems. After the definition of the local interpolants, their merging back to a global one is essential in order to avoid distortion and non-realistic deformations, which may occur in applications involving computational meshes with different resolutions. In addition, the implementation of such a local strategy ameliorates the force distribution over the interface; the area of influence of each flow node is limited only to its nearby ones of the structural grid, unlike the corresponding area used with the complete method.

3 MESH DEFORMATION

The formulation of RBFs in case of mesh motion is different from this of data transfer, since the two problems have different requirements. In case of data transfer the inclusion of polynomial terms of at least first order is mandatory to ensure conservation of total force and moment, while for mesh deformation these functions are omitted to avoid its undesirable global coverage, which can cause the whole domain to move rather than deform. Hence, the dependence matrix $\boldsymbol{C}_{cc}$ is reduced to $\boldsymbol{M}_{cc}$, and the weighting coefficients are obtained via the following expression

$$\boldsymbol{a}^{x} = \boldsymbol{M}_{cc}^{-1} \boldsymbol{x}_{c_i}\,, \tag{15}$$

with $\boldsymbol{a}^{x,y,z}$ denoting the vector containing the scalar coefficients in each direction and $\boldsymbol{M}_{\boldsymbol{cc}}$ the matrix defined in Equation (9).

The required displacements of the volume mesh nodes are calculated according to those of the boundary points. As such, the deformation of the volume mesh can be computed as follows

$$\boldsymbol{s}(x_a) = \boldsymbol{A}_{cv} \boldsymbol{M}_{cc}^{-1} \boldsymbol{x}_{c_i} = \boldsymbol{H} \boldsymbol{x}_{c_i}\,, \tag{16}$$

where index v represents the volume nodes of the flow grid, and $\boldsymbol{A}_{\boldsymbol{cv}}$ a $n_v \times n_c$ matrix, described as:

$$\boldsymbol{A_{cv}} = \begin{bmatrix} \boldsymbol{\Phi_{11}} & \boldsymbol{\Phi_{12}} & \cdots & \boldsymbol{\Phi_{1n_c}} \\ \boldsymbol{\Phi_{21}} & \boldsymbol{\Phi_{22}} & \cdots & \boldsymbol{\Phi_{2n_c}} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Phi_{n_v1}} & \boldsymbol{\Phi_{n_v2}} & \cdots & \boldsymbol{\Phi_{n_vn_c}} \end{bmatrix}. \tag{17}$$

In order to overcome the increased computational cost of the method, a relatively recently developed by the authors methodology is followed, which considers the agglomeration of the adjacent boundary nodes [19]. The idea behind this strategy is the generation of successively coarser grids via the fusion of adjacent control volumes of the finer ones. The construction of the control areas of the boundary nodes is based on a node-centered scheme, where each node's control area is defined by connecting the barycenters of surrounding triangles with the mid-points of triangle edges. It is this fusion process that can be implemented on the deformed solid boundary control cells, along with an RBFs-based mesh deformation technique, to successively reduce the set of surface points and, consequently, the dimensions of equation systems to be solved.

4 COMPUTATIONAL METHODS

4.1 Flow Solver

The simulation of unsteady incompressible flow phenomena is performed with the use of a recently developed academic code, named *Galatea-I* [20, 29, 30], which employs the Navier-Stokes equations, coupled with the artificial compressibility method [31-33]. The discretization of the governing equations is succeeded with a node-centered finite-volume scheme on unstructured hybrid computational grids, comprised of tetrahedra, prisms and pyramids [20, 29]. The calculation of the inviscid fluxes is performed with the Roe's approximate Riemann solver [34], along with a second-order spatial accurate scheme based on the MUSCL approach [35]. For the calculation of the viscous fluxes the gradients of velocity components have to be evaluated at the interfaces of the control cells. Two methodologies have been incorporated in *Galatea-I*; an element-based approach, utilizing the edge-dual control volume scheme and a simpler nodal-averaging one [20, 29]. For turbulence prediction the RANS (Reynolds-Averaged Navier-stokes) approach is implemented, along with the SST (Shear Stress Transport) turbulence model [36]. At the inlet and outlet regions characteristics-based boundary conditions are applied [30], while at the solid walls free-slip or no-slip conditions can be applied depending on the type of the encountered flow (inviscid or viscous). For the time accurate computation a dual-time stepping scheme has been included [31]. According to this scheme the relaxation of the governing equations is achieved with a four-stage Runge-Kutta (RK(4)) method over pseudo-time [37], while the time accurate solution is obtained with the use of a second-order finite-difference scheme over real-time [31]. Besides using edge-based data structures and a local time-stepping technique [35], *Galatea-I* solver has incorporated the capability of parallel processing, based on the domain decomposition approach [20], and the implementation of an agglomeration multigrid scheme, based on the FAS (Full Approximation Scheme) approach [20], for the acceleration of the solution procedure. With the first method the initial computational grid is divided into smaller sub-domains, each of which is attributed on a single computer core [20]. Interaction between neighboring partitions is achieved with the MPI (Message Passing Interface) library functions. With the latter methodology successively coarser grids are generated, derived by the fusion of adjacent control volumes. The relaxation of the governing equations is performed on all meshes in a V-cycle type that accelerates solution convergence [20].

4.2 Structural Solver

As far as the structural computations are concerned, the open-source software package Calculix [21] is utilized, which employs Finite-Element Method (FEM) for linear or non-linear calculations on structured or unstructured three-dimensional computational grids. The solver has the capability to encounter a variety of mechanical, thermal, coupled thermo-mechanical and contact problems in a serial or parallel computational environment; for the latter one either MPI (Message Passing Interface) or OpenMP library functions can be used.

5 NUMERICAL RESULTS

The assessment of the pre-described methodologies was performed with the Commonwealth Advisory Aeronautical Research Council (CAARC) Standard Tall Building Model, which is in fact a specification for wind tunnel experiments on tall building geometries [22, 38]. The full-scale geometry of the building is specified as a rectangular shaped prism with base size equal to 100ft (30.8m) by 150ft (45.72m) and height 600ft (183.88m). It is described as a flat-topped building without parapets; its external walls are flat without any geometric disturbances [22]. The case has been encountered by several researchers, who have performed numerical tests of wind flow around the building, as well as FSI simulations [23, 39-41]; the proposed geometry has been tested in wind tunnel experiments too [22, 42]. A sketch of the geometric model that was prepared for use with the *Galatea-I* solver is shown in Figure 1.

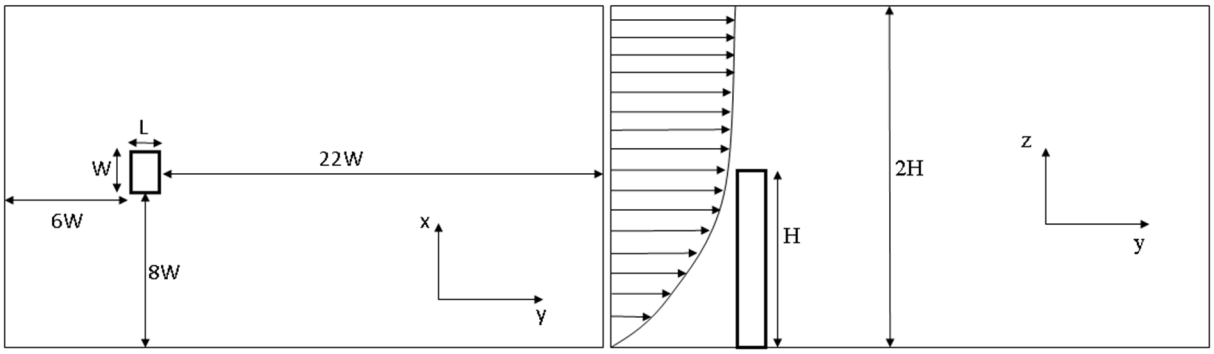


Figure 1: Sketch of CAARC test case setup.

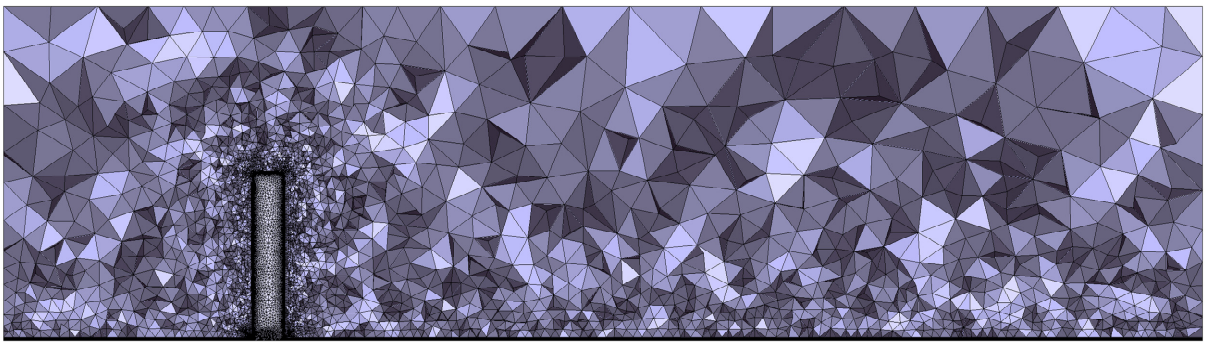


Figure 2: Computational mesh around the CAARC standard tall building model.

The computational mesh, generated for the simulation of wind flow around the CAARC building, is presented in Figure 2. As the flow around the building develops in three-dimensional fashion and is unsteady in nature, the whole building geometry had to be meshed. In order to reduce computation time a relatively coarse mesh was generated that consisted of 434,082 nodes, 623,801 tetrahedra and 640,685 prisms. Prismatic inflation was applied on the

building walls, as well as at the base of the geometry that represents the ground around the building, in order no slip boundary conditions to be applied at these areas and the corresponding boundary layer to be effectively predicted. The prismatic inflation consisted of 20 layers with the first layer height equal to W/2500. At the top and side walls of the geometry free slip conditions were imposed.

The Reynolds number was set equal to 380,000, based on the building height H and the velocity c_H at height z=H [40]. In order to simulate wind flow at such a low height, the velocity profile at the inlet was defined by the following relation

$$\frac{|\vec{c}|}{c_H} = \left(\frac{z}{H}\right)^{a}, \tag{18}$$

where α is the power law exponent, which in this case was set equal to 0.3. Moreover, the turbulence boundary conditions had to be calculated, based on the turbulence intensity in the approaching flow, as it has a significant effect on the accuracy of the numerical results [23, 39, 40]. In Figure 3 the velocity profile at the inlet, as well as the turbulence intensity profile are presented. The values of turbulence intensity at the inlet of the computational domain were interpolated from those presented in Figure 3 [39]. Finally, all values were normalized with H and c_H for use with the dimensionless solver.

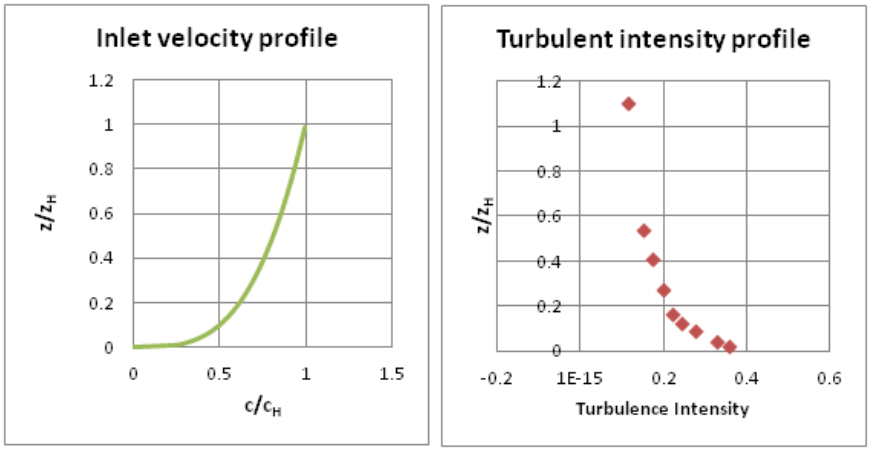


Figure 3: Inlet velocity profile (left) and turbulence intensity profile (right).

The simulation was performed on a workstation with an AMD FX™ 8150 Eight-Core processor at 3.62 Ghz. Acceleration via parallel processing was applied by partitioning the initial mesh into six sub-domains, while due to the small size of the computational mesh the multigrid method was not used. The time-accurate equations were solved over a total of 1300 real time iterations, with a dimensionless real time step equal to 0.05. In each real time step 50 pseudo-time iterations were performed, with a CFL number equal to 0.5. The artificial compressibility parameter was set equal to 10.0. For the quantitative evaluation of the extracted results the mean-pressure coefficient on the wall of the building was calculated from the obtained time-accurate results. In Figure 4 the distribution of the mean-pressure coefficient along the building wall at height z=2/3H is compared with the corresponding ones found in the literature [39, 40]. While a very good agreement is achieved between the current and reference data at the windward and leeward sides of the building, there seems to be some disagreement between numerical and experimental results at the side of the building. This is attributed to the method of turbulence evaluation; a more accurate method (e.g., LES-Large Eddy Simulation) could provide more accurate results at these regions.

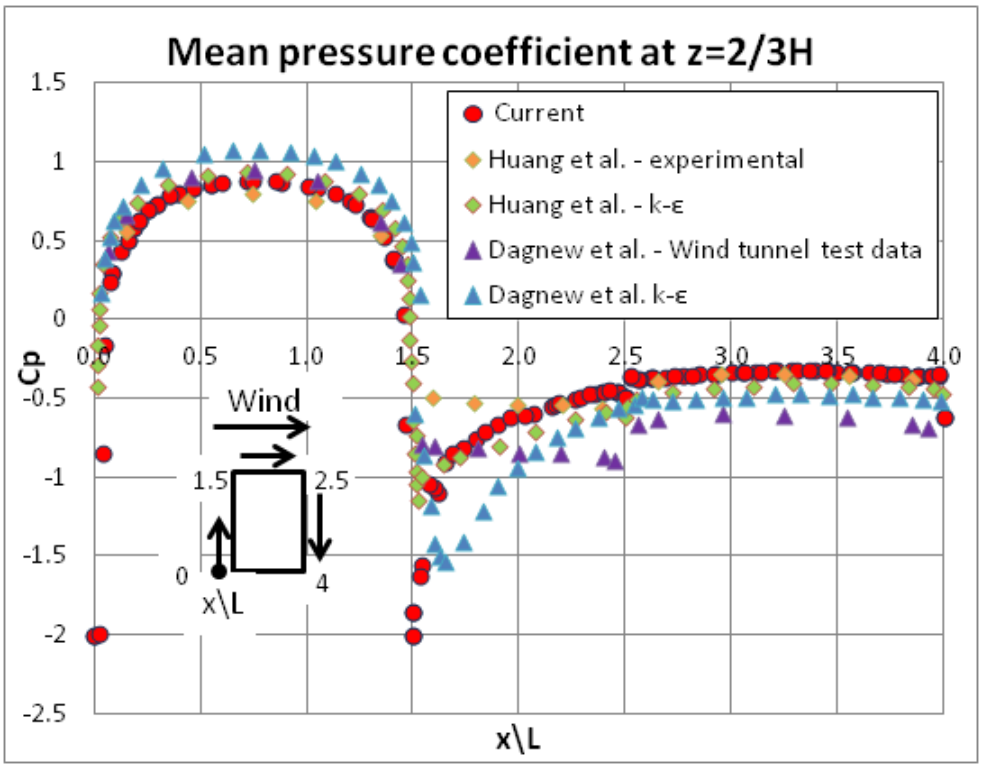


Figure 4: Mean pressure coefficient on the CAARC building wall at z=3/2H.

In Figure 5 the mean pressure coefficient distribution on the building walls and the "ground" around it is presented. The maximum pressure value is located near the building top at the windward side, while under-pressure is developed at the leeward side of the building.

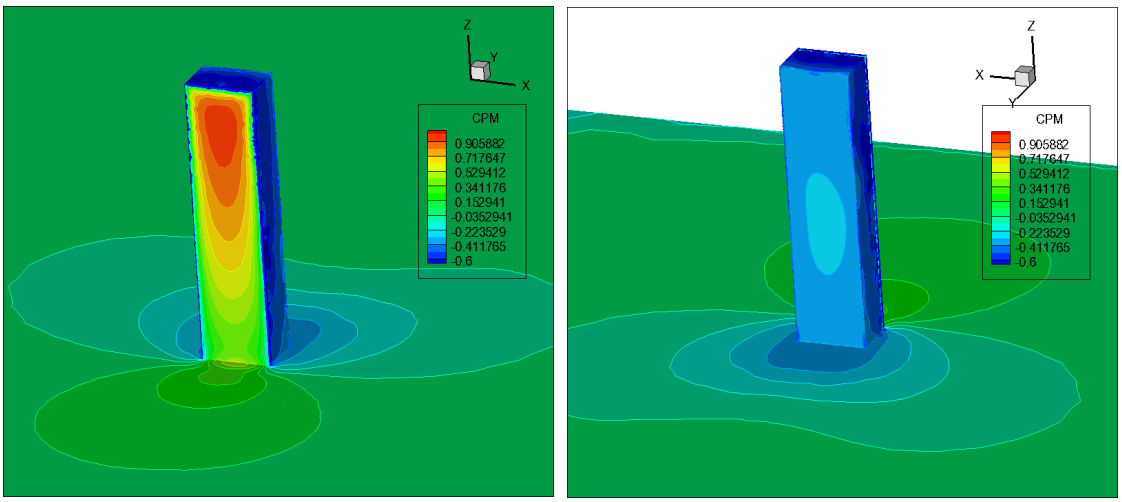


Figure 5: Mean pressure coefficient distribution on the CAARC building walls and surrounding "ground".

The force field around the CAARC standard tall building model was then calculated from the dimensionless results provided by the *Galatea-I* solver with regard to the flow conditions, reported by Braun et al. [23]. These properties of the flow can be found in Table 1. In order to use them with the CSD code, the exported force field had to be interpolated between the two different grids, i.e., the unstructured and the structural one. For the coupling of the two grids with the RBFs PoU method, the nearest nodes of the structural mesh have to be identified for each aerodynamic grid point at the building wall surface. Nevertheless, such an identification

is not a straightforward procedure; a brutal linear search method would require an enormous amount of time to perform, especially in the case of very dense meshes. Therefore, custom Octree data structures were developed to find the nearest structural grid point to each aerodynamic one, while at next the remaining nodes required by the proposed methodology were gathered with the aid of the structural mesh topology. Applying this approach, whereby the number of structural points is given *a priori*, ensures the sparseness of the coupling matrix, as the user can specify the number of these points according to the problem needs. Due to this PoU method, a support radius needs not be defined; it is set equal to the distance of each aerodynamic grid point from the uttermost in the specific patch.

Specific mass – ρ [Kgr/m^3]	1.25
Dynamic viscosity – μ [N·s/m^2]	$7.03 \cdot 10^{-2}$
Reference velocity – c_H [m/s]	100
Reference length – H	180

Table 1: Flow field properties.

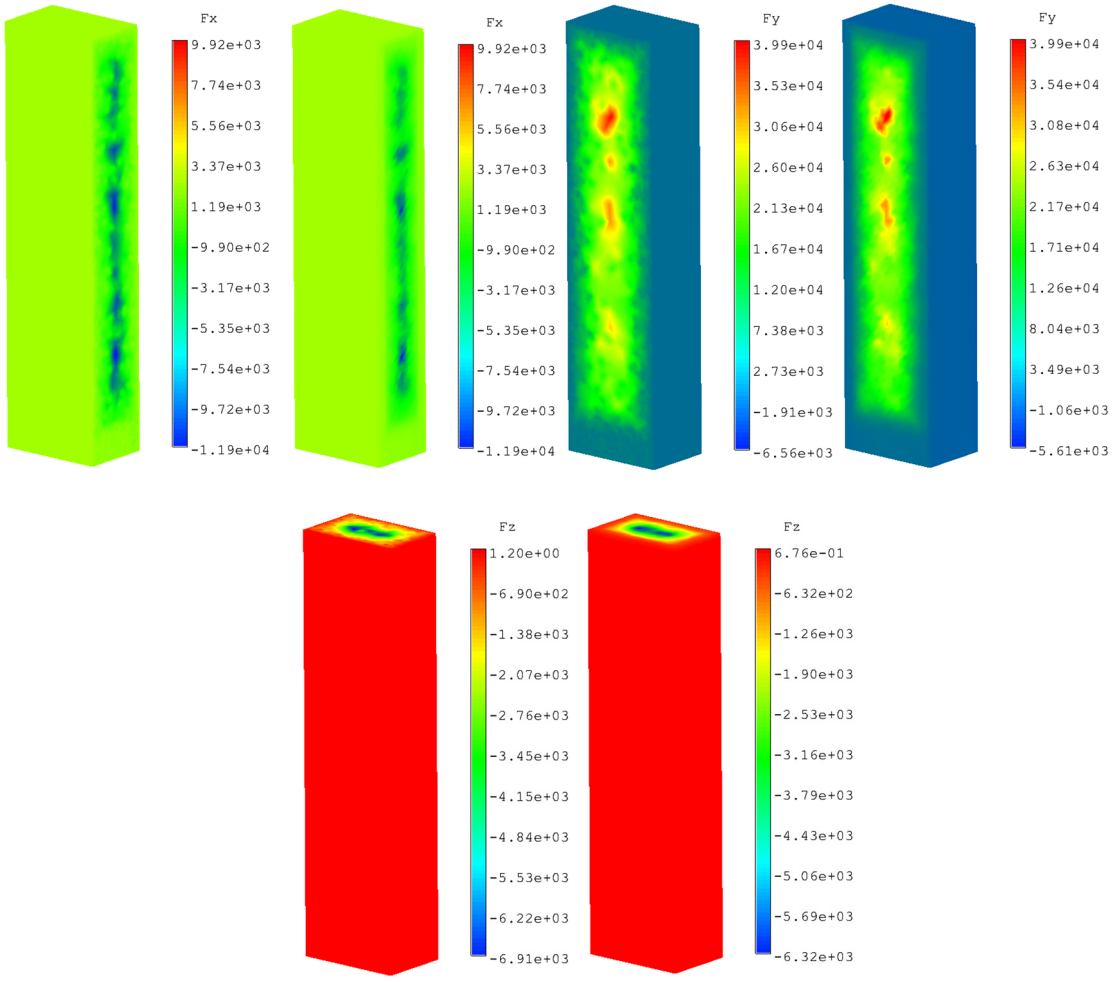


Figure 6: Comparison of force component distribution between CFD (left) and CSD (right) meshes.

The utilized structural grid is composed of *38,675* nodes and *43,048* hexahedral eight-node brick elements, fully integrated (2x2x2 integration points). For each boundary node of the CFD mesh the 20 closest points of the CSD mesh, were used, while for each set of CSD nodes a RBFs problem was set deriving the corresponding required coefficients. While the coupling of

the CFD with the CSD mesh may be a rather tedious procedure, it only has to be completed once (at the beginning of the FSI simulation) even for a dynamic problem. The computational cost of the coupling procedure is slightly affected by the size of the structural mesh, as the number of RBFs problems to be solved remains the same; only the nearest nodes identification process will differ. Comparisons between the force components distributions over the boundary surfaces, for the flow and structural computational grids, are illustrated in Figure 6.

These qualitative results reveal the ability of the proposed methodology to provide accurate transfer of the force field in an efficient manner though. As it can be observed, the PoU approach offers the advantage of physical distribution of forces over the interface, as opposed to the full method, where the transpose of the coupling matrix would destroy the local character of the force distribution [17].

With the loading (i.e., the force distribution) provided by the CFD code, a modal analysis was performed with the CalculiX open source software. The structural properties of the model were based on the CAARC standard tall building specifications [22], while the Young's modulus was set equal to that reported by Braun et al. [23]. All the employed structural properties are given in Table 2.

Specific mass – ρ	160 Kg/m^3
Natural frequency – n	0.2 Hz
Young's modulus – E	2.3 ·10^8 N/m^2
Poisson's ratio - m	0.25
Structural damping - ξ	1%

Table 2: Mechanical properties of the structure.

In Figure 7 the von Mises stresses and the stresses at the three main axes are presented, while in Figure 8(a) the deformation of the structural mesh according to the results of the aforementioned analysis is illustrated (magnified x10) along with the un-deformed boundary surface of the fluid mesh. The maximum displacement was 46.9 cm on the y axis (as expected) because of the influence of the first bending mode. Moreover, in Figure 8(b) the deformed boundary surface of the CFD mesh obtained by transferring the displacement field from the CSD one, is depicted; no distortion was observed on the boundary surface of the flow grid, resulting in a smooth deformed surface. The normalized along-wind displacement (σ_y/L) as a function of the reduced velocity (c_H/nW), obtained with the proposed methodology, is compared to the corresponding results of other reference works in Table 3; a satisfactory agreement is observed.

Current	0.015633
Braun et al. [23]	0.011595
Thepmongkorn et al. [43]	0.013835
Melbourne [22]	0.048892

Table 3: Comparison of normalized maximum displacement as a function of reduced velocity.

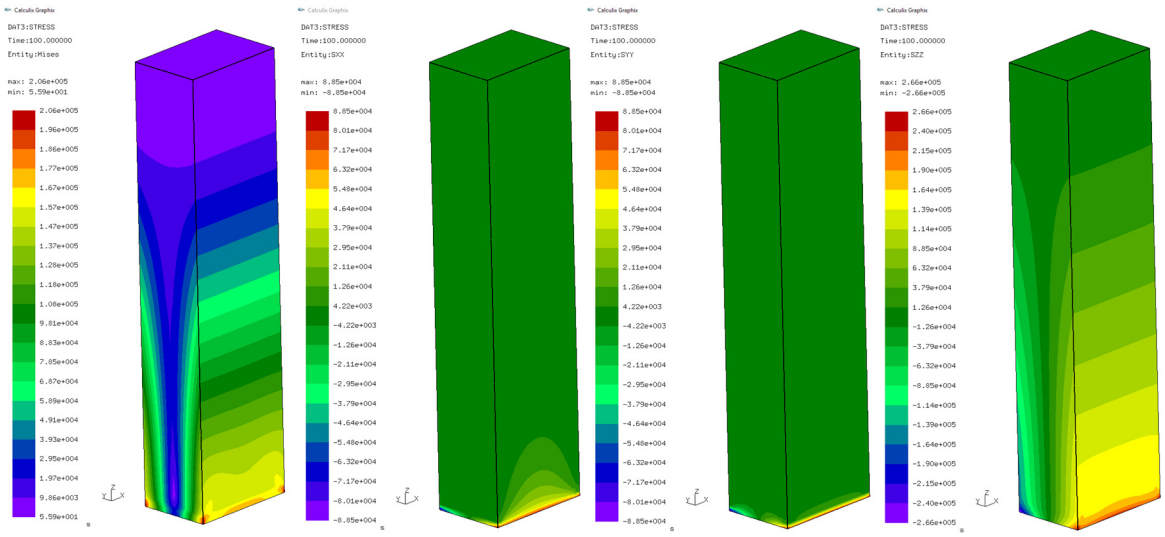
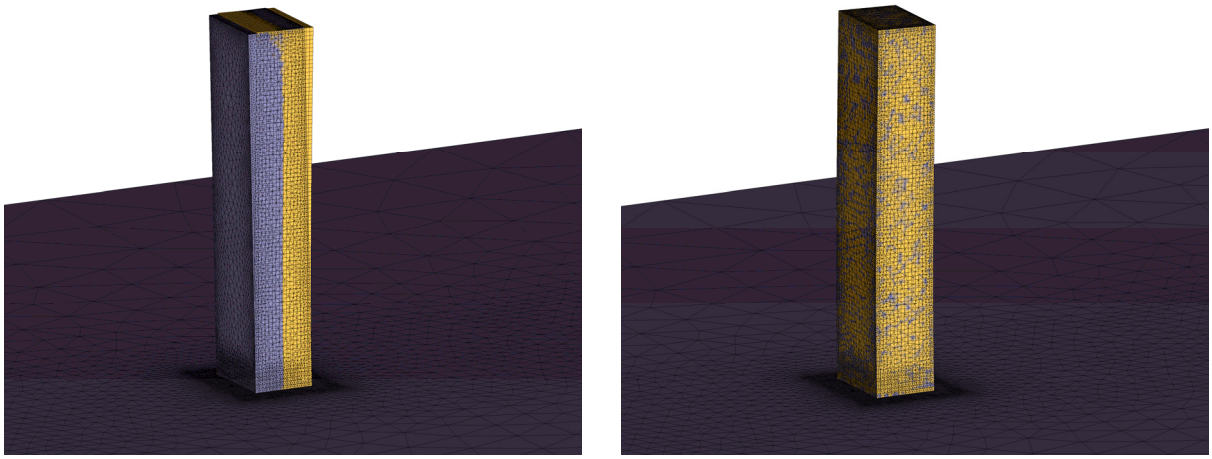
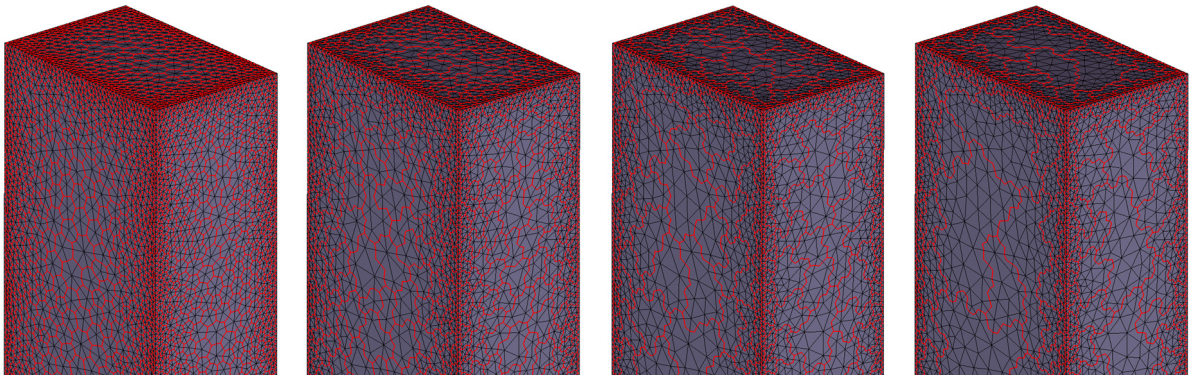


Figure 7: Von Misses, S_{XX}, S_{YY}, S_{ZZ} stresses of the CAARC standard tall building model.



(a) (b)

Figure 8: Deformation of the structural grid (a) and corresponding deformation of the CFD boundary grid (b).

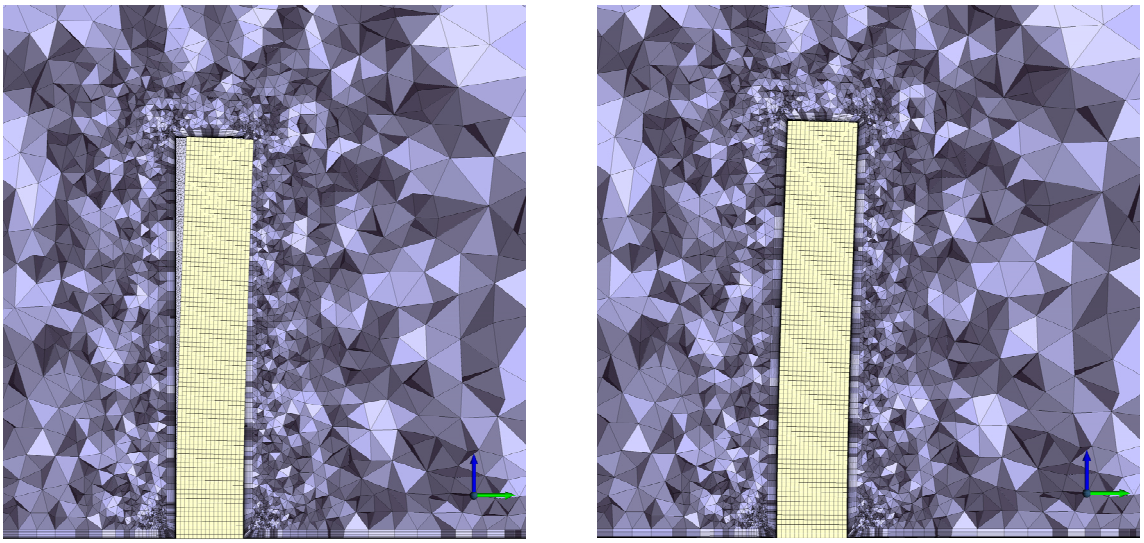


Initial (*9,724* nodes) Level 1 (*3,494* nodes) Level 2 (*2,199* nodes) Level 3 (*1,995* nodes)

Figure 9: Fine and coarser grids generated with the agglomeration procedure.

Since this stage was accomplished, the deformation of the flow grid had to be performed, according to the deflections of the boundary. For the acceleration of the procedure four coarser grids were obtained, implementing the proposed agglomeration process on the surface nodes

[19]. In Figure 9 the boundary control areas of each level are depicted. For all of them the node with the minimum distance from their area-weighted centers is selected to be used as the base point for the RBFs method [19]. These basis points form the interpolant, required for the final deformation of the volume grid. In Figure 10, the flow grid prior and after the implementation of the deformation procedure is illustrated.



(a) (b)

Figure 10: Before (a) and after (b) deformation states of the CFD grid.

The deformed CFD grid was subsequently used with the *Galatea-I* solver for a new flow simulation with the same flow conditions as with the original one. The derived mean pressure coefficient distribution around the CAARC standard tall building, along with the velocity streamlines are illustrated in Figure 11. The successfully performed simulation demonstrates the deformation methodology capabilities in providing quality grids.

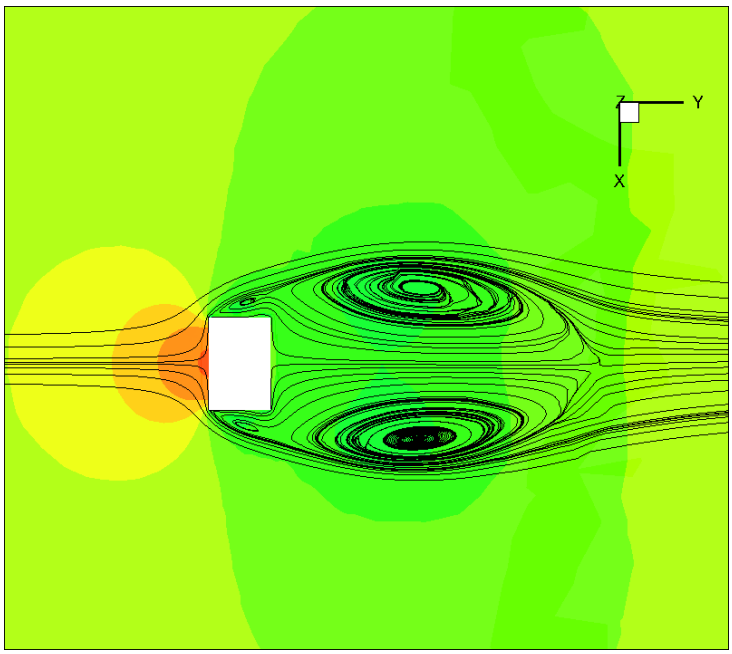


Figure 11: Mean pressure coefficient and velocity streamlines around the deformed CFD grid.

The aforementioned analysis was based on a specification for a tall building model which however differs a lot from a realistic one. For the completion of the current work, the results from a dynamic analysis of a similar 45-storey building are presented. The building has the same external geometry as the CAARC standard tall building model, but the interior is not solid. It is consisted of beams, slabs, columns, walls, and a lift core. Moreover, two different materials, steel for the structural members (green color), and glass for all the outer walls (red color) were used (Figure 12). Table 4 presents the properties of the steel and glass material.

	Glass	Steel
Specific mass – ρ	2500 Kg/m^3	7850 Kg/m3
Natural frequency – n	0.2 Hz	0.2 Hz
Young's modulus – E	$7 \cdot 10^{10}$ N/m^2	$2.1 \cdot 10^{11}$ N/m^2
Poisson's ratio - m	0.25	0.25
Structural damping - ξ	1%	1%

Table 4: Mechanical properties of the structure.

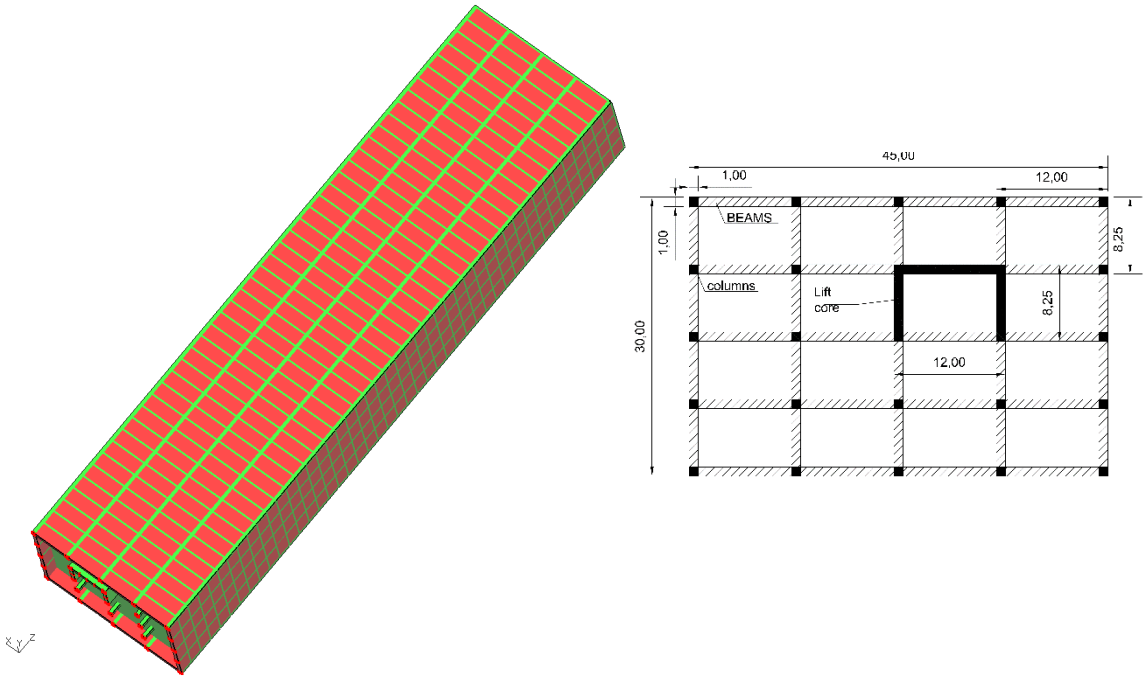


Figure 12: Realistic CAARC building model (structural members: green color, walls: red color).

In this case a dynamic analysis takes place, in which the wind loading is considered the same to the previous one (Figure 6). The deformation is also similar to the previous case but the displacements are much smaller, an expected, taking into account that the elements properties are stiffer. Figure 13 presents the extracted von Mises stresses as well as the three main stresses. Additionally, comparing to the previous problem, the stresses appear to be more realistic. The columns produce larger stresses than the walls, because of their larger stiffness, while as the height increases the stresses get smaller. In general, larger stresses were extracted (comparing to the previous case) as the same force was imposed at a smaller surface though.

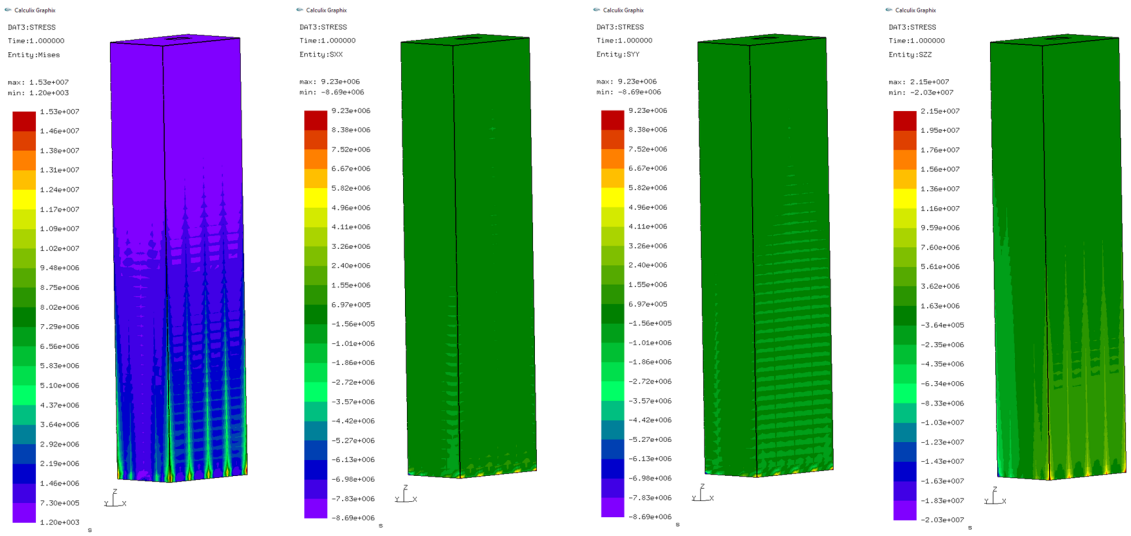


Figure 13: Von Misses, S_{XX}, S_{YY}, S_{ZZ} stresses of the CAARC realistic building model.

6 CONCLUSIONS

In this paper a recently developed partitioned FSI methodology was reported. The whole procedure involved the coupling of an in-house academic CFD code, named *Galatea-I* [20], with the Calculix open source CSD software [21], which was succeeded employing the Radial Basis Functions (RBFs) based Partition of Unity (PoU). With this approach the conservation of energy, momentum and force is ensured over the interface of the flow and structural mesh as a result of the radial functions properties. The use of the PoU methodology improves the efficiency of the data transfer procedure providing simultaneously a physical formulation of the force distribution. The FSI methodology was applied on the CAARC standard tall building model, which was specified for similar procedures [22, 23]. The effect of the obtained pressure field around the building was successfully transferred to the structural mesh with the PoU procedure, deriving subsequently the desired deformation; the latter agreed sufficiently with the corresponding experimental and numerical results, being available in the literature. As a result, the capabilities of the proposed methodology for such test cases in terms of accuracy and efficiency were demonstrated. A mesh deformation approach, based on the RBFs methodology but improved though by a fusion procedure, resembling the agglomeration procedure used in agglomeration multigrid methodology [19], was applied on the initial CFD mesh. The resulting deformed mesh, was subsequently imported in the *Galatea-I* solver, successfully providing a new flow field and hence demonstrating the deformation method's capabilities of generating quality meshes.

ACKNOWLEDGMENTS

This study was funded by Wacker Bauwerksaerodynamik GmbH, DENCO Developments & Engineering Consultants S.A. (DENCO S.A.) and the EUREKA-EUROSTARS Project No E!7987 ''Model-Based Aeroelastic Analysis of Long-Span Bridges on the HPC Cloud (BridgeCloud)''.

REFERENCES

[1] C. F. Serge Piperno, Partitioned procedures for the transient solution of coupled aeroelastic problems part I: Model problem, theory and two-dimensional application. *Computer Methods in Applied Mechanics and Engineering*, **124**, 79–112, 1995.

[2] C. Farhat, M. Lesoinne, and P. LeTallec, Load and motion transfer algorithms for fluid/structure interaction problems with non-matching discrete interfaces: Momentum and energy conservation, optimal discretization and application to aeroelasticity. *Computer Methods in Applied Mechanics and Engineering*, **157.1**, 95–114, 1998.

[3] A. Quarteroni and A. Valli, *Domain decomposition methods for partial differential equations*, Vol. 10. Clarendon Press, 1999.

[4] C. A. Felippa and K. C. Park, Staggered transient analysis procedures for coupled mechanical systems: Formulation. *Computer Methods in Applied Mechanics and Engineering*, **24**, 61–111, 1980.

[5] A. de Boer, A. H. van Zuijlen, and H. Bijl, Radial basis functions for interface interpolation and mesh deformation. *Advanced Computational Methods in Science and Engineering*, B. Koren and K. Vuik, Eds. Springer Berlin Heidelberg, 143–178, 2009.

[6] P. Thévenaz, T. Blu, and M. Unser, Interpolation revisited. *IEEE Transactions Medical Imaging*, **19.7**, 739–758, 2000.

[7] N. Maman and C. Farhat, Matching fluid and structure meshes for aeroelastic computations: A parallel approach. *Computers & Structures*, **54.4**, 779–785, 1995.

[8] P. Masarati and P. Mantegazza, A conservative mesh-free approach for fluid structure problems. *Coupled Problems,* 24–27, *Santorini, 2005.*

[9] A. de Boer, A. H. van Zuijlen, and H. Bijl, Review of coupling methods for non-matching meshes. *Computer Methods in Applied Mechanics and Engineering*, **196.8**, 1515–1525, 2007.

[10] G. S. L. Goura, K. J. Badcock, M. A. Woodgate, and B. E. Richards, A data exchange method for fluid-structure interaction problems. *Aeronautical Journal*, **105**, 215–221, 2001.

[11] M. Sadeghi and F. Liu, Coupled fluid-structure simulation for turbomachinery blade rows. *43rd Aerospace Sciences Meeting and Exhibit*, American Institute of Aeronautics and Astronautics, Nevada, 1995.

[12] P. Chen and I. Jadic, Interfacing of fluid and structural models via innovative structural boundary element method. *AIAA Journal*, **36.2**, 282–287, 1998.

[13] M. J. Smith, D. H. Hodges, and C. E. Cesnik, An evaluation of computational algorithms to interface between CFD and CSD methodologies. *American Institute of Aeronautics and Astronautics, AIAA 96-1400*, 1995.

[14] M. J. Smith, D. H. Hodges, and C. E. S. Cesnik, Evaluation of computational algorithms suitable for fluid-structure interactions. *Journal of Aircraft*, **37.2**, 282–294, 2000.

[15] A. Beckert and H. Wendland, Multivariate interpolation for fluid-structure-interaction problems using radial basis functions. *Aerospace Science and Technology*, **5.2**, 125–134, 2001.

[16] A. de Boer, M. S. van der Schoot, and H. Bijl, Mesh deformation based on radial basis function interpolation. *Computers & Structures*, **85.11–14**, 784–795, 2007.

[17] A. B. R. Ahrem, A meshless spatial coupling scheme for large-scale fluid-structure interaction problems. *Computer Modeling in Engineering & Sciences*, **12.2**, 2006.

[18] T. C. S. Rendall and C. B. Allen, Improved radial basis function fluid–structure coupling via efficient localized implementation. *International Journal for Numerical Methods in Engineering*, **78.10**, 1188–1208, 2009.

[19] G. A. Strofylas, G.N. Lygidakis, I.K Nikolos, Accelarating RBF-Based Mesh Deformation by Implementing An Agglomeration Strategy. *ASME International Mechanical Engineering Congress & Exposition (IMECE 2015),* Houston, Texas, 2015, IMECE2015-50902.

[20] G. N. Lygidakis, S. S. Sarakinos, I. K. Nikolos. Comparison of different agglomeration multigrid schemes for compressible and incompressible flow simulations, *Advances in Engineering Software*, in press, 2016.

[21] G. Dhondt, *CalculiX CrunchiX User's Manual*, version 2.7, 2014.

[22] W. H. Melbourne. Comparison of measurements on the CAARC standard tall building model in simulated model wind flows, *Journal of Wind Engineering and Industrial Aerodynamics*, **6.1**, 73-88, 1980.

[23] A. L. Braun, A. M. Awruch. Aerodynamic and aeroelastic analyses on the CAARC standard tall building model using numerical simulation, *Computers & Structures,* **87.9**, 564-581, 2009.

[24] M. H. L. Hounjet and J. J. Meijer, Evaluation of elastomechanical and aerodynamic data transfer methods for non-planar configurations in computational aeroelastic analysis. *National Aerospace Laboratory NLR*, 11.1–11.24, 1995.

[25] A. Beckert, Coupling fluid (CFD) and structural (FE) models using finite interpolation elements. *Aerospace Science and Technology*, **4.1**, 13–22, 2000.

[26] H. Wendland, *Scattered data approximation,* Cambridge University Press, 2004.

[27] S. Jakobsson and O. Amoignon, Mesh deformation using radial basis functions for gradient-based aerodynamic shape optimization. *Computers & Fluids*, **36.6**, 1119–1136, 2007.

[28] T. C. S. Rendall and C. B. Allen, Efficient mesh motion using radial basis functions with data reduction algorithms. *Journal of Computational Physics*, **228.17**, 6231–6249, 2009.

[29] G. N. Lygidakis, S. S. Sarakinos, I. K.Nikolos. A parallel agglomeration multigrid method for incompressible flow simulations, *Proceedings of the 9th International Conference on Engineering Computational Technology (ECT 2014)*, Naples, Italy, 2014.

[30] S.S. Sarakinos, G.N. Lygidakis, I.K. Nikolos, Evaluation of a parallel agglomeration multigrid finite-volume algorithm, named Galatea-I, for the simulation of incompressible flows on 3D hybrid unstructured grids, *Proceedings of the International Mechanical Engineering Congress and Exposition (IMECE 2014)*, Montreal, Canada, 2014, IMECE2014-39759.

[31] Y. Kallinderis, H.T. Ahn. Incompressible Navier-Stokes method with general hybrid grids, *Journal of Computational Physics*, **210**, 75-108, 2005.

[32] A.J. Chorin. A numerical method for solving incompressible viscous flow problems," *Journal of Computational Physics*, **2**, 12-26, 1967.

[33] W.W. Kim, S. Menon, An unsteady incompressible Navier-Stokes solver for large eddy simulation of turbulent flows. *International Journal for Numerical Methods in Fluids*, **31**, 983-1017, 1999.

[34] P. Roe. Approximate Riemann solvers, parameter vectors and difference schemes, *Journal of Computational Physics*, **43**, 357-371, 1981.

[35] J. Blazek, Computational Fluid Dynamics: Principles and Applications, *Elsevier Science*, Kidlington, 2001.

[36] F.R. Menter, Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications, *AIAA Journal,* **32**, 1598-1605, 1994.

[37] M.H. Lallemand, Etude de Schemas Runge-Kutta a 4 pas pour la Resolution Multigrille des Equations d' Euler 2D, *Raport de Recherche*, INRIA, 1988.

[38] R. L. Wardlaw, G. F. Moss. A standard tall building model for the comparison of simulated natural winds in wind tunnels*, Report CC-662 Tech25*, 1970.

[39] S. Huang, Q. S. Li, S. Xu. Numerical evaluation of wind effects on a tall steel building by CFD, *Journal of Constructional Steel Research,* **63.5***,* 612-627, 2007.

[40] A. K. Dagnew, G. T. Bitsuamalk, R. Merrick. Computational evaluation of wind pressures on tall buildings. *11th American conference on Wind Engineering*, San Juan, Puerto Rico, 2009.

[41] A. K. Dagnew, G. T. Bitsuamlak. LES evaluation of wind pressures on a standard tall building with and without a neighboring building, *The 5th International Symposium on Computational Wind Engineering*, Chapel Hill, North Carolina, USA, 2010.

[42] E. D. Obasaju. Measurement of forces and base overturning moments on the CAARC tall building model in a simulated atmospheric boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics,* **40.2**, 103-126, 1992.

[43] S. Thepmongkorn, K. C. S. Kwok, Wind-induced responses of tall buildings experiencing complex motion. *J Wind Eng Ind Aerodynamics*, **90**, 515–526, 2002.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MONOLITHIC FLUID-STRUCTURE INTERACTION METHOD, APPLICATION TO A PISTON PROBLEM

Felix Ischinger[1], **Martijn Anthonissen**[2], **and Barry Koren**[2]

[1]Eindhoven University of Technology
e-mail: f.ischinger@student.tue.nl

[2] Eindhoven University of Technology
Department of Mathematics and Computer Science
P.O. Box 513, 5600 MB Eindhoven
The Netherlands
e-mail: {b.koren, m.j.h.anthonissen}@tue.nl

Keywords: fluid-structure interaction, monolithic approach, Euler equations of gas dynamics, acoustic equations, mass-spring system, Discontinuous Galerkin, immersed boundary method, Lyapunov stability

Abstract. *A monolithic FSI method is presented. A standard piston problem is considered as test case. The piston problem's fluid domain is represented by a closed tube filled with air. One end of the fluid tube is formed by a piston connected to a spring. We use the Euler equations of gas dynamics as well as a linear simplification of these, the acoustic equations, to model the gas dynamics in the tube. A Discontinuous Galerkin method is applied to discretize the fluid-flow equations, together with an immersed-boundary method to account for the moving piston. A monolithic formulation of the coupled system is derived and analyzed. It is proven that the semidiscrete formulation is stable, if two correction terms are used at the coupling interface. We use Lyapunov functions to prove stability of the semi-discrete monolithic formulation. Further, different time-integration methods are considered, analyzed and tested. The numerical results are very accurate; they correspond very well to analytical approximations. The theoretical prediction of the eigenfrequency can be reproduced very accurately. Moreover, the amplitude of the spring oscillation is conserved very well.*

1 INTRODUCTION

The computer simulation of Fluid-Structure Interaction (FSI) has become a major research topic over the last two decades and is of interest to many applications. There are two main FSI approaches, the partitioned and the monolithic approach. The idea of the partitioned approach is to calculate the fluid flow and the structural movement separately. This has the advantage that already existing and specialized simulation tools can be used to calculate the dynamics of each system. The solvers are then applied alternatingly in time. The boundary conditions for one system are given by the state of the other system. Therefore, there is a time lag between the boundary conditions and the system. To avoid the time lag, monolithic methods are to be preferred over partitioned methods. In monolithic methods, the fluid and the structural part are evolved simultaneously in time. The challenge for monolithic methods lies in the correct mathematical formulation of the system. Both systems and the coupling between them have to be formulated in a single expression.

Another challenge in FSI computations is the numerical tracking of moving boundaries. In the Arbitrary-Lagrangian-Eulerian (ALE) approach, the computational grid is adapted at every time step to fit the geometry of the modeled problem. This may become computationally expensive. In contrast, there are Immersed Boundary (IB) methods, introduced by Peskin [1]. Over the last decades a lot of modifications were proposed for IB methods. The main idea is to use a non-deforming grid, which does not necessarily fit with the model geometry. IB methods are very well suited for complex geometries and large deformations. Furthermore, using an IB method can save a lot of computational time, because the computational grid does not have to be adapted at every time step. In this work, we use an IB method.

As a fluid-structure interaction problem we consider the piston problem from [2]. The model consists of a closed tube filled with air. The right end of the tube is closed by a piston that is attached to a spring and hence is movable (Figure 1). The pressure outside the tube is set constant to $p_0 > 0$. The left end of the tube is at $x_{\text{LB}} \in \mathbb{R}$ and the piston wall is at position $x_{\text{wall}} \in \mathbb{R}$. The equilibrium position of the spring is denoted as $l_0 \in \mathbb{R}$. The velocity of the piston is $\dot{x}_{\text{wall}} = u_{\text{wall}} \in \mathbb{R}$.

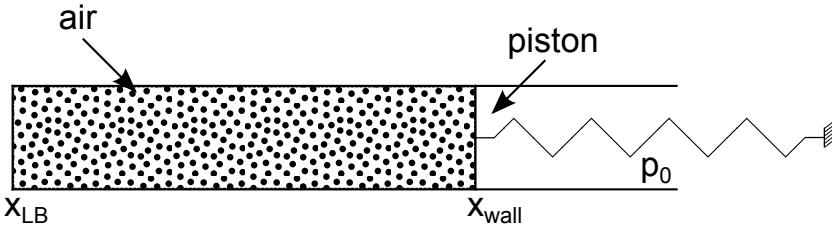


Figure 1: Sketch of the piston problem.

Some publications addressed the piston problem with an ALE approach. In [2] and [3] partitioned approaches are introduced. A first order accurate monolithic implicit scheme is proposed in [4]. Further work is presented in [5], [6] and [7]. In these publications, a time-discontinuous Galerkin method is used for the discretization of the fluid equations. This is related to an ALE approach. As mentioned, we will address the piston problem with an IB method.

The structural part of the FSI problem, the spring and the piston, is described by Newton's law. Two forces act on the piston, a restoring force F_r which is determined by Hooke's law and

a force F_p resulting from the pressure difference over the piston, i.e.

$$m\ddot{x}_{\text{wall}}(t) = F_p - F_r = A\left(p_{\text{wall}} - p_0\right) - k\left(x_{\text{wall}} - l_0\right), \tag{1}$$

with m the mass of the piston, A the cross sectional area of the piston, p_{wall} the pressure in the fluid at the left of the moving wall and k the stiffness of the spring. (1) can be rewritten into a system of first order ordinary differential equations:

$$\frac{d}{dt}\begin{pmatrix} x_{\text{wall}} \\ u_{\text{wall}} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{pmatrix}\begin{pmatrix} x_{\text{wall}} \\ u_{\text{wall}} \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{k}{m}l_0 + \frac{A}{m}\left(p_{\text{wall}} - p_0\right) \end{pmatrix}. \tag{2}$$

The air flow in the tube is modeled with a strictly hyperbolic system of conservation laws. We use the Euler equations and a linear simplification, the acoustic equations. The Euler equations describe the conservation of mass, momentum and energy of the gas. In one spatial dimension they read

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} + \frac{\partial}{\partial x}\begin{pmatrix} \rho u \\ \rho u^2 + p \\ u\left(E + p\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3}$$

ρ denotes the density, u denotes the velocity in x-direction and E is the total energy per unit volume. p is the pressure. An equation of state is necessary to close the system. For an ideal gas this is

$$p = \left(E - \frac{1}{2}u^2\rho\right)\left(\gamma - 1\right), \tag{4}$$

with the ratio of specific heats $\gamma = 1.4$. The Euler equations can also be written in quasi-linear form:

$$\frac{\partial}{\partial t}\boldsymbol{u} + \boldsymbol{F}\left(\boldsymbol{u}\right)\frac{\partial}{\partial x}\boldsymbol{u} = \boldsymbol{0}, \tag{5}$$

with the Jacobian of the flux $\boldsymbol{F}$

$$\boldsymbol{F}\left(\boldsymbol{u}\right) = \frac{d\boldsymbol{f}\left(\boldsymbol{u}\right)}{d\boldsymbol{u}} = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2}\left(\gamma - 1\right)u^2 & \left(3 - \gamma\right)u & \gamma - 1 \\ -\frac{\gamma\rho u E}{\rho^2} + \left(\gamma - 1\right)\left(u\right)^3 & \frac{\gamma E}{\rho} - \frac{3}{2}\left(\gamma - 1\right)u^2 & \gamma u \end{pmatrix}. \tag{6}$$

The acoustic equations are a linear simplification of the Euler equations. In fluids, acoustics is defined as small, isentropic perturbations. Therefore, perturbations $\tilde{\rho}$ around a constant density $\bar{\rho}$ (we choose $\bar{\rho} = 1.3^{\text{kg}}/_{\text{m}^3}$) and perturbations $\tilde{u}$ around a constant velocity $\bar{u}$ are considered. The gas is assumed to be at rest ($\bar{u} = 0^{\text{m}}/_{\text{s}}$). This gives the expressions $\rho = \tilde{\rho} + \bar{\rho}$ and $u = \tilde{u}$.

The isentropic speed of sound is given by

$$c = \sqrt{\left(\frac{\partial p}{\partial \rho}\right)\Big|_s}, \tag{7}$$

the square root of the derivative of the pressure with respect to the density at constant entropy s.

Substituting the expressions for density and velocity, as well as the isentropic speed of sound into the Euler equations and neglecting higher order terms gives

$$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \bar{\rho}u \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ c^2 & 0 \end{pmatrix}\frac{\partial}{\partial x}\begin{pmatrix} \rho \\ \bar{\rho}u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{8}$$

The quasi-linear formulation can also be transformed into characteristic variables $\boldsymbol{\xi}$ with the help of diagonalization matrices $\boldsymbol{R}_f$:

$$\boldsymbol{R}_f^{-1}\frac{\partial}{\partial t}\boldsymbol{u} + \boldsymbol{R}_f^{-1}\boldsymbol{F}\boldsymbol{R}_f\boldsymbol{R}_f^{-1}\frac{\partial}{\partial x}\boldsymbol{u} = \boldsymbol{0} \tag{9}$$

$$\Leftrightarrow \qquad \frac{\partial}{\partial t}\boldsymbol{\xi} + \boldsymbol{\Lambda}\frac{\partial}{\partial x}\boldsymbol{\xi} = \boldsymbol{0}. \tag{10}$$

The diagonalization matrix for the Euler equations is denoted as $\boldsymbol{R}_{f,\text{Euler}}$ and for the acoustic equations as $\boldsymbol{R}_{f,\text{ac}}$. The matrix $\boldsymbol{\Lambda}$ contains the eigenvalues λ_j of the Jacobian matrix $\boldsymbol{F}$. The characteristic variables and the diagonalization matrices will appear later on for the numerical flux calculation and the stability analysis.

In [4] an analytical approach is introduced to predict the eigenfrequency of the piston problem when the acoustic equations are used. In contrast to the referred work, we do not neglect the cross sectional area A of the piston and get the equation

$$-m\omega^2 + k + A\bar{\rho}\omega c\frac{1}{\tan\left(\frac{\omega}{c}L\right)} = 0. \tag{11}$$

ω describes the eigenfrequency of the coupled piston system. L is the mean length of the tube and defined as $L = l_0 - x_{\text{LB}}$. (11) can now be made dimensionless. We introduce the three dimensionless numbers

$$X = \frac{\bar{\rho}AL}{m}, \qquad Y = \sqrt{\frac{k}{m}}\frac{L}{c}, \qquad Z = \frac{\omega L}{c}. \tag{12}$$

(11) then reads

$$-Z^2 + Y^2 + XZ\frac{1}{\tan(Z)} = 0. \tag{13}$$

X describes the mass ratios between the mass of the air in the tube ($\bar{\rho}AL$) and the mass of the piston (m). If the mass of the piston dominates the mass of the air inside the fluid tube, X becomes small and can eventually be assumed to be zero. (13) can then be rewritten into an explicit formulation for ω:

$$\omega^2 = \frac{k}{m}. \tag{14}$$

This is just the eigenfrequency of a simple mass-spring system. We conclude that if the piston mass is dominating the mass of the air, then the eigenfrequency of the system is only determined by the spring and the fluid tube does not have any influence. Therefore, it could be reasonable for small values of X to make the FSI coupling only in one direction, from the spring onto the fluid. In this case it is assumed that the fluid does not have any influence on the spring.

We choose the parameters for our test problem as $m = 0.8\text{kg}$, $k = 0.8 \cdot 100^{\text{kg}}/_{\text{s}^2}$, $c = 328.17^{\text{m}}/_{\text{s}}$, $L = 1\text{m}$ and $A = 1\text{m}^2$. This results in an angular frequency of $\omega \approx 341.6068^{1}/_{\text{s}}$, respectively an eigenfrequency of $f = \frac{\omega}{2\pi} \approx 54.3684^{1}/_{\text{s}}$. Further, we get $X = 1.625$, thus a strong interplay between the fluid and the structure. Therefore a monolithic coupling is advisable.

2 NUMERICAL SCHEME

2.1 Notation

First we need to introduce some notation. For the piston problem, it is necessary to define some special notation. Two domains are defined. The first one is called the enclosing domain

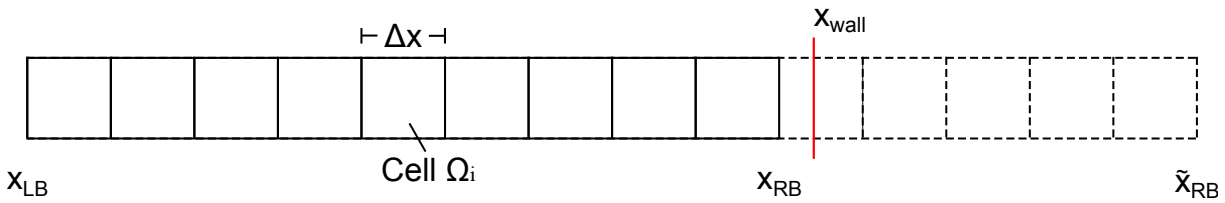


Figure 2: Sketch of the two domains. The computational domain Ω is drawn with solid lines. The enclosing domain Ω_{tot} is defined by all cells, indicated by the solid and dashed lines together.

Ω_{total} and represents the largest possible domain for the fluid. The second one is called Ω and it is a subset of the enclosing domain $\Omega \subseteq \Omega_{\text{total}}$. Ω is the computational domain. The left boundary of Ω_{total} and Ω is always the same as this end is immovable. It has the coordinate x_{LB}. The right boundary of the enclosing domain Ω_{total} is also fixed. It has the coordinate $\tilde{x}_{\text{RB}}$. It is required that the right boundary of Ω is always within the enclosing domain Ω_{total}, that is $x_{\text{wall}} \leq \tilde{x}_{\text{RB}}$. The right boundary of the computational domain is called x_{RB}.

For the numerical discretization the enclosing domain Ω_{total} is split into $N_{\text{total}} \in \mathbb{N}$ cells Ω_i with equal lengths Δx. Therefore, the spatial discretization size becomes $\Delta x := \frac{\tilde{x}_{\text{RB}} - x_{\text{LB}}}{N_{\text{total}}}$. The boundaries of a cell i have the coordinates $x_{i,\text{L}}$ for the left cell boundary and $x_{i,\text{R}}$ for the right one. The number of cells in the computational domain is $N \leq N_{\text{total}}$.

The i-th cell is called Ω_i. Every cell contains $p+1$ integration points $x_{i,k}$ where $i = 1, \ldots, N$ is the index of the cell number and $k = 1, \ldots, p+1$ indicates the integration point on which the solution of the Discontinuous Galerkin method is represented. $p \in \mathbb{N}_0$ is the polynomial degree of the solution. A Discontinuous Galerkin method with $p = 0$ may be interpreted as a Finite-Volume method.

2.2 Discontinuous Galerkin method

The general expression for the Euler and the acoustic equations is

$$\frac{\partial}{\partial t}\boldsymbol{u}(x,t) + \frac{\partial}{\partial x}\boldsymbol{f}\left(\boldsymbol{u}(x,t)\right) = \boldsymbol{0}. \tag{15}$$

The numerical solution of the DG method to approximate $\boldsymbol{u}(x,t)$ consists of N polynomials of order p for each conserved variable, referred to as $\boldsymbol{U}_i\left(x,t\right) \in \left(\mathbb{P}^p\left(\Omega_i\right)\right)^d$ with i indicating the respective cell Ω_i and $\mathbb{P}^p\left(\Omega_i\right)$ the space of polynomials of order p on Ω_i. For the Euler equations, a system of three equations, $\boldsymbol{U}_i(x,t)$ is a vector of three polynomials $U_{i,j}(x,t)$, representing the conserved variables ρ, ρu and E. At each cell interface the solution is not unique but defined twice for each variable, because the cells are closed intervals. A detailed derivation and analysis of the Discontinuous Galerkin method can be found in [8].

We are using the nodal Discontinuous Galerkin approach. This means that the solution for one variable j on one cell i has the structure

$$U_{i,j}(x,t) = \sum_{k=1}^{p+1} U_{i,j}\left(x_k,t\right) l_k(x) = \sum_{k=1}^{p+1} U_{i,j,k} l_k(x). \tag{16}$$

The Legendre-Gauß-Lobatto points are used as integration points $x_{i,k}$. They have the advantage that they always lie on the boundaries of the cells which will be helpful for the flux calculations. $l_k(x)$ are the Lagrange polynomials defined on cell Ω_i. They are defined with the help of the

integration points:

$$l_k(x) = \prod_{\substack{m=1 \\ m \neq k}}^{p+1} \frac{x - x_{i,m}}{x_{i,k} - x_{i,m}}, \qquad \text{for } x \in \Omega_i. \tag{17}$$

$U_{i,j,k}$ are the integration coefficients and can be written more compactly in the vector $\boldsymbol{U} \in \mathbb{R}^{Nd(p+1)}$. This vector contains values for every integration point, every unknown variable and every cell. In total, the solution vector of the nodal DG method looks as follows:

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_1 \\ \vdots \\ \boldsymbol{U}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}_{1,1} \\ \vdots \\ \boldsymbol{U}_{1,d} \\ \vdots \\ \vdots \\ \boldsymbol{U}_{N,d} \end{pmatrix} = \begin{pmatrix} U_{1,1,1} \\ \vdots \\ U_{1,1,p+1} \\ \vdots \\ \vdots \\ U_{N,d,p+1} \end{pmatrix}. \tag{18}$$

For the Discontinuous Galerkin discretization, we replace the exact solution in (15) by the nodal approach, multiply with test functions from $\mathbb{P}^p(\Omega_i)$ and integrate over cell Ω_i. The flux term is simply approximated by

$$\boldsymbol{f}(\boldsymbol{U}_i)_j \approx \sum_{k=1}^{p+1} \boldsymbol{f}\left(U_{i,j,k}(t)\right)_j l_k(x) =: \sum_{k=1}^{p+1} F_{i,j,k}(t)\, l_k(x). \tag{19}$$

Using integration by parts, we get

$$\begin{aligned} &\int_{\Omega_i} \frac{\partial}{\partial t} \left(\sum_{k=1}^{p+1} U_{i,j,k}(t) l_k(x) \right) l_m(x) d\Omega_i - \int_{\Omega_i} \left(\sum_{k=1}^{p+1} F_{i,j,k}(t) l_k(x) \right) \frac{\partial}{\partial x} l_m(x) d\Omega_i \\ &\qquad = \delta_{1,m} \left[g_j(\boldsymbol{U}_{i-1}, \boldsymbol{U}_i) \right] + \delta_{p+1,m} \left[-g_j(\boldsymbol{U}_{i-1}, \boldsymbol{U}_i) \right], \qquad m = 1, \ldots, p+1. \end{aligned} \tag{20}$$

The flux at the right-hand side, that is evaluated at the boundary of the cell, is approximated with a numerical flux g_j, known from the Finite-Volume methods. As the Legendre-Gauss-Lobatto points are used, the cell boundaries are interpolation points. Therefore, as $l_i(x_j) = \delta_{ij}$, the value at the cell interface is exclusively defined by the outer interpolation points. δ_{ij} denotes the Kronecker delta. We use a flux-splitting approach to calculate the numerical fluxes:

$$g_j(\boldsymbol{U}_{i-1}, \boldsymbol{U}_i) = \left(\boldsymbol{F}^+ \boldsymbol{U}_{i-1} + \boldsymbol{F}^- \boldsymbol{U}_i \right)_j. \tag{21}$$

The positive and negative fluxes are defined as

$$\boldsymbol{F}^\pm := \boldsymbol{R}_f \boldsymbol{\Lambda}^\pm \boldsymbol{R}_f^{-1} = \boldsymbol{R}_f \begin{pmatrix} \lambda_1^\pm & & \\ & \ddots & \end{pmatrix} \boldsymbol{R}_f^{-1} = \boldsymbol{R}_f \begin{pmatrix} \frac{1}{2}(\lambda_1 \pm |\lambda_1|) & & \\ & \ddots & \end{pmatrix} \boldsymbol{R}_f^{-1}. \tag{22}$$

(20) is called the weak formulation of the DG method. After one more integration by parts, using the assumption that the values only depend on values from inside the cell, we get to the strong formulation of the DG scheme:

$$\begin{aligned} &\int_{\Omega_i} \frac{\partial}{\partial t} \left(\sum_{k=1}^{p+1} U_{i,j,k} l_k(x) \right) l_m(x) + \left(\frac{\partial}{\partial x} \sum_{k=1}^{p+1} F_{i,j,k} l_k(x) \right) l_m(x) d\Omega_i \\ &\qquad = \delta_{1,m} \left(-F_{i,j,1} + g_j(\boldsymbol{U}_{i-1}, \boldsymbol{U}_i) \right) + \delta_{p+1,m} \left[F_{i,j,p+1} - g_j(\boldsymbol{U}_i, \boldsymbol{U}_{i+1}) \right]. \end{aligned} \tag{23}$$

In the remainder of this work we consider the strong formulation only. Next, the integrals will be projected on the reference cell $I = [-1, 1]$ and we call the reference variable $r \in I$. The interpolation points on I are called r_k. The following mapping from the reference variable r to the actual variable x is used:

$$x(r) = x_{i,\mathrm{L}} + \frac{1+r}{2}\Delta x, \qquad -1 \leq r \leq 1, \qquad \text{for } x \in \Omega_i. \tag{24}$$

The transformation between r and x for some function f under the integral is

$$\int_{x_{i,\mathrm{L}}}^{x_{i,\mathrm{R}}} f(x)\,dx = \int_{-1}^{1} f(x(r))\frac{dx(r)}{dr}dr = \frac{\Delta x}{2}\int_{-1}^{1} f(r)dr. \tag{25}$$

The transformation introduces a scaling factor $\frac{dx(r)}{dr}$ with $x(r)$ from (24). For the second integral, this scaling factor is compensated by the change of the derivative $\frac{d}{dx}$ to $\frac{d}{dr}$. The transformation becomes especially useful if the cells do not have the same size.

The integration of the multiplied Lagrange polynomials can be summarized in a mass matrix $\boldsymbol{\mathcal{M}}$ with entries

$$\boldsymbol{\mathcal{M}}_{pq} = \int_{-1}^{1} l_p(x(r))l_q(x(r))dr = \int_{-1}^{1} \prod_{\substack{m=1\\m\neq p}}^{p+1} \frac{r-r_m}{r_p-r_m} \prod_{\substack{m=1\\m\neq p}}^{p+1} \frac{r-r_m}{r_q-r_m}dr. \tag{26}$$

The same holds for the stiffness matrix $\boldsymbol{\mathcal{S}}$:

$$\boldsymbol{\mathcal{S}}_{pq} = \int_{-1}^{1} l_p(x(r))\frac{d}{dr}l_q(x(r))dr. \tag{27}$$

The strong formulation of the DG scheme can then be written very compactly for every cell:

$$\frac{d}{dt}\boldsymbol{U}_{i,j} = \frac{2}{\Delta x}\boldsymbol{\mathcal{M}}^{-1}\left(-\boldsymbol{\mathcal{S}}\boldsymbol{F}_{i,j} + \boldsymbol{F}_{i,j}|_{\partial\Omega_i} + \boldsymbol{g}_j\right). \tag{28}$$

The inner boundary fluxes from the second integration by parts is represented in the vector $\boldsymbol{F}_{i,j}|_{\partial\Omega_i}$ and the numerical fluxes over the cell boundaries are summarized in the vector $\boldsymbol{g}_j$. The first entry contains the flux of the j-th variable between cell $i-1$ and i and the last entry the numerical flux between i and $i+1$. The other entries are zero. Note that the mass matrix is in $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{(p+1)^2}$ and thus can be inverted with reasonable computational cost.

2.3 Limiter

Numerical schemes with order higher than one in the spatial discretization produce unphysical oscillations at steep gradients and discontinuities in the solution. These must be eliminated or at least controlled, because they might cause breakdowns in the computation (e.g. negative densities). The solution to this problem is to introduce limiters to make the numerical solution total variation bounded (TVB). This is done for every variable separately and therefore the subscript j is dropped in this section.

We will make use of the modified minmod function. In its general form this is

$$\bar{m}(a_1, \ldots, a_m) := \text{minmod}\left(a_1, a_2 + M\Delta x\text{sign}(a_2), \ldots, a_m + M\Delta x^2\text{sign}(a_m)\right). \tag{29}$$

The number $M \geq 0$ depends on the problem and is not known a-priori. If M is chosen too small, the local dissipation and order reduction are increased. If, on the other hand, M is chosen too large, oscillations occur [8].

The first step in the limiting procedure is to look for cells in which oscillations might occur and mark them as so-called troubled cells. The high order accuracy of the DG method should not be limited in smooth regions of the solution. This is done with the generalized limiter

$$U_i^{\text{lim},L} = \bar{U}_i - \text{minmod}(\bar{U}_i - U_{i,\cdot,1}, \bar{U}_{i+1} - \bar{U}_i, \bar{U}_i - \bar{U}_{i-1}), \tag{30}$$

$$U_i^{\text{lim},R} = \bar{U}_i + \text{minmod}(U_{i,\cdot,p+1} - \bar{U}_i, \bar{U}_i - \bar{U}_{i+1}, \bar{U}_{i-1} - \bar{U}_i). \tag{31}$$

Then, if $U_i^{\text{lim},L} = U_{i,\cdot,1}$ and $U_i^{\text{lim},R} = U_{i,\cdot,p+1}$ the cell is not marked as a troubled cell. In this case, all three input arguments of the minmod function have the same sign and the jump from the average solution in cell i to its boundary value has the smallest absolute value.

The second step is the actual limiting of the troubled cells. The problem that the limiter destroys high order accuracy in regions of smooth extrema remains present, since local extrema are detected as oscillations. In order to improve this, we limit the solution with a so-called modified minmod limiter. We use the Monotone Upstream-centered Scheme for Conservation Laws (MUSCL) limiter [8]

$$\Pi U_i(x) = \bar{U}_i + (x - x_{i,c})\,\text{minmod}\left((U_i)_x, \frac{\bar{U}_{i+1} - \bar{U}_i}{\Delta x}, \frac{\bar{U}_i - \bar{U}_{i-1}}{\Delta x}\right). \tag{32}$$

Using the limited solution $\Pi U_i(x)$ at every numerical time integration step guarantees a total variation bounded solution.

2.4 Fluid boundaries

Boundary conditions for the Discontinuous Galerkin scheme are modeled with so-called ghost cells. These are artificial cells at the ends of the computational domain Ω. The left end of the tube (x_{LB}) does not move. Static reflective boundary conditions are used to model this. For the acoustic equations, this is modeled with the ghost cell state values

$$\boldsymbol{U}_0^{\text{ac}} = \begin{pmatrix} \rho_1 \\ -\bar{\rho} u_1 \end{pmatrix}. \tag{33}$$

The left boundary treatment for the Euler equations is correctly modeled with the ghost cell values

$$\rho_0 = \rho_1, \quad u_0 = -u_1, \quad p_0 = p_1, \tag{34}$$

respectively for the conservative variables

$$\boldsymbol{U}_0^{\text{Euler}} = \begin{pmatrix} \rho_1 \\ -\rho_1 u_1 \\ E_1 \end{pmatrix}. \tag{35}$$

It can be shown that solving a Riemann problem between the left inner cell of the domain and the ghost cell gives exactly the desired zero velocity [9].

The right end of the tube is defined by the position of the piston. Therefore, a boundary condition for moving reflective boundaries must be introduced. We interpret this as the limit of

a subsonic outflow boundary condition. This reads [9]

$$\boldsymbol{U}_{N+1}^{\text{ac}} = \begin{pmatrix} \rho_N \\ -\bar{\rho} u_N + 2\rho_N u_{\text{wall}} \end{pmatrix}, \qquad \boldsymbol{U}_{N+1}^{\text{Euler}} = \begin{pmatrix} \rho_N \\ -(\rho u)_N + 2\rho_N u_{\text{wall}} \\ E_N \end{pmatrix}. \tag{36}$$

The computational mesh is not changed, so at some point, the moving wall will reach a new cell. If a new cell has to be added, it gets the initial values

$$\boldsymbol{U}_{\text{new}}^{\text{ac}} = \begin{pmatrix} \rho_N \\ \bar{\rho} u_{\text{wall}} \end{pmatrix}, \qquad \boldsymbol{U}_{\text{new}}^{\text{Euler}} = \begin{pmatrix} \rho_\star \\ \rho_\star u_\star \\ \frac{1}{2}\rho_\star u_\star^2 + \frac{p_\star}{(\gamma-1)\rho_\star} \end{pmatrix} \tag{37}$$

for the acoustic equations and the Euler equations, respectively. The star values are the exact solution values of the boundary Riemann problem defined in [9].

2.5 Monolithic formulation

In this section, we introduce the monolithic formulation for the piston problem using the acoustic equations. In its general semi-discrete formulation it has the structure

$$\frac{d}{dt}\boldsymbol{X} = \boldsymbol{M}\boldsymbol{X} + \boldsymbol{F}_{\text{ext}}, \tag{38}$$

with the vector of unknowns

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}. \tag{39}$$

$\boldsymbol{U}$ represents the fluid variables and $\boldsymbol{V}$ the spring variables. $\boldsymbol{M}$ is the system matrix. The system matrix is a block matrix describing the fluid and the spring discretization as well as the coupling between the two systems.

First of all, we need to introduce some general matrices. We write the identity matrix as $\boldsymbol{I}^N$. The superscript indicates the dimension as $N \times N$. The matrices $\boldsymbol{I}_{-1}^N$ and $\boldsymbol{I}_{+1}^N$ have ones on the sub- and superdiagonal respectively. $\boldsymbol{I}^{p+1}|_{\partial\Omega_i}$ has the entries -1 in the top left corner and 1 in the bottom right corner. Finally, we use the Kronecker product $\otimes$ to define the matrices

$$\boldsymbol{I}_{11} = \begin{pmatrix} 0 & \cdots & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \boldsymbol{I}_{12} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$\boldsymbol{I}_{21} = \begin{pmatrix} 0 & \cdots & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{I}_{22} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

The DG discretization for the single cells is defined with the following matrices:

$$\boldsymbol{A}_c = \frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[-\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{S}}\right)\left(\boldsymbol{F} \otimes \boldsymbol{I}^{p+1}\right) + \left(\boldsymbol{F} \otimes \boldsymbol{I}^{p+1}|_{\partial\Omega_i}\right)\right], \tag{40}$$

$$\boldsymbol{G}_{l+} = \frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\boldsymbol{F}^+ \otimes \boldsymbol{I}_{11}\right], \tag{41}$$

$$\boldsymbol{G}_{l-} = \frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\boldsymbol{F}^- \otimes \boldsymbol{I}_{12}\right], \tag{42}$$

$$\boldsymbol{G}_{r+} = \frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\boldsymbol{F}^+ \otimes \boldsymbol{I}_{21}\right], \tag{43}$$

$$\boldsymbol{G}_{r-} = \frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\boldsymbol{F}^- \otimes \boldsymbol{I}_{22}\right]. \tag{44}$$

Together, they define the matrix for the fluid flow discretization. However, the boundary treatment is not included yet:

$$\boldsymbol{A} := \left(\boldsymbol{I}^N \otimes \boldsymbol{A}_c\right) + \left(\boldsymbol{I}^N_{-1} \otimes \boldsymbol{G}_{l+}\right) + \left(\boldsymbol{I}^N \otimes \boldsymbol{G}_{l-}\right) - \left(\boldsymbol{I}^N \otimes \boldsymbol{G}_{r+}\right) - \left(\boldsymbol{I}^N_{+1} \otimes \boldsymbol{G}_{r-}\right) \tag{45}$$

$$= \begin{pmatrix} \boldsymbol{A}_c + \boldsymbol{G}_{l-} - \boldsymbol{G}_{r+} & -\boldsymbol{G}_{r-} & & & \\ \boldsymbol{G}_{l+} & \boldsymbol{A}_c + \boldsymbol{G}_{l-} - \boldsymbol{G}_{r+} & -\boldsymbol{G}_{r-} & & \\ & \boldsymbol{G}_{l+} & \boldsymbol{A}_c + \boldsymbol{G}_{l-} - \boldsymbol{G}_{r+} & -\boldsymbol{G}_{r-} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{46}$$

The static boundary at the left end, as described in (33), becomes the matrix

$$\boldsymbol{G}_{\mathrm{LB}} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \otimes \left[\frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\begin{pmatrix} \frac{c}{2} & -\frac{1}{2} \\ \frac{c^2}{2} & -\frac{c}{2} \end{pmatrix} \otimes \boldsymbol{I}_{12}\right]\right]. \tag{47}$$

The numerical modeling for the right, movable wall was described in (36). It is split into two parts. The first one is

$$\boldsymbol{G}_{\mathrm{RB}} := \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \otimes \left[\frac{2}{\Delta x}\left(\boldsymbol{I}^d \otimes \boldsymbol{\mathcal{M}}^{-1}\right)\left[\begin{pmatrix} -\frac{c}{2} & -\frac{1}{2} \\ \frac{c^2}{2} & \frac{c}{2} \end{pmatrix} \otimes \boldsymbol{I}_{21}\right]\right], \tag{48}$$

and models the reflections that are independent of the spring. The second part models the coupling from the spring to the fluid. It is

$$\boldsymbol{B} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \otimes \left[\left(\boldsymbol{I}^d \otimes \frac{2}{\Delta x}\boldsymbol{\mathcal{M}}^{-1}\right)\begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & -\bar{\rho} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & c\bar{\rho} \end{pmatrix}\right]. \tag{49}$$

The matrix for the spring is defined as

$$\boldsymbol{D} := \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{pmatrix}. \tag{50}$$

The coupling from the fluid to the spring is done with the matrix $\boldsymbol{C} \in \mathbb{R}^{2\times Nd(p+1)}$. Remember that ρ actually is a disturbance in the mean density and is denoted as $\tilde{\rho}$. For the coupling, this must be taken into account and the contribution of the mean density has to be added. The total contribution must be $\frac{1}{m}c^2\rho_{N,p+1} + \frac{1}{m}c^2\bar{\rho}$. Thus, the coupling matrix is

$$\boldsymbol{C} := \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \frac{c^2}{m} & 0 & \cdots & 0 \end{pmatrix}, \tag{51}$$

where the non-zero entry is in row 2 and in column $Nd(p+1)-(p+1)$. The contribution of the mean density goes into the term $\boldsymbol{F}_{\text{ext}} \in \mathbb{R}^{Nd(p+1)+2}$. Besides this, the external pressure p_0 and the position of rest of the spring l_0 are taken into account:

$$\boldsymbol{F}_{\text{ext}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{k}{m}l_0 - \frac{1}{m}p_0 + \frac{c^2}{m}\bar{\rho} \end{pmatrix}. \tag{52}$$

For reasons that will become clear in the stability analysis, we introduce two correction terms $\boldsymbol{\mathcal{C}}_1$ and $\boldsymbol{\mathcal{C}}_2$. They account for the difference between the velocity of the piston u_{wall} and the last inner velocity of the gas in the tube $u_{N,p+1}$, which introduces instabilities in the system. They are defined as

$$\boldsymbol{\mathcal{C}}_1 = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \otimes \left[\boldsymbol{I}^d \otimes \frac{2}{\Delta x}\boldsymbol{\mathcal{M}}^{-1} \left[\begin{pmatrix} 0 & 0 \\ 0 & -c \end{pmatrix} \otimes \boldsymbol{I}_{21} \right] \right], \tag{53}$$

and

$$\boldsymbol{\mathcal{C}}_2 := \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \otimes \left[\boldsymbol{I}^d \otimes \frac{2}{\Delta x}\boldsymbol{\mathcal{M}}^{-1} \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & -c\bar{\rho} \end{pmatrix} \right]. \tag{54}$$

The stability analysis and a more detailed motivation for this choice of correction can be found in the next section.

We conclude with the monolithic formulation, using the above derived matrices to assemble the system matrix $\boldsymbol{M}$:

$$\frac{d}{dt}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A} + \boldsymbol{G}_{\text{LB}} - \boldsymbol{G}_{\text{RB}} + \boldsymbol{\mathcal{C}}_1 & \boldsymbol{B} + \boldsymbol{\mathcal{C}}_2 \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} + \boldsymbol{F}_{\text{ext}}. \tag{55}$$

3 STABILITY ANALYSIS

We now want to address the subject of numerical stability, thus the response of the system to small perturbations. This is of interest, because during numerical simulations, there will always be a truncation error due to the limited computer precision. Roughly speaking, a system is stable, if small perturbations only have a small impact on the results. For the analysis we use Lyapunov's concept, which can be found in e.g. [10] in more detail.

3.1 Lyapunov's stability concept

Let us first consider the ordinary differential equation

$$\dot{\boldsymbol{x}} = \boldsymbol{f}\left(\boldsymbol{x}\left(t\right)\right), \qquad \boldsymbol{x} \in \mathbb{R}^n. \tag{56}$$

$\boldsymbol{x}_e \in \mathbb{R}^n$ is an equilibrium point of (56), if

$$\boldsymbol{f}\left(\boldsymbol{x}_e\right) = \boldsymbol{0}, \quad \forall t \geq 0. \tag{57}$$

The equilibrium point $\boldsymbol{x}_e$ is said to be (Lyapunov) stable, if

$$\forall \epsilon > 0, \exists \delta > 0 : ||\boldsymbol{x}\left(0\right) - \boldsymbol{x}_e|| < \delta \Rightarrow ||\boldsymbol{x}(t) - \boldsymbol{x}_e|| < \epsilon, \quad \forall t > 0. \tag{58}$$

Let $\boldsymbol{x}_e$ be an equilibrium point of (56). Further, if there exists a C^1-function $\mathcal{E} : \mathcal{U} \rightarrow \mathbb{R}$ defined on a neighborhood $\mathcal{U}$ of $\boldsymbol{x}_e$ with the properties

i) $\mathcal{E}\left(\boldsymbol{x}_e\right) = 0$ and $\mathcal{E}(\boldsymbol{x}) > 0, \forall \boldsymbol{x} \in \mathcal{U} \setminus \{\boldsymbol{x}_e\}$,

ii) $\dot{\mathcal{E}}\left(\boldsymbol{x}(t)\right) \leq 0, \forall \boldsymbol{x} \in \mathcal{U} \setminus \{\boldsymbol{x}_e\}$,

then $\mathcal{E}\left(\boldsymbol{x}\right)$ is called a Lyapunov function and the equilibrium point $\boldsymbol{x}_e$ is stable.

3.2 Stability of the semi-discrete piston problem

Lyapunov's stability concept can be simplified for the piston problem with acoustic equations. We consider the evolution of an error $\boldsymbol{\delta}$ in the initial condition. Using the monolithic formulation (55), this becomes

$$\frac{d}{dt}\mathbf{err}(t) = \boldsymbol{M}\mathbf{err}(t), \qquad \mathbf{err}(0) = \boldsymbol{\delta}. \tag{59}$$

We choose the Lyapunov function candidate $\mathcal{E}\left(\boldsymbol{X}\right) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{E}\boldsymbol{X}$ and investigate the equilibrium point $\mathbf{err}(t) = \boldsymbol{X}_e = \boldsymbol{0}$. $\boldsymbol{E}$ is called the energy matrix of the system and $\boldsymbol{X} = \begin{pmatrix}\boldsymbol{U}, & \boldsymbol{V}\end{pmatrix}^T$ is the vector of unknowns. Inspired by [2] we choose the energy matrix to be

$$\boldsymbol{E} = \begin{pmatrix}\boldsymbol{E}_f & \\ & \boldsymbol{E}_s\end{pmatrix} = \begin{pmatrix}\boldsymbol{I}^N \otimes \left(\left(\overline{\boldsymbol{R}}_{f,\text{ac}}^{-T}\boldsymbol{R}_{f,\text{ac}}^{-1} + \boldsymbol{R}_{f,\text{ac}}^{-T}\overline{\boldsymbol{R}}_{f,\text{ac}}^{-1}\right) \otimes \frac{1}{\bar{\rho}}\frac{\Delta x}{2}\boldsymbol{\mathcal{M}}\right) & \\ & m\left(\overline{\boldsymbol{R}}_s^{-T}\boldsymbol{R}_s^{-1} + \boldsymbol{R}_s^{-T}\overline{\boldsymbol{R}_s}^{-1}\right)\end{pmatrix}. \tag{60}$$

Then, $\boldsymbol{E}$ is a real symmetric positive definite matrix and the criteria $\mathcal{E}\left(\boldsymbol{X}_e\right) = 0$ and $\mathcal{E}\left(X\right) > 0$ are fulfilled. It is left to show that $\dot{\mathcal{E}}\left(\boldsymbol{X}(t)\right) \leq 0$, which turns out to consist of three subproblems:

$$\begin{aligned}\boldsymbol{X}^T\boldsymbol{E}\boldsymbol{M}\boldsymbol{X} &= \begin{pmatrix}\boldsymbol{U}\\ \boldsymbol{V}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{E}_f & \\ & \boldsymbol{E}_s\end{pmatrix}\begin{pmatrix}\boldsymbol{A}+\boldsymbol{G}_{LB}-\boldsymbol{G}_{RB}+\boldsymbol{\mathcal{C}}_1 & \boldsymbol{B}+\boldsymbol{\mathcal{C}}_2\\ \boldsymbol{C} & \boldsymbol{D}\end{pmatrix}\begin{pmatrix}\boldsymbol{U}\\ \boldsymbol{V}\end{pmatrix}\\ &= \underbrace{\boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{A}\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{G}_{LB}\boldsymbol{U} - \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{G}_{RB}\boldsymbol{U}}_{1} + \\ &\quad \underbrace{\boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{\mathcal{C}}_1\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{B}\boldsymbol{V} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{\mathcal{C}}_2\boldsymbol{V} + \boldsymbol{V}^T\boldsymbol{E}_s\boldsymbol{C}\boldsymbol{U}}_{2} + \underbrace{\boldsymbol{V}^T\boldsymbol{E}_s\boldsymbol{D}\boldsymbol{V}}_{3}.\end{aligned} \tag{61}$$

The single terms have the following interpretation:

1. The DG discretization with two reflecting walls has to be stable. It can be proven that
$$\boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{A}\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{G}_{LB}\boldsymbol{U} - \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{G}_{RB}\boldsymbol{U} \leq 0. \tag{62}$$

2. The coupling between the fluid and the spring has to be stable:
$$\boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{\mathcal{C}}_1\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{B}\boldsymbol{V} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{\mathcal{C}}_2\boldsymbol{V} + \boldsymbol{V}^T\boldsymbol{E}_s\boldsymbol{C}\boldsymbol{U} - \bar{\rho}cu^2_{N,p+1} = 0. \tag{63}$$
The last term is a left-over from the inner DG discretization. The stability is only guaranteed with a correction.

3. The spring system has to be stable, i.e.
$$\boldsymbol{V}^T\boldsymbol{E}_s\boldsymbol{D}\boldsymbol{V} = 0. \tag{64}$$

In total, we get
$$\frac{d}{dt}\mathcal{E} = \boldsymbol{X}^T\boldsymbol{E}\boldsymbol{M}\boldsymbol{X} \leq 0. \tag{65}$$
Therefore, the coupled system is (Lyapunov) stable.

Without the use of the correction terms $\boldsymbol{\mathcal{C}}_1$ and $\boldsymbol{\mathcal{C}}_2$, the energy analysis results in
$$\begin{aligned}
&\boldsymbol{V}^T\boldsymbol{E}_s\boldsymbol{C}\boldsymbol{U} + \boldsymbol{U}^T\boldsymbol{E}_f\boldsymbol{B}\boldsymbol{V} - \bar{\rho}cu^2_{N,p+1} \\
&\quad = \rho_{N,p+1}c^2u_{\text{wall}} - \rho_{N,p+1}c^2u_{\text{wall}} + \bar{\rho}cu_{N,p+1}u_{\text{wall}} - \bar{\rho}cu^2_{N,p+1} \\
&\quad = \bar{\rho}cu_{N,p+1}\left(u_{\text{wall}} - u_{N,p+1}\right).
\end{aligned} \tag{66}$$
The velocity of the wall u_{wall} and the velocity in the last cell $u_{N,p+1}$ are very close to each other. The difference between them will become smaller, if the spatial discretization length Δx is reduced. Nevertheless, the difference might be positive and introduce instabilities to the system. By introducing the two correction terms $\boldsymbol{\mathcal{C}}_1$ and $\boldsymbol{\mathcal{C}}_2$ it is possible to eliminate this error.

4 RESULTS

We will compare different time integration methods for the piston problem. The Forward Euler method with time step Δt for the monolithic formulation simply reads
$$\boldsymbol{X}^{n+1} = \boldsymbol{X}^n + \Delta t\boldsymbol{M}\boldsymbol{X}^n + \Delta t\boldsymbol{F}_{\text{add}}. \tag{67}$$
Further, we will use a Runge-Kutta 2 and a Runge-Kutta 3 method:
$$\begin{aligned}
\boldsymbol{X}^{(1)} &= \boldsymbol{X}^n + \Delta t\boldsymbol{M}\boldsymbol{X}^n + \Delta t\boldsymbol{F}_{\text{add}}, \\
\boldsymbol{X}^{n+1} &= \frac{1}{2}\left(\boldsymbol{X}^n + \boldsymbol{X}^{(1)} + \Delta t\boldsymbol{M}\boldsymbol{X}^{(1)} + \Delta t\boldsymbol{F}_{\text{add}}\right),
\end{aligned} \tag{68}$$
and
$$\begin{aligned}
\boldsymbol{X}^{(1)} &= \boldsymbol{X}^n + \Delta t\boldsymbol{M}\boldsymbol{X}^n + \Delta t\boldsymbol{F}_{\text{add}}, \\
\boldsymbol{X}^{(2)} &= \frac{1}{4}\left(3\boldsymbol{X}^n + \boldsymbol{X}^{(1)} + \Delta t\boldsymbol{M}\boldsymbol{X}^{(1)} + \Delta t\boldsymbol{F}_{\text{add}}\right), \\
\boldsymbol{X}^{n+1} &= \frac{1}{3}\left(\boldsymbol{X}^n + 2\boldsymbol{X}^{(2)} + 2\Delta t\boldsymbol{M}\boldsymbol{X}^{(2)} + 2\Delta t\boldsymbol{F}_{\text{add}}\right).
\end{aligned} \tag{69}$$
Besides, we also test the Backward Euler method:
$$\left(\boldsymbol{I}^{Nd(p+1)} - \Delta t\boldsymbol{M}\right)\boldsymbol{X}^{n+1} = \boldsymbol{X}^n + \Delta t\boldsymbol{F}_{\text{add}}. \tag{70}$$
The last time integration method tested is a simple partitioned method. It consists of two steps:

1. The fluid system is evolved in time:

$$\boldsymbol{U}^{n+1} = \boldsymbol{U}^n + \Delta t \left[\left(\boldsymbol{A} + \boldsymbol{G}_{\text{LB}} - \boldsymbol{G}_{\text{RB}} + \boldsymbol{\mathcal{C}}_1 \right) \boldsymbol{U}^n + \left(\boldsymbol{B} + \boldsymbol{\mathcal{C}}_2 \right) \boldsymbol{V}^n \right]. \tag{71}$$

2. The spring system is evolved in time:

$$\boldsymbol{V}^{n+1} = \boldsymbol{V}^n + \Delta t \left(\boldsymbol{C}\boldsymbol{U}^{n+1} + \boldsymbol{D}\boldsymbol{V}^n \right) + \hat{\boldsymbol{F}}_{\text{ext}}. \tag{72}$$

Unless stated otherwise, a Runge-Kutta 3 method is used for the computations. We use the parameters chosen in Section 1 for the following simulations. The initial conditions are chosen to be

$$\rho_{i,j,k} = \bar{\rho}, \qquad \bar{\rho}_{i,j,k} = 0^{\text{kg}}/_{\text{m}^3}, \qquad x_{\text{wall}} = 0\text{m}, \qquad u_{\text{wall}} = 20^{\text{m}}/_{\text{s}}.$$

The oscillation of the piston is depicted in Figure 3. Since there is no damping in the system, no energy should dissipate at all. This means that the oscillation amplitude should remain constant in the mean over time. It can be clearly seen that in Figure 3 the amplitude of the oscillation is conserved quite well.

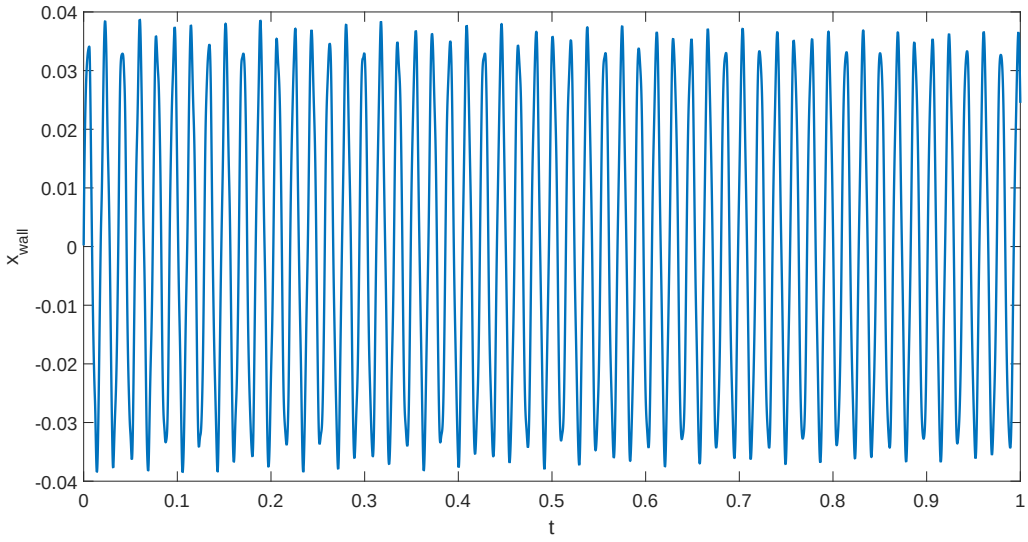


Figure 3: Position of the piston x_{wall} over time.

The eigenfrequency of the simulated system is $f_{\text{sim}} \approx 54.3763^1/_{\text{s}}$. This is very close to the eigenfrequency of the analytical approach in Section 1, which gives an eigenfrequency of $f \approx 54.3684^1/_{\text{s}}$.

The results look different when the correction terms $\boldsymbol{\mathcal{C}}_1$ and $\boldsymbol{\mathcal{C}}_2$ are not used. Then, the amplitude of the oscillation is increasing (Figure 4). The more cells are used, the less the impact of the instability is. Nevertheless, the amplitude keeps increasing.

In the left plot of Figure 5, the Lyapunov function of the monolithic formulation is plotted over time. The red line represents the case in which no new cells are added and removed when the piston is moving. In the other case (blue line), the computational domain is adapted to fit the piston position. Even though, the wall is moving, the domain is not adapted. This is an assumption in the analytical approach. In the right plot, the Lyapunov functions for the separated fluid (blue line) and solid (red line) parts as well as their sum (yellow line) are plotted

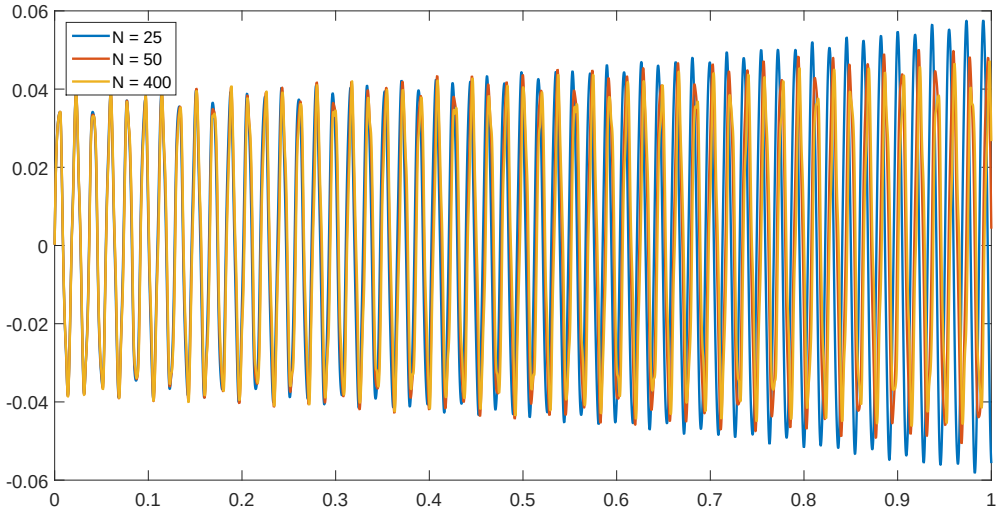


Figure 4: Position of the piston x_{wall} without the use of the correction terms $\boldsymbol{\mathcal{C}}_1$ and $\boldsymbol{\mathcal{C}}_2$.

over time. Even though, the Lyapunov functions of the separated regions look rather irregular, their sum has a smooth shape. The decline of the Lyapunov functions over time comes from numerical dissipation during the computation and also manifests itself in a decrease of the amplitude of the piston oscillation.

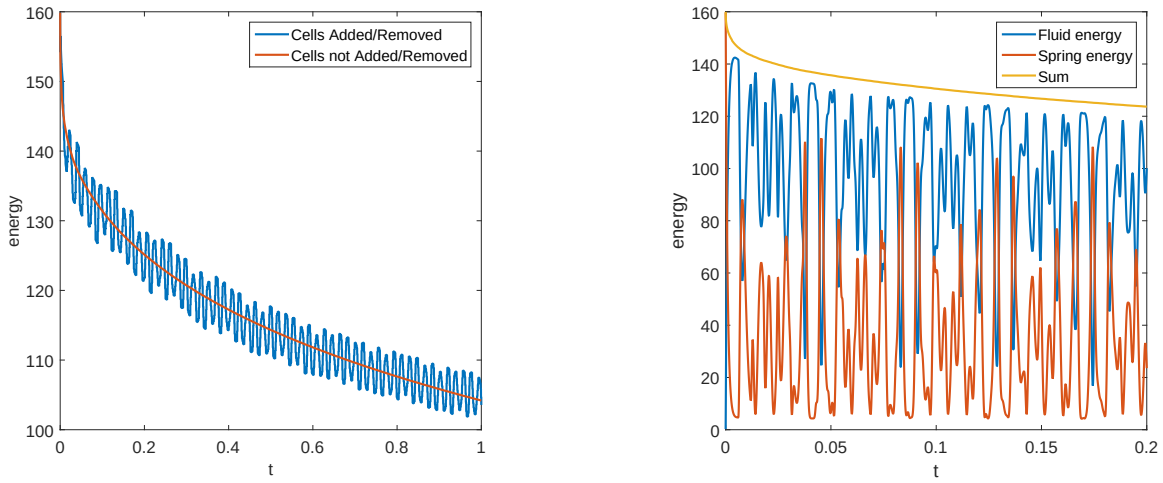


Figure 5: Left: Lyapunov functions $\mathcal{E}$ when cells are added and removed (blue line) and when the computational domain is set constant (red line). Right: Lyapunov functions for the fluid part (blue) and the structural part (red).

This (unphysical) decrease of the amplitude is small when the Discontinuous Galerkin method is used. In Figure 6 the results of a Discontinuous Galerkin method and a Finite-Volume method are compared. The Finite-Volume method can be interpreted as a Discontinuous Galerkin method with polynomial degree $p = 0$. It can be seen that the higher polynomial order gives much better results as the numerical dissipation is reduced considerably.

We now want to compare the different time integration schemes introduced in the beginning of this chapter. As can be seen in Figure 7, and as expected, the simple partitioned and the Forward Euler method are unstable. The Lyapunov functions increase rapidly in the very be-

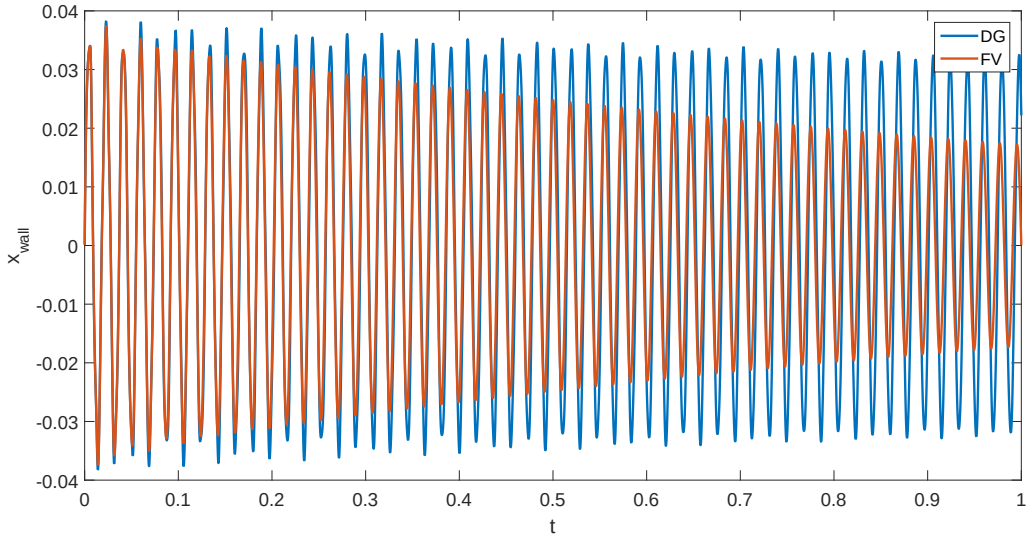


Figure 6: Position of the piston x_{wall} using the Discontinuous Galerkin method ($p = 1$; blue line) and the Finite-Volume method ($p = 0$; red line).

ginning. This effect goes along with a strongly increasing amplitude of the piston oscillation. In [2] some stable partitioned methods are presented. The Backward Euler, the Runge-Kutta 2 and the Runge-Kutta 3 methods can be numerically stable. They distinguish themselves by the amount of energy that is dissipated. We see in Figure 7 that the Runge-Kutta 3 method causes the least numerical dissipation.

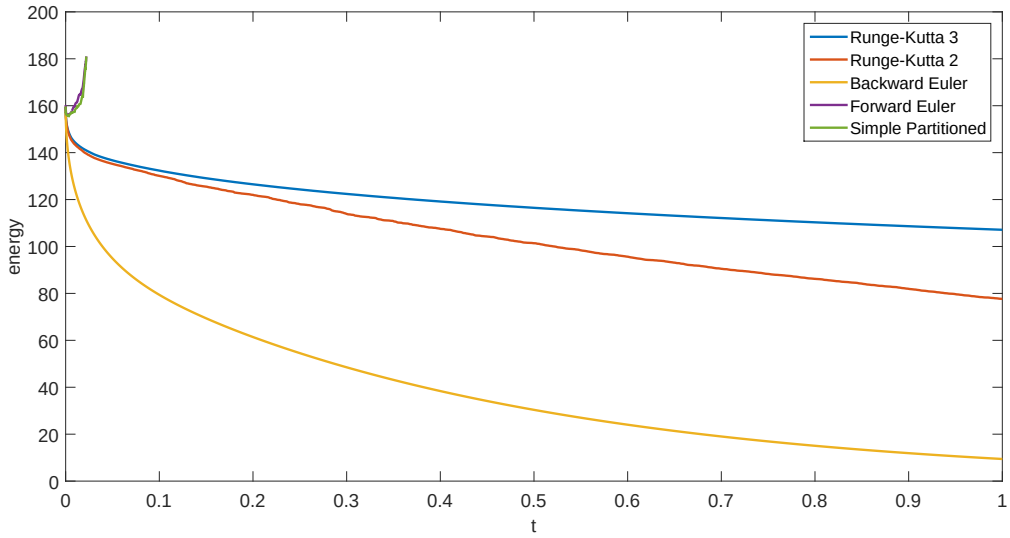


Figure 7: Lyapunov functions $\mathcal{E}$ for different time integration methods.

In the last plot, Figure 8, the oscillation of the piston is plotted over time, using the acoustic equations (blue line) and the Euler equations (red line). Especially in the beginning, the results are very much alike.

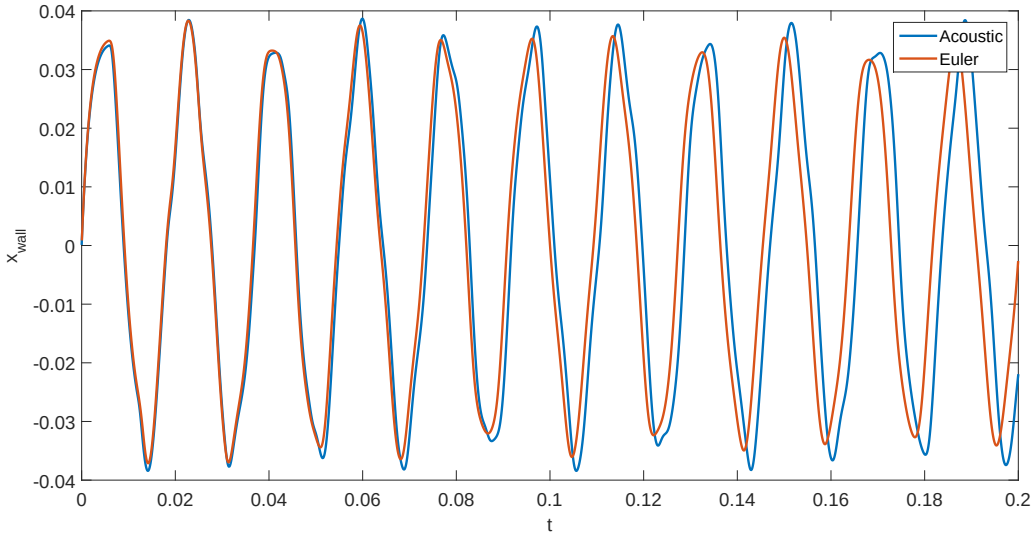


Figure 8: Position of the piston over time using the acoustic equations (blue line) and the Euler equations (red line)

5 CONCLUSION

We finish this work with a short summary:

- We successfully developed a monolithic simulation code for the piston problem.
- An analytical approximation of the frequency of the oscillation could be reproduced very well with our code.
- We introduced two correction terms $\mathcal{C}_1$ and $\mathcal{C}_2$ in order to achieve numerical stability for the coupling between the fluid and the structural part.
- The Discontinuous Galerkin method shows clear improvements towards the Finite-Volume method in terms of numerical dissipation.
- We tested different time integration methods on stability and numerical dissipation. In our tests, the Runge-Kutta 3 method shows the best results.

For the future, a more sophisticated boundary treatment and an extension to higher dimensions are desirable.

REFERENCES

[1] C.S. Peskin, Flow patterns around heart valves: A numerical method. *Journal of Computational Physics*, **10**, 252–271, 1972.

[2] S. Piperno, C. Farhat, B. Larrouturou, Partitioned procedures for the transient solution of coupled aeroelastic problems – Part I: Model problem, theory and two-dimensional application. *Computer Methods in Applied Mechanics and Engineering*, **124**, 79–112, 1995.

[3] S. Piperno, C. Farhat, Partitioned procedures for the transient solution of coupled aeroelastic problems – Part II: Energy transfer analysis and three-dimensional applications. *Computer Methods in Applied Mechanics and Engineering*, **190**, 3147–3170, 2001.

[4] F.J. Blom, A monolithical fluid-structure interaction algorithm applied to the piston problem. *Computer Methods in Applied Mechanics and Engineering*, **167**, 369–391, 1998.

[5] E.H. van Brummelen, S.J. Hulshoff, R. de Borst, Energy conservation under incompatibility for fluid–structure interaction problems. *Computer Methods in Applied Mechanics and Engineering*, **192**, 2727–2748, 2003.

[6] C. Michler, S.J. Hulshoff, E.H. van Brummelen, R. de Borst, The relevance of conservation for stability and accuracy of numerical methods for fluid–structure interaction. *Computer Methods in Applied Mechanics and Engineering*, **192**, 4195–4215, 2003.

[7] C. Michler, S.J. Hulshoff, E.H. van Brummelen, R. de Borst, A monolithic approach to fluid–structure interaction. *Computers & Fluids*, **33**, 839–848, 2004.

[8] J.S. Hesthaven, T. Warburton, *Nodal Discontinuous Galerkin Methods*. Springer-Verlag New York, 2008.

[9] E.F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction, 3rd Edition*. Springer-Verlag, 2009.

[10] S. Wiggins, *Introduction to Applied Nonlinear Dynamical Systems and Chaos, Vol. II, 2nd Edition*. Springer-Verlag New York, 2003.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SLAMMING IMPACT SIMULATION OF 2D WATER ENTRY FOR RIGID STRUCTURES

O. Hashim Hassoon[1], M. Tarfaoui[2], A. El Malki[3]

[1]ENSTA Bretagne, MSN/LBMS/DFM, 2 Rue François Verny, 29806, Brest Cedex 9, France
omar_hashim.hassoon@ensta-bretagne.fr

[2]ENSTA Bretagne, MSN/LBMS/DFM, 2 Rue François Verny, 29806, Brest Cedex 9, France
mostapha.tarfaoui@ensta-bretagne.fr

[3]ENSTA Bretagne, MSN/LBMS/DFM, 2 Rue François Verny, 29806, Brest Cedex 9, France
Aboulghit.el_malki_alaoui @ ensta-bretagne.fr

Keywords: slamming impact, 2D water entry, ALE, pressure distribution, hydrodynamic forces.

Abstract. *Slamming impact is considered important factor for ship design engineers to predict the structural reliability under maximum pressure and hydrodynamic forces in a small duration of time. In this paper, we deal with rigid bodies by simulated 2D water entry problem based on Arbitrary Lagrangian Eulerian (ALE) built-in explicit finite element in Abaqus software. To validate the efficiency of models, various deadrise angles are applied from 4° to 80° for wedge. The prediction of maximum pressure and hydrodynamic forces are compared with the analytical formulations of the rigid body. Many factors are affecting on slamming phenomenon due to high deformation in the fluid. Therefore we have delineated the effect of the contact stiffness factor between the fluid and structure in penalty contact method, which influences on the results. Fluid domain mesh sizes convergence are applied with take in consideration the hourglass control effect, showed that the effect of the density of the mesh has high influence on the amplitudes stability for both pressure and force.*

1 INTRODUCTION

Slamming is the water impact loads that considered important in the structural design of all high-speed vessels and can cause both local and global effects in the ship structure, the global effect is often called whipping. Hydrodynamic may be important for global loads but also for local loads in the case of very high slamming pressure with very short duration [1].

The loads occur during the slamming of the body onto a water surface have relevance for the design of many vessels, unfortunately, the analysis of this scenario is challenging for several reasons. During impact, the free surface is characterized by a thin jet with velocities that are much larger that the body velocity. This means that free surface undergoes stretching and its topology becomes extremely complex [2].

The major interesting of the ship engineers is to find the analytical solution capable to provide the hydrodynamic load and the impact pressure distribution when the slamming event happens, since these pressures can be used during the design process [3]. The slamming pressure distribution and hydrodynamic forces can be used in the static structure analysis that locates the local impact induced stresses. This done when the local deadrise impact angle between the fluid and the structure is not very small at the impact position, and suppose that hydrodynamic calculations for a rigid body [1]. Several assumptions are adapted to calculate the pressure for the rigid body such as the airflow. Usually, these methods assume that the fluid is incompressible and irrotational. Von Karman was considered the first author attempted to calculate analytically the slamming phenomenon based on the mass conservation, but is not accounted the effect of the jet up-rise flow [4]. The method based on Wagner theory allowed taking in account the effect of the jet up-rise flow of the fluid, which assume the impact of the blunt body [5].

Kaushik et al [6] studied the water slamming of deformable sandwich panels using the commercial FE software LS-DYNA with the ALE formulation. All geometric nonlinearities are considered when studying deformations of the composite panel, assumed the fluid to be compressible. However, the authors have accounted inertia effects in the fluid and the solid, and examined delamination between the core and the face sheets. The pressure distribution on the wetted panel surface was found to be oscillatory. Peseux et al [7] have solved numerically two and three dimensional Wagner theory using finite element method for the rigid and deformable structures. A series of tests were performed experimentally for different deadrise angles and thickness, and analysis the distribution and evolution of hydrodynamic pressure. Stenius et al. [8] used numerical simulations of the hydroelastic problem related to panel-water impacts for high-speed craft, by applying different deadrise angles and velocities and different boundary conditions and panel properties. Siyauan & Mahfuz [9] used the finite element analysis to study the fluid structure interaction (FSI) for the sandwich structure. This performs by coupling between the finite element analysis (FEA) model with the computational fluid dynamic (CFD) model. The global model was used to construct the composite and the foam of the sandwich structure, then transporting the force and displacement for this model to the sub-model with refine mesh. The interlaminar and the shear stress distribution are determined at the core and faces. Yang and Qiu [10] has been computed slamming forces on two and three dimensional bodies entering the calm water with constant velocity, employing the Constrained Interpolation Profile (CIP) method. The nonlinear water entry problem is governed by the Navier-Stokes equations and solved by a CIP-based finite difference method on a fixed Cartesian grid. Wang and Soares [11] have investigated the water impact problem for three-dimensional buoys (hemisphere and cones) using ALE solver based on Eulerian formulation in LS-DYNA software. They concluded that the mesh density of the fluid is very influ-

ence factor on determine the hydrodynamic pressures. Furthermore, they noted that the influence of the constant impact velocity is larger than ones from the drop velocity when compared its total impact force and hydrodynamic pressure.

In this paper, a parametric study is carried out to provide the hydrodynamic load and the impact pressure distribution when the slamming event is happened with constant velocity. Indeed, the response of the structure will depend on the parameters of the numerical calculation inputs for describing the fluid-structure coupling.

2 MATHEMATICAL FORMULATION FOR ALE MODEL

ALE model consists of both the Lagrangian and Eulerian model, materials arbitrary coordinates are defined corresponding to the reference or global coordinate system. The relationship between these coordinates and the reference coordinate relative with the time are constructed the ALE model as described by equation (1) [12]

$$\frac{df(X_i,t)}{dt} = \frac{df(x_i,t)}{dt} + w_i \frac{df(x_i,t)}{dxi} \tag{1}$$

Where X_i, xi and $w_i = (v - u)$ are the Lagrangian coordinate, Eulerian coordinate and relative velocity respectively. From definition of relative velocity as a difference between the material velocity (v) and the mesh velocity (u) , formulations of the model are depended on the conservation equations as followed:

2.1 Conservation of the mass

For point of view of the conservation of the mass which means that the material density is constant relative with time and can be described as:

$$\frac{d\rho}{dt} = \rho \frac{dv_i}{dx_i} - w_i \frac{d\rho}{dx_i} \tag{2}$$

Where ρ is the density of the fluid.

2.2 Conservation of momentum

In conservation of the energy, the governing equation must be specified the appropriated initial boundary conditions to represent the ALE description corresponding to Navier-stock equations:

$$\rho \frac{dv_i}{dt} = \sigma_{ij,j} + \rho b_i - \rho w_i \frac{dv_i}{dx_j} \tag{3}$$

The stress tensor σ_{ij} is described as follows:

$$\sigma_{ij} = -p\delta_{ij} + \mu(v_{i,j} + v_{j,i}) \tag{4}$$

Where μ, is the dynamic viscosity and δ_{ij} is the Kronecker delta function and p is the pressure.

Define the initial boundary conditions for the fluid and the body domain as shown in figure (1), the equations can be solved.

$v_i = U_i^0 \ \ on\ \Gamma 1$ $\qquad$ $\sigma_{ij} n_j = 0 \ \ on \ \ \Gamma 2$

$\Gamma 1 \cap \Gamma 2 \ = \ 0$ $\qquad$ $\Gamma 1 \cup \Gamma 2 \ = \ 0$

$V_i(x_i, 0) \ = \ 0$ $\qquad$ on $t = 0$

Where

Γ1 is the boundary of the body,

Γ2 is the boundary of the fluid,

nj is the outer unit normal vector on the boundary of Γ2,

δ_{ij} is the Kronecker's delta function.

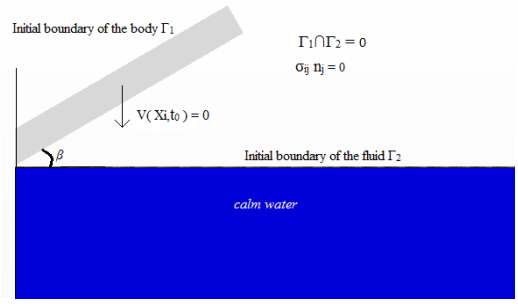


Figure 1: Initial boundary conditions for fluid and body domain

2.3 Energy equation

$$\rho \frac{dE}{dt} = \sigma_{ij} v_{i,j} + \rho b_i v_i - \rho w_j \frac{dE}{dx_j} \tag{5}$$

The velocity of reference coordinate is assumed zero in Eulerian equation, while the relative velocity are computed in advection phase to transfer from previous mesh to the new ma-

terial mesh. For this reason, ALE equations are considered much more difficult numerically than the Lagrangian equations, where the relative velocity is zero.

ALE governing equations are implemented in two phases: the first is applied to the Lagrangian phase that changes of the velocity in the mesh moves with the material. Consequently an internal energy due to the internal and external forces is calculated. The equilibrium equations are:

$$\rho \frac{dv_i}{dt} = \sigma_{ij,j} + \rho b_i \tag{6}$$

$$\rho \frac{dE}{dt} = \sigma_{ij} v_{i,j} + \rho b_i v_i \tag{7}$$

Mass is conserved in the Lagrangian phase, since no material flows across the element boundaries. The advection phase is followed to determine a transport of mass, internal energy and momentum across cell boundaries. This may be thought of as remapping the displaced mesh at the Lagrangian phase back to its initial or arbitrary position.

3 NUMERICAL INVESTIGATION OF THE SLAMMING PHENOMENON

For the slamming event with small period duration, large deformations of the fluid are happening, which caused the mesh distortion. Therefore, an automatic re-mesh method requires for the fluid domain, which consumes a high computational time. For this reason the suitable simulation of the interaction between the fluid and the structure is implemented. Arbitrary Lagrangian Eulerian (ALE) solver in ABAQUS software are used to allow the mesh to move and back to reference position during translating to another step. The material in the ALE region flowed with the mesh move. Hence, to prevent the element mesh destroy, a remap of the mesh are adapted using an advection phase to update fluid velocity and material variables for the new mesh.

ALE is a mixture of Lagrangian and Eulerian discretization. The region in the fluid that close to the impact area has a refine meshes size to govern the large deformation of the fluid in this position. While the outer region of this domain used another mesh size because the deformation in the outer domain of impact position is still moderate. Figure 2 illustrates the boundary conditions and the mesh part model for the water entry problem, which are tested to reduce the computational time as reason of high time consuming of coupling between the fluid and structure.

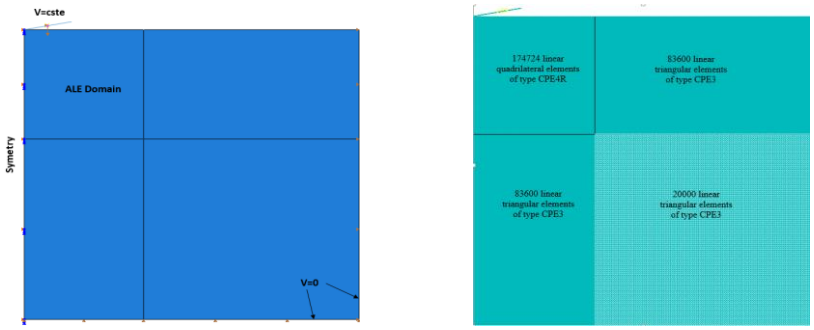


Figure 2: Boundary conditions and mesh regions for the slamming of two-dimensional wedge model.

For the slamming of the symmetric wedge, symmetrical boundary condition is applied about the symmetric plane with breadth of section 0.25 m and length 0.4 m. To prevent reflected waves effect from external boundaries, fluid domain length is specified as [13]:

$$L \geq \frac{c_0 T}{2} \tag{8}$$

Where L is the distance between the impact position and the external surface of the water domain, T is the time of the slamming event and c0 is the speed of the sound in the fluid. In Abaqus explicit, EOS can be defined by linear Us - Up formulation of the Mie-Gruneisen equation of state which can be exploited in water entry problem. Pressure can be determined as a function of the density and the internal energy as:

$$P = f(\rho, E_m) \tag{9}$$

$$P = PH\left(1 - \frac{\Gamma_0\ \eta}{2}\right) + \Gamma_0\ \rho_0 E_m \tag{10}$$

Where PH, Γ_0 , η =$1 - \rho_0/\rho$ and ρ_0 is the Hugoniot pressure, material constant, nominal volumetric compressive strain and fluid reference density respectively.

$$PH = \frac{\rho_0 c_0 \eta}{(1 - s\eta)^2} \tag{11}$$

Where s is represented the linear relationship between shock velocity Us and particle velocity Up (Us= c0 + s Up).

Then the pressure of the fluid can be calculated as:

$$P = \frac{\rho_0 c_0 \eta}{(1 - s\eta)^2}\left(1 - \frac{\Gamma_0\ \eta}{2}\right) + \Gamma_0\ \rho_0 E_m \tag{12}$$

The water parameters are illustrated in table (1). For the modelling of the contact solid/fluid, the exponential soft contact with frictionless tangential behavior has been adopted. The mesh convergence is investigated in the region close to the impact position with element size 0.5mm for the fluid domain.

C_0 (m/sec)	μ (-)	s (-)	Γ_0 (-)	ρ(kg/m^3)
1420	0.001	0	0	1000

Table 1: Parameters of the water in the model.

3.1 Effect of the hourglass control

The mesh of Lagrangian domain deforms with the material, while the Eulerian mesh is fixed in the space, add the material flows from one cell to another. Lagrangian is easier to handle (particularly the definition of boundary conditions), but with large mesh deformation and highly distorted leading to inaccurate calculations, or even failure.

Hourglassing is essentially a spurious deformation mode of a Finite Element Mesh, resulting from the excitation of zero-energy degrees of freedom. It typically manifests as a patchwork of zig-zag or hourglass like element shapes (Figure 3), where individual elements are severely deformed, while the overall mesh section is non-deformed.

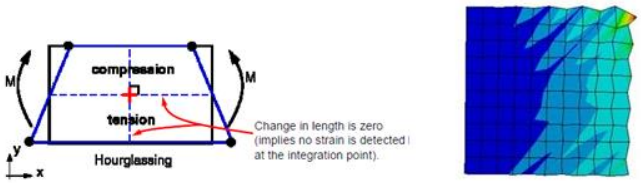


Figure 3: Hourglass behaviour for a single first-order reduced-integration element with no strain.

In this section the hourglass effect is investigated, built in artificial stiffness of the elements to prevent this kind of deformation. There are five element controls (combine, Enhanced, relax-stiffness, stiffness and viscous) that are added as artificial stiffness for the elements. By verify the total artificial energy that used to control hourglass is small (<1) relative to the total internal energy, as shown in table (2).

contact	Hourglass control	Artificial energy (ALLAE)	Internal energy (ALLIE)	Relative difference <1
Kinematic method	Combine	19,288	46,878	**0,588**
	Enhanced	0,006	14,288	0,999
	Relax-stiffness	0,006	14,356	0,999
	stiffness	0,007	14,277	0,999
	viscous	43,809	67,109	**0,347**
Penalty method (scale factor = 0.05)	Combine	37,138	56,518	**0,342**
	Enhanced	0,017	15,571	0,998
	Relax-stiffness	0,017	15,571	0,998
	stiffness	0,004	15,795	0,999
	viscous	44,559	59,940	**0,256**

Table 2: Comparison of the energy histories of hourglass element controls.

By comparison, values of the energy histories (artificial and internal energy) for two types of the contact methods with different hourglass controls are illustrated in table (2). Total hydrodynamic forces and the hydrodynamic pressure along the length of the body contact surface are also compared for their amplitude values and oscillation as shown in figure 4. From these results, we can conclude that the suitable contact coupling is penalty contact method with two best hourglass controls (combine and viscos). As a result, they have minimum difference of energy and minimum vibration for both hydrodynamic pressure and hydrodynamic force. The simulation results have a good agreement with analytical approaches.

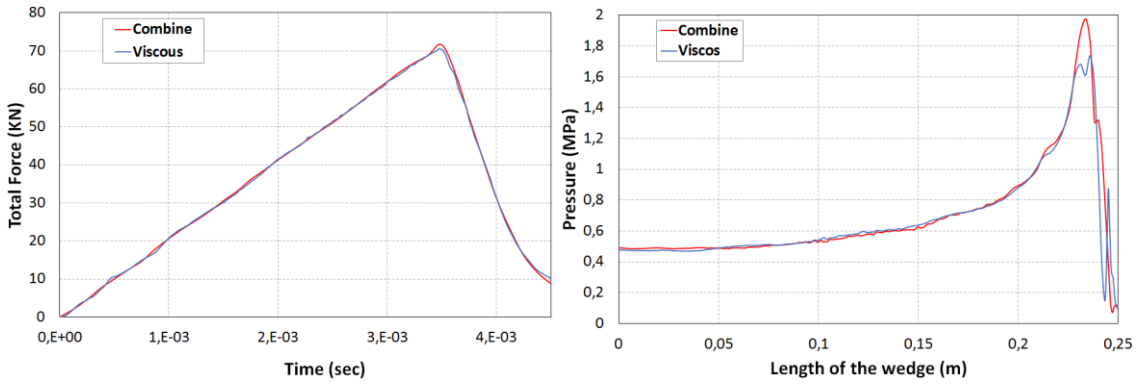


Figure 4: Total Force and Pressure distribution due to the contact pressure on the wedge surface with penalty contact method with deadrise angle 10°, impact velocity 8 m/s.

3.2 Effect of the penalty stiffness factor (PFAC)

The penalty stiffness factor (PFAC) is a factor for scaling the contact stiffness of the two pairs of the contact surfaces, which applies to nodes of the water-wedge interface to estimate the coupling force. The penalty coupling method is working as a spring system in each of both ends that attached to the structure and fluid nodes. The coupling force calculates relative with the penetration rate.

$$F = Kd \tag{13}$$

Where *K* and *d* are representing the spring system stiffness and penetration respectively.

The spring stiffness is depended on the scaling factor, bulk modulus of the fluid and the mesh size of the fluid as descripted in equation (14):

$$K = p_f \frac{k\,A}{V} \tag{14}$$

Where p_f , A and V are penalty scale factor, the average area of the structure element and the volume of the fluid element, which are in the coupling state.

By applying a range of penalty scale factor (1, 0.2, 0.1 and 0.05), a comparison study on slamming force and pressure distributions is performed. In this model, the default value 0.05 is applied as optimum value according to amplitude and profile of total force and pressure distribution as shown in figure 5. Notes this factor has small influence on results compared with mesh size density effect. When used a high value of scale factor, coupling stiffness has been increased between nodes of fluid-structure interface. Consequently, it causes a high oscillation in pressure and the force without penetration. In the other hand, low value of scale factor leads to decreasing stiffness. In this situation, a non-physical penetration appears and leakage will be done. The free water surface elevation and pressure histories for 10° deadrise angles is illustrated in figure 6.

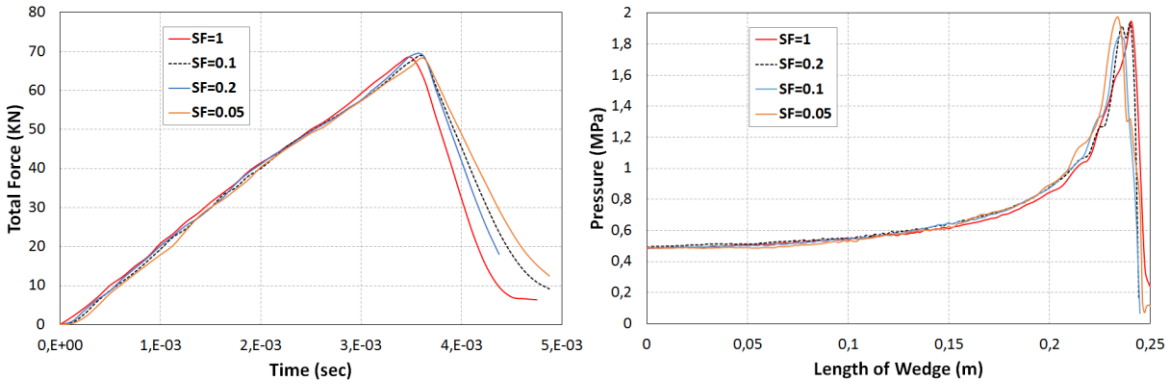


Figure 5: Total Force and Pressure on the wedge surface with different stiffness scale factors of penalty contact method with deadrise angle 10°, impact velocity 8 m/s.

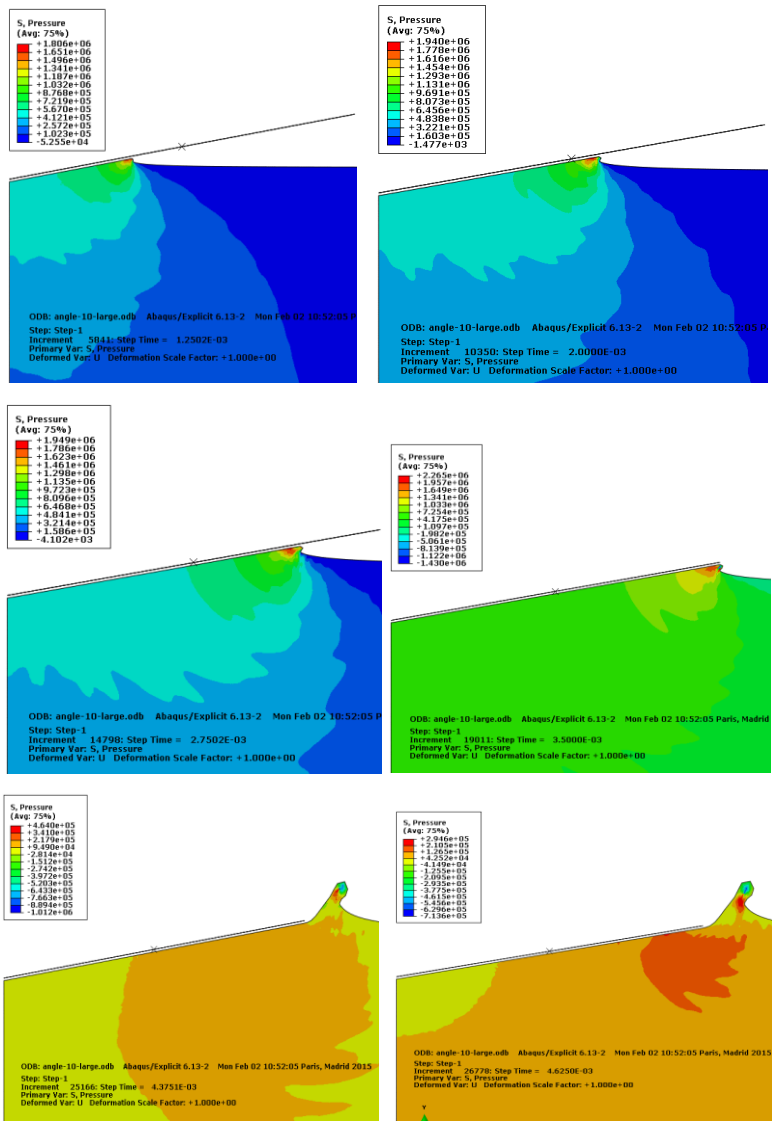


Figure 6: Free water surface and pressure histories for deadrise angle 10°, impact velocity 8 m/s.

3.3 Deadrise angles effect on the hydrodynamic pressure

Slamming phenomena happens in short time duration. For this situation, a high magnitude peak pressure occurs, which is considered very important for the ships design in the naval applications. This phenomenon can be cause global and local damages in the structure due the interaction between the structure and the fluid. The pressure is travelling along the panel width that largely dependent on the velocity impact and the angle between the structure and the fluid. Hydrodynamic effects are important when the deadrise angle between the body and the water surface is small [3].

The kinematic boundary of the fluid-structure is changing when the structure responds to the dynamic pressure, this kinematic coupling between the fluid and the structure is a type of hydroelastic effects. For the rigid body, the pressure decreases relatively with increasing of the deadrise angle under constant velocity. Indeed, that the wedge cuts more easily through the water, which is shown in figures 7 and 8. In these figures, a good agreement is observed with the analytical methods [14], obviously that the coefficient pressure is decreasing according to increase the deadrise angle. In equation (15), the non-dimensional pressure corresponding to the deadrise angles has been calculated as:

$$Cp_{max} = \frac{p_{max}}{0.5\rho\, v^2} \tag{15}$$

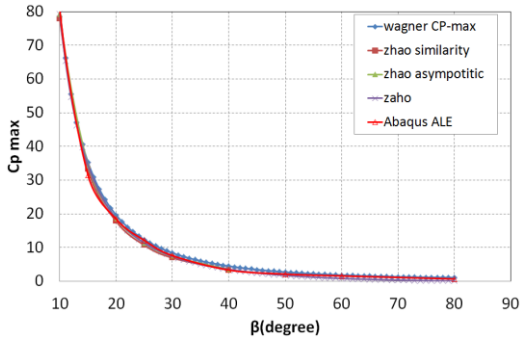
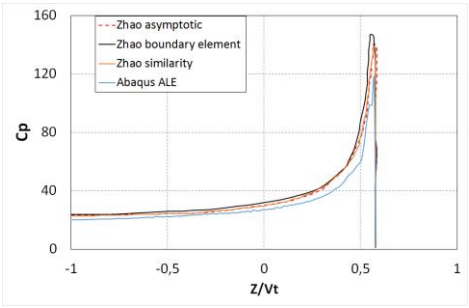
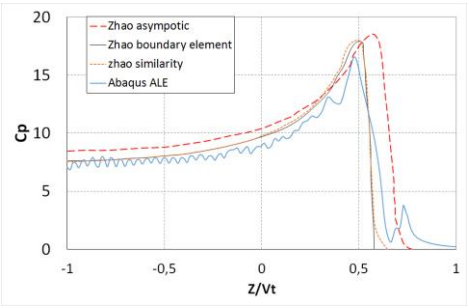


Figure 7: Non-dimensional coefficient pressure Cp_{max} according the deadrise angles



(a) 7.5°



(b) 20°

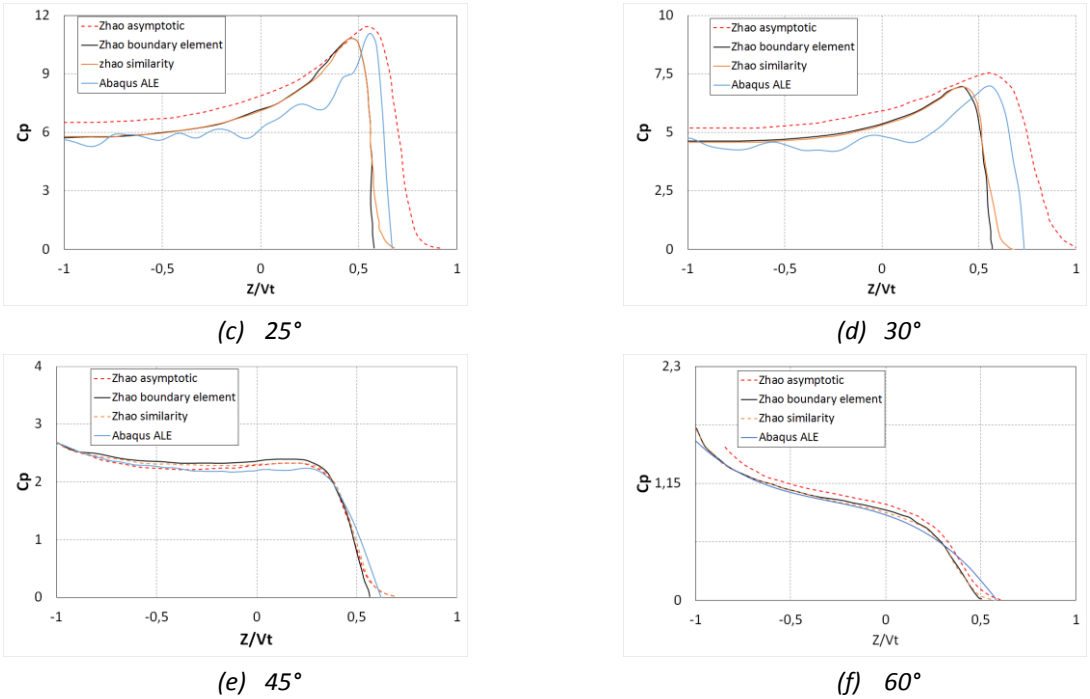


Figure 8: Coefficient pressure distribution for arrange of deadrise angles with constant velocity 13 m/s.

4 CONCLUSIONS

In this study we are simulated the water entry for 2D rigid structure by using ALE (Arbitrary Lagrangian-Eulerian) solver, which can be represented the solid deformation as fluid with water properties. ALE combines the properties of Lagrangian and pure Eulerian using the adaptive mesh technique, and can be remap the mesh with each step increment to prevent the distortions of the mesh due to high deformations. This technique is depending on the coupling contact, therefore, penalty contact method have been used. By defining the appropriate parameters for this method such as stiffness scaling factor and mesh density, it can be satisfied the stability and prevented the penetration between the rigid body and the water. When compared these simulations with previous methods, a good agreement can be observed between the measured impact pressures for deadrise angles large than 30° and the analytical methods. Noted that high pressure peak close to move to the keel of the wedge, and it was dropped with increase the deadrise angle. For the deadrise angles less than 10°, some difference are happen compared with the analytical methods, and pressure with sharply peak occurs close to the jet flow region. The pressure peak for a wedge body is stay unchanged approximately along the wedge-water interface with time histories and it is reducing after the flow separation happens.

REFERENCES

[1] M.Faltinsen, «Slamming, Whipping, and Springing,» chez *Hydrodynamics of High-Speed Marine Vechicles*, New york, USA, Cambrige University Press, 2005, pp. 302-311.

[2] Kevin J.Maki, DongheeLee, ArminW.Troesch & Nickolas Vlahopoulos, «Hydroelastic impact of a wedge-shaped body,» *Ocean Engineering*, vol. 38, p. 621–629, (2011).

[3] R. Panciroli, «Hydrodynamic impact of deformable wedges,» chez *Dynamic Failure of Composite and Sandwich Structrure*, New Yourk, London, Springger Science, Business Media Dordrecht, 2013, pp. 1-45.

[4] von Ka´rma´n, T., «The impact of seaplane floats during landing,» NACA TN., 1929.

[5] H. Wagner, «Uber stoss und gleitvorgange an der oberflach von flussigkeiten,» *Zeitschrift fuer Angewandte Mathematik und Mechanik,* vol. 12, pp. 193- 215, 1932.

[6] Batra, Kaushik Das & Romesh C., «Local water slamming impact on sandwich composite hulls,» *Journal of Fluids and Structures,* vol. 27, pp. 523- 551, 2011.

[7] B. Peseux, L. Gornet & B. Donguy, «Hydrodynamic impact: Numerical and experiment investigations,» *Journal of Fluids and Structures,* vol. 21, pp. 277- 303, 2005.

[8] I. Stenius, A.Rose & J.Kuttenkeuler, «Hydroelastic interaction in panel-water impacts of high-speed craft,» *Ocean Engineering,* vol. 38, pp. 371- 381, 2011.

[9] Hassan Mahfuz, Siyuan Ma. , «Finite element simulation of composite ship structures with fluid structure interaction,» *Ocean Engineering,* vol. 52, p. 52–59, 2012.

[10] Q. Yang & W. Qiu, «Numerical simulation of water impact for 2D and 3D bodies,» *Ocean Engineering,* vol. 43, pp. 82- 89, 2012.

[11] S. Wang & C.G.Soares, «Numerical study on the water impact of 3D bodies by an explicit finite element method,» *Ocean Engineering,* vol. 78, pp. 73-88, 2014.

[12] N. Aquelet & M. Souli, «A new ALE formulation for sloshing analysis,» *Structural Engineering and Mechanics,* vol. 16, n°. 14, pp. 000-000, 2003.

[13] A. Tassin, N.Jacques, A. El malki, A. Neme & B. Leblé, «Assessment and comparison of several analytical,» *Int. Jnl. of Multiphysics,* vol. 4, n°. 12, pp. 125-140, 2010.

[14] R.Zhao, O. Faltinsen, «Water entry of two-dimensional bodies,» *Fluid Mech,* vol. 246, pp. 593-612, 1993.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A MESH MORPHING BASED FSI METHOD USED IN AERONAUTICAL OPTIMIZATION APPLICATIONS

M. Andrejašič[1], D. Eržen[1], E. Costa[2], S. Porziani[2], M. E. Biancolini[3], and C. Groth[3]

[1] Pipistrel d.o.o. Ajdovščina, R&D, Department of Aerodynamics, Slovenia
e-mail: matej.andrejasic@pipistrel.si, david.erzen@pipistrel.si

[2] D'Appolonia S.p.A. Italy
e-mail: emiliano.costa@dappolonia.it, stefano.porziani@dappolonia.it

[3] University of Rome Tor Vergata (UTV), Italy
e-mail: biancolini@ing.uniroma2.it, corrado.groth@uniroma2.it

Keywords: Fluid Structure Interaction, Modal Analysis, Radial Basis Function, Mesh Morphing, Aeronautical Optimization.

Abstract. *In this paper a fast and efficient mesh morphing based technique to perform FSI analyses for aeroelastic design and optimization applications is presented. The procedure is based on the finite volume CFD solver (OpenFOAM® and SU2) coupled with the RBF Morph™ tool capable of deforming the surface and volume mesh of the computational domain according to the mode superposition method. Structural vibration modes of the geometry of interest are calculated in a pre-processing stage by means of a FEM solver and later imported into the RBF Morph™ tool to create a set of individual basic deformations. Aerodynamic loads calculated with a CFD solver are then projected onto the accounted structural modes to get modal loads and modal coordinates which are applied to the computational model in order to obtain the deformed configuration. An FSI cycle incorporating a CFD simulation and morphing of its mesh can be iteratively repeated upon convergence to the final deformed shape. Since the modal parameterization and the mesh calculation have to be prepared only once per FSI analysis, its computation time is drastically reduced compared to a standard two-way coupling method in which a structural analysis has to be done at each cycle. Present procedure was applied to two geometries, HIRENASD fuselage-wing geometry for the purpose of testing the procedure and a Pipistrel's electric aircraft propeller for the purpose of optimization of its shape. By utilizing a DoE and a response surface method an increase of four percent of propeller efficiency was obtained by converging to a most favourable propeller pitch and twist configuration incorporating also FSI deformation. The abovementioned procedure was developed in the framework of the EU-funded RBF4AERO project (Grant Agreement No: 605396) and is available through the RBF4AERO platform.*

1 INTRODUCTION

An increasing demand for a faster design phase, optimized final product and reduction of the cost of production in today's world requires a constant development of high fidelity computer-aided engineering (CAE) tools and embedded numerical methods. Multidisciplinary numerical optimization procedures, where multiple physical phenomena are treated at the same time, are therefore desired because they let the engineers to determine their combined effect with a single analysis.

Fluid-structure interaction (FSI) is an example of such multi-physics phenomena where an interaction of a surrounding or internal fluid flow with the movable or deformable structures occurs [1]. FSI is in particular important in aeronautical design, since the safety issue is of the most important concern. All aerodynamic surfaces deform to some extend under aerodynamic forces. Putting aside large scale deformations that cause a malfunction of an airplane, even smaller deformations can cause the airplane to fly at non-optimal configuration. These deformations therefore shouldn't be neglected during the design phase.

In general, all FSI approaches can be grouped according to their governing equations onto monolithic approaches (equations governing the flow and the displacement of the structure are coupled and solved simultaneously) and partitioned approaches (equations are solved separately with two distinct solvers), and according to the treatment of the meshes onto matching and non-matching mesh approaches [2]. For a strong interaction between the fluid and the structure monolithic approaches should be used [3], whereas in cases when the interaction is week a simpler partitioned approach, with a typical two-way approach example [4], can be used. In order to tackle FSI problems of small deformations the FSI strategy can be simplified even further by incorporating a modal approach and thus avoiding the iteration between the computational fluid dynamics (CFD) and finite element method (FEM) solver [5].

The idea behind the modal approach is that the geometry of interest deforms according to a combination of a limited number of structural mode shapes. FEM calculation therefore has to be performed only once, which leads to a faster overall FSI analysis. Structural deformations are then taken into account directly by morphing the shape of the geometry and the volume mesh surrounding it. Since radial basis functions (RBFs) have been proven in the past as a robust choice for mesh morphing [6-8], they are used as an interpolation basis also in the present work. Morphing is done by exploitation of RBF Morph™ software that has been shown as a powerful and effective tool for solving challenging aerospace [9-12] and non-aerospace [13, 14] engineering applications.

A mode-superposition based FSI strategy is presented in this paper. The strategy is applied to a real engineering problem and thus proven as a robust and effective optimization method. The theory behind the modal analysis and RBF-based morphing is presented in Section 2 and 3, respectively. Overall FSI strategy with three distinctive stages is presented in Section 4. The strategy is validated in Section 5 and exploited in an aeronautical optimization application in Section 6.

2 MODE-SUPERPOSITION METHOD

Mode-superposition method is a powerful and well-established theory in structural engineering. It is applicable to both free vibration and forced vibration dynamic problems, regardless of the type of the system (continuum or discrete) being treated.

The basic idea of the theory is to use undamped free vibration modes of a structure in order to uncouple the equations of motion. Each vibration mode consists of a natural mode shape, characterizing one of the displacement patterns, and its associated natural frequency. Modal

analysis is therefore a linear theory that enables a determination of the structural static and dynamic response of a discretized system.

The method basically consists of three steps. First, mode shapes and frequencies of a system need to be found performing a modal analysis on the FEM model. The response of each mode under aerodynamic loading is then calculated in the second step. The third step concludes with a superposition of a chosen number of responses to find the full modal response to a given loading.

General equations of motion of an undamped discrete system with n degrees of freedom (DOF) can be written as

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f}(t), \tag{1}$$

where $\boldsymbol{x}$ represents a displacement vector, $\boldsymbol{M}$ is the mass matrix of the system, $\boldsymbol{K}$ is its stiffness matrix and $\boldsymbol{f}(t)$ is external excitation vector or loading vector. Assuming an exponential solution $\boldsymbol{x} = \boldsymbol{u}\, e^{i\omega t}$ Eq. (1) can be rewritten into the following eigenvalue problem [15]

$$\boldsymbol{K}\boldsymbol{u} = \omega^2 \boldsymbol{M}\boldsymbol{u}, \tag{2}$$

where no external forces were considered. By solving Eq. (2) a complete set of eigenvectors $\boldsymbol{u}$, that represent mode shapes, and eigenvalues ω^2, that lead to natural frequencies ν as $\omega = 2\pi\nu$, for all DOF can be obtained. Eq. (2) therefore states that a vibration mode is a configuration in which a balance between elastic resistance and inertial loads occurs.

According to the eigendecomposition operation all eigenvectors, i.e. mode shapes, are orthogonal and form a basis of the eigenvalue problem [16]. The dynamic response of a mechanical system can therefore be represented by a superposition of all mode shapes as

$$\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\eta}, \tag{3}$$

where matrix $\boldsymbol{U}$ and vector $\boldsymbol{\eta}$ represent a complete set of modal shapes and modal coordinates, respectively. Mode superposition is therefore a weighted sum of all mode shapes, where modal coordinates represent the weights.

A convenient normalization can be obtained by imposing a unit modal mass for each mode shape $\boldsymbol{u}_m$, which, taking into account also the orthogonality of the basis, results into

$$\boldsymbol{u}_m^T \boldsymbol{M}\boldsymbol{u}_m = 1, \qquad m = 1,2,\dots,n. \tag{4}$$

By inserting Eq. (4) into Eq. (2) an additional simplification can be obtained as

$$\boldsymbol{u}_m^T \boldsymbol{K}\boldsymbol{u}_m = \omega_m^2. \tag{5}$$

Equations of motion (Eq. (1)) can be, using Eqs. (3), (4) and (5), finally rewritten into n uncoupled and independent single DOF systems [17]

$$\ddot{\eta}_m + \omega_m^2 \eta_m = \boldsymbol{u}_m^T \boldsymbol{f}(t) \tag{6}$$

that can be solved one at the time in order to obtain all needed modal coordinates η_m . Each nodal force defined as

$$N_m(t) = \boldsymbol{u}_m^T \boldsymbol{f}(t) \tag{7}$$

can be therefore obtained by multiplying mode eigenvector $\boldsymbol{u}_m^T$ by a real physical force distribution $\boldsymbol{f}$(t). The complete response of a system to a given loading can finally be obtained using Eq. (3), with all entities known.

The number of mode shapes of a structure is in reality equal to its total number of DOF, i.e. infinite for a continuum system. In practice it is not necessary to employ all mode shapes in order to describe the system. Since the energy associated with each mode decreases with in-

creasing frequency, a good approximation can be obtained by superimposing only first few most energetic mode shapes n_{modes} that tend to provide the greatest contribution to structural response [15].

Modal approach is typically exploited in dynamic analyses in which the number of modes retained is defined on the basis of excited frequencies. Nevertheless, it can be used even for approximating a static solution. In this case, Eq. (6) can be simplified to

$$\omega_m^2 \eta_m = N_m. \tag{8}$$

3 RBF-BASED MORPHING

Proposed FSI strategy employs mesh morphing done in RBF Morph™ tool. The morphing is based on the RBFs defined at discrete points (source points) that are in this way able to interpolate an n dimensional scattered data. The interpolation ensures a smooth morphing of the computational CFD mesh and at the same time, an exactly prescribed displacement of the source points.

After defining a set of source points, whether on the surface of the geometry of interest or in the volume space surrounding it, with their displacement, the solution of the RBF mathematical problem is sought by the computation of the scalar parameters of a linear system of the order equal to the number of considered source points. Once the RBF system coefficients have been computed, the displacement of an arbitrary node of the mesh, either inside or outside of the domain, can be expressed as a sum of the radial contribution of each source point. In such a way, a desired modification of the mesh nodes position can be rapidly applied preserving mesh topology.

RBFs can be classified on the basis of the type of the support (global or compact) they have, meaning the domain where the chosen RBF is non zero-valued. The behaviour and the quality of the interpolation therefore depend on the chosen RBF.

An interpolation function composed of an RBF φ and a polynomial H of order $h - 1$ (h is the order of φ), introduced with the aim to guarantee the compatibility with the rigid motion, is defined as

$$I(\boldsymbol{r}) = \sum_{i=1}^{S} \gamma_i \, \varphi\left(\left\|\boldsymbol{r} - \boldsymbol{r}_{k_i}\right\|\right) + H(\boldsymbol{r}), \tag{9}$$

where $\boldsymbol{r}$ is a vector identifying the position of a generic node belonging to the surface or volume mesh, $\boldsymbol{r}_{k_i}$ is the i-th source point position vector (S is their total number) and $\|\cdot\|$ is the Euclidean norm, namely the distance between two points. The minimal degree of polynomial H depends on the choice of the RBF. A unique interpolant exists if the RBF is conditionally positive definite function [18]. In that case a linear polynomial

$$H(\boldsymbol{r}) = \beta_1 + \beta_2 x + \beta_3 y + \beta_4 z \tag{10}$$

can be used with RBFs of order $h \leq 2$ [19]. A radial basis fit exists if the coefficients γ_i of the RBF and the weights of the linear polynomial vector β_i can be found such that the desired values of displacement g_i are obtained at source points and the polynomial terms give no contributions at source points, i.e.

$$I\left(\boldsymbol{r}_{k_i}\right) = g_i, 1 \leq i \leq S; \; \sum_{i=1}^{S} \gamma_i q\left(\boldsymbol{r}_{\boldsymbol{k}_i}\right) = 0 \tag{11}$$

for all polynomials q with a degree less or equal to that of polynomial H. Coefficients γ_i and β_i can be obtained by solving the system

$$\begin{pmatrix} \boldsymbol{R} & \boldsymbol{P} \\ \boldsymbol{P}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \gamma \\ \beta \end{pmatrix} = \begin{pmatrix} \boldsymbol{g} \\ \boldsymbol{0} \end{pmatrix}, \tag{12}$$

where $\boldsymbol{R}$ is the interpolation matrix defined by calculating all the radial interactions between source points

$$R_{ij} = \varphi\left(\left\|\boldsymbol{r}_{k_i} - \boldsymbol{r}_{k_j}\right\|\right), \qquad 1 \le i \le S, 1 \le j \le S \tag{13}$$

and $\boldsymbol{P}$ is a constraint matrix that arises to balance the polynomial contribution and contains a column of "1" and the x, y, z positions of the source points in the other three columns

$$\boldsymbol{P} = \begin{pmatrix} 1 & x_{k_1} & y_{k_1} & z_{k_1} \\ 1 & x_{k_2} & y_{k_2} & z_{k_2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{k_N} & y_{k_N} & z_{k_N} \end{pmatrix}. \tag{14}$$

RBF interpolation works for scalar fields, hence a system of the form of Eq. (12) has to be solved for each of the three spatial directions.

Using RBF method for interpolation purposes has several advantages that makes it very attractive for mesh morphing. Since it is a meshless method only grid points are moved regardless of the type of the volume cells they belong to. This makes the method suitable for parallel implementation that can potentially handle meshes with large number of cells [20]. In fact, once the solution is known and shared in the memory, each computer core has the ability to morph the nodes on its own mesh partition without knowing what is happening outside. The reason for that lies in a global range of the interpolation function where the continuity at interfaces is implicitly guaranteed. The method is, in spite of its meshless nature, able to exactly prescribe desired deformations of the geometry of interest. This is achieved by using all the surface mesh nodes as RBF source points with prescribed displacements, including the simple zero displacement condition at the surface that should be left by the morphing tool as undeformed.

4 FSI STRATEGY

FSI strategy proposed in the present paper is composed out of three sequential stages: vibration modes calculation, RBF solution generation using RBF Morph™ tool and at the end an iterative FSI cycle employing CFD solver and mesh morphing through RBF solutions.

At the first stage, structural mode shapes and associated natural frequencies of deformable parts, that are utilized in the mode superposition phase, are calculated by means of a FEM solver. For the present paper, Ansys® APDL solver was used in the validation test case (Section 5), whereas Abaqus™ FEA software was adopted in the propeller optimization test case (Section 6). The number of physically most prominent mode shapes that are used in the subsequent stages has to be chosen at this point. It should be mentioned here that the FEM mesh doesn't have to be conformal to the CFD mesh (used at the third stage) at the geometry surface. For a typical industrial problem, CFD simulation needs a few times finer mesh at the surface (in order to simulate or model boundary layer effects) compared to the mesh used at the FEM calculation. FEM mesh nodes therefore in general don't coincide with CFD nodes. Because of this reason an interpolation step between both meshes is introduced at the third stage that leads to as universal FSI strategy as possible.

RBF solution for each mode shape is calculated at the second stage. This is done by the means of the deformable surface displacement distribution of the corresponding mode shape obtained in previous stage. Parts of the geometry that are considered as rigid are kept constrained. In order to obtain an adequate level of CFD mesh quality after each FSI cycle morphing iteration, a special attention should be given to each RBF solution. For this reason a proper source point's density and morphing domain dimensions should be used.

The third stage of the proposed strategy employs an FSI cycle composed out of CFD simulation, modal coordinates calculation and mesh morphing steps. FSI cycle is basically an iterative process which deforms the baseline geometry in such a way that the internal tension forces of the structure counteracts the aerodynamic loads on the geometry surface. Since it is an iterative process the cycle is finished when the resultant force on the surface converges below a certain threshold, defined by the user.

A surface distribution of aerodynamic forces $\boldsymbol{f}$ that act on the geometry of interest, as a consequence of the airflow passing by, is calculated via a CFD simulation. This aerodynamic load, comprised out of two contributions (pressure force and viscous force), deforms the geometry. In order to apply the deformation in a modal space, a modal force for each mode m can be calculated using Eq. (7) as

$$N_m = \sum_{i=1}^{n_{surf}} u_{m_i}^T f_i. \tag{15}$$

Each modal force is therefore a scalar obtained by integrating the external load field over the entire structure (all surface nodes) weighted by the m-th mode eigenvector. According to Eq. (8) each modal coordinate can then be calculated as

$$\eta_m = \frac{N_m}{\omega_m^2}. \tag{16}$$

The last step of the FSI cycle represents the morphing of the mesh and the geometry. A linear superposition of all chosen mode shapes (RBF solutions) weighted by the modal coordinates is applied to the CFD mesh in order to deform it. This crucial step is in the modal space represented by the following relation

$$\boldsymbol{r}_{CFD} = \boldsymbol{r}_{CFD_0} + \sum_{m=1}^{n_{modes}} \eta_m u_m, \tag{17}$$

where $\boldsymbol{r}_{CFD_0}$ and $\boldsymbol{r}_{CFD}$ represent the position of CFD mesh nodes before and after the deformation, respectively. As a consequence of this very same geometry deformation, the aerodynamic loads are altered at the same time. Next FSI cycle iteration therefore requires another CFD simulation on altered geometry that calculates new force distribution and the FSI cycle can continue upon convergence. The proposed FSI approach bypasses several complexities related to the implementation of the alternative two-way FSI procedures. Its advantage is that after the first stage, no further iteration with the structural solver is required, which drastically reduces the time of the analysis.

5 HIRENASD TEST CASE

A model employed to accomplish a validation of the present FSI approach is one of the configurations supplied at the 1st Aeroelastic Prediction Workshop (AePW) [21] launched by NASA. In particular, the wind tunnel model of this HIgh REynolds Number Aero-Structural

Dynamics (HIRENASD) configuration [22] tested in the Cologne European Transonic Wind tunnel (Figure 1), consists of a tapered 34° aft-swept wing with a BAC3-11/RES/30/21 supercritical airfoil profile and a wing span of 1.286m. The numerical and experimental investigations of the test case have been performed under $Mach = 0.8$, $Re = 7 \cdot 10^6$ (based on the 0.345m reference wing root chord), air density $\rho = 1.22 kg/m^3$ and angle of attack (AoA) = 1.5° flow conditions (ETW132 data point [22]).



Figure 1: HIRENASD wind tunnel model.

5.1 Mode shapes and RBF solutions

Mode shapes and natural frequencies have been calculated using the ANSYS® APDL solver from the FEM model, consisted of the wing, the excitation system and the balance, in NASTRAN format made available by the AeWP committee [21]. For the FSI analysis the first six vibration modes were chosen; their shapes, together with the corresponding natural frequencies, are depicted in Fig. 2, starting with Mode 1 at the top left and ending with Mode 6 at the bottom right. According to the AePW classification, label B is assigned to the out-of-plane bending and label Fa to the in-plane fore-and-aft bending.

In the second stage, an RBF solution has been generated for each of the six modes through the RBF Morph™ tool. Since in both the wind tunnel and the FEM model the fuselage aerodynamic fairing is mechanically uncoupled from the wing root, a slight deformation of the fairing is allowed. This is achieved by removing fixed source points from the portion of the fuselage near the wing root (Fig. 3 left) leaving the rest of the fuselage undeformable. Additional box-shaped encapsulation domain, presented in Fig. 3 right, is utilized to limit the morphing action to the CFD mesh volume close to the HIRENASD geometry.

5.2 FSI cycle

FSI analysis was divided in two subsequent simulation phases that ran on a mixed 1.5 million cell SOLAR unstructured grid made available by the AePW committee [21]. Both phases employed a steady CFD simulation using a compressible flow SU2 solver with a low-Reynolds Spalart-Allmaras turbulence model. A fully developed airflow passing a stiff geometry (baseline) was obtained in the first phase, whereas the elastic behaviour of the geometry was taken into account in the second phase through the FSI cycle described in the Section 4. Since the baseline flow field is not far from the final deformed one, the proposed two-phase approach accelerated the FSI cycle convergence through smaller number of CFD iterations in the second phase. Aerodynamic load distribution, integrated to obtain all modal forces

through Eq. (15), was composed only out of pressure force contribution. Shear stresses were neglected.

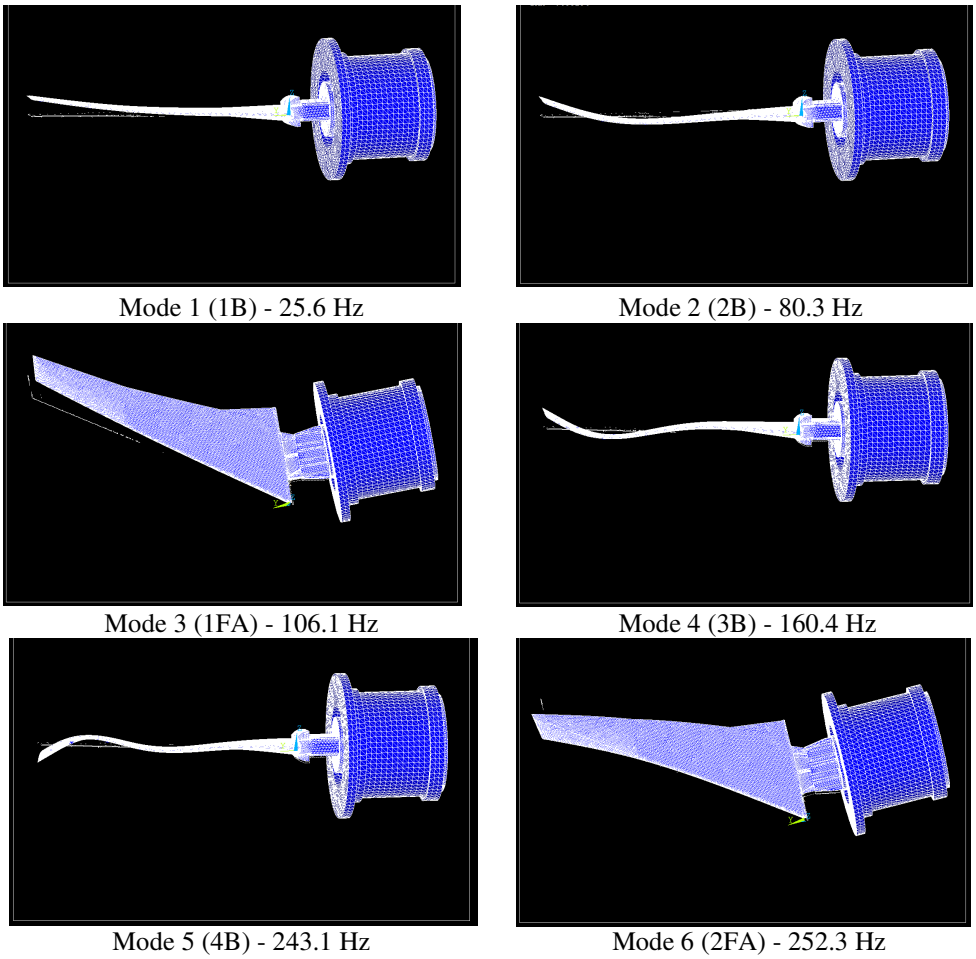


Figure 2: First six mode shapes and natural frequencies of the HIRENASD model.

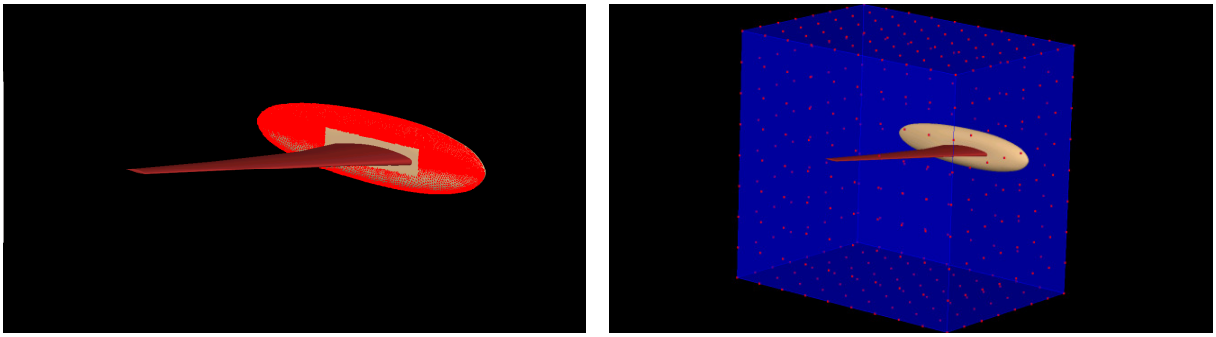


Figure 3: Fixed (left) and encap domain (right) source points.

The surface pressure distribution for the baseline configuration is visualised in Fig. 4 left, with the position of the shock wave quickly noticeable along the wing span. The y^+ surface distribution with values under 1, can be seen in Fig. 4 right.

The validation of the proposed FSI strategy was performed by comparing three different calculation outputs with respect to either numerical or experimental data, which can be found in [23]. The first output was the vertical displacement of the wing tip of the fully converged wing (deformed) with respect to the baseline. A comparison of the baseline geometry (grey

colour at the bottom) and the deformed one (green colour on top) can be seen in Fig. 5 left. A zoom-in on the wing tip is added in order to easily compare its position at different FSI cycle iterations. As the wing's shape of the third and fourth cycle almost coincides, the convergence of the static aeroelastic solution can be judged as reached. The convergence of the vertical displacement value at the trailing edge wing tip point during the FSI study is depicted in Figure 5 right. The converged value is 14.60 mm, which is very close to the numerical results found in [23] for two different grid setups (13.74 mm for Grid A and 14.02 mm for Grid B).

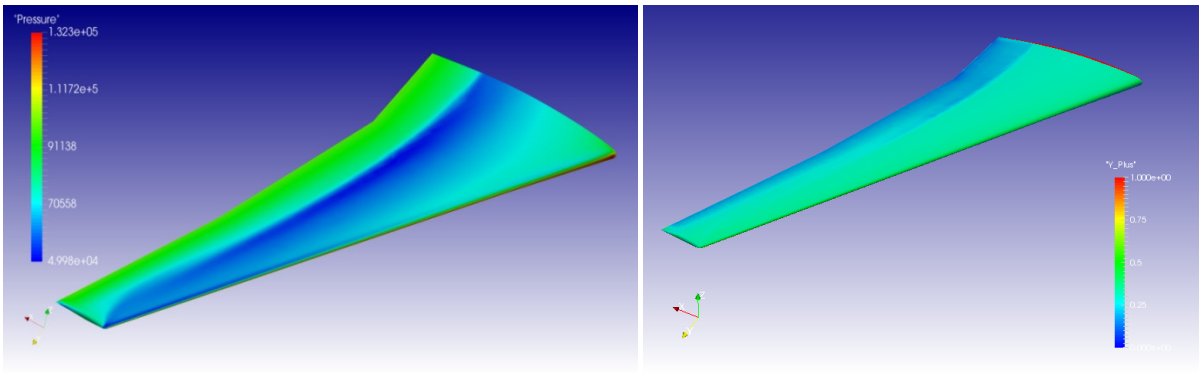


Figure 4: Steady state pressure (left) and y^+ (right) distribution on the HIRENASD baseline configuration.

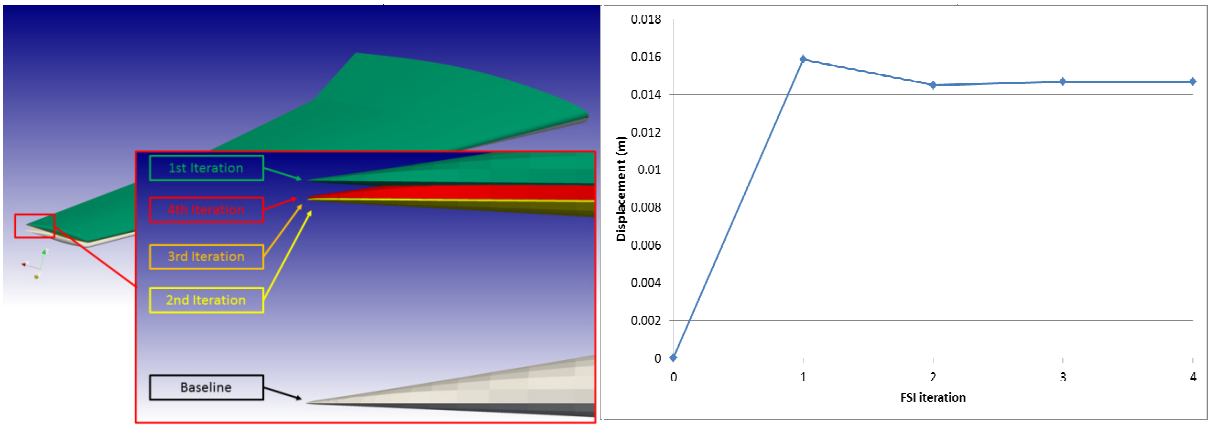
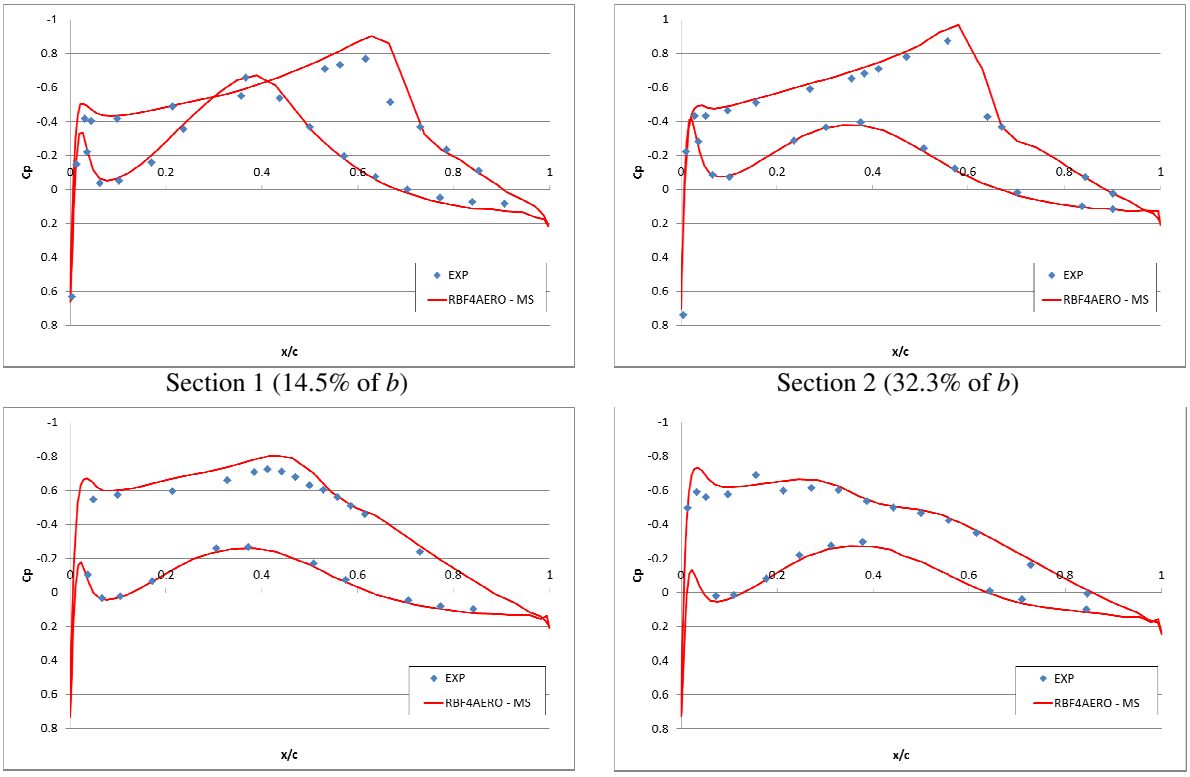


Figure 5: Comparison of wing baseline (grey) and its deformed geometry (green) together with a wing tip zoom-in detail (left). Wing tip vertical displacement at consecutive FSI cycle iterations (right).

The second FSI strategy validation was performed by comparing the pressure coefficient distribution with respect to the experimental data presented in [23] for a deformed geometry. The comparison at four different wing-span locations (in percentage of the wing span b) is depicted in Fig. 6. Even though some deviations of the numerical results from the experimental measurements can be observed on the pressure suction side the comparison can be still treated as satisfactory.

Furthermore, a good agreement between the proposed approach and the literature numerical results [23] of the lift and the drag aerodynamic coefficients for the baseline and the deformed geometry can be observed in Table 1. The results also show a strong influence of the deformation at both lift and drag coefficient. Based on the presented results of all three validation criteria the proposed FSI strategy validation can be considered as positively accomplished.



Section 1 (14.5% of *b*) Section 2 (32.3% of *b*)

Section 5 (65.5% of *b*) Section 7 (95.3% of *b*)

Figure 6: Comparison of numerical (RBF4AERO-MS) and experimental (EXP) pressure distribution at four different wing section locations.

Aerodynamic coefficient	Baseline			Deformed		
	Proposed approach	Grid A (NASA)	Grid B (NASA)	Proposed approach	Grid A (NASA)	Grid B (NASA)
C_l	0.3593	0.3542	0.3544	0.3407	0.3373	0.3366
C_d	0.0151	0.0173	0.0154	0.0144	0.0166	0.0147

Table 1: Aerodynamic coefficients comparison between the proposed approach and documented numerical results for two different grid set-ups [23] for the baseline and the deformed geometry.

6 PROPELLER OPTIMIZATION

In this case, spinning aircraft propeller deformations due to aerodynamic loads are investigated. As a consequence of the deformation propeller's aerodynamic characteristics change with respect to the ones calculated on the initial geometry. Since the power needed to spin the propeller and the thrust the propeller is capable of differ from the expected ones, the propeller designer must take into account such deformations. The objective of the present test case was to increase the efficiency of the chosen propeller taking into account its deformed shape by means of the proposed FSI strategy through design of experiments (DoE) analysis.

Test case employs a two-bladed fixed pitch propeller measuring 1.6m in diameter (Fig. 7 left). The propeller is a result of an in-house design and it is intended for a particular Pipistrel's electric aircraft - WATTsUP (Fig. 7 right). In order to simplify the analysis only a spinner, extended into a water drop like shape, was used substituting the complete aircraft. All CFD simulations including the calculations of the aerodynamic loads were done using finite

volume open source software OpenFOAM® where a multi reference frame (MRF) approach was used. In this way an absolute velocity field is calculated in a rotating frame of reference and the problem reduces to a steady-state calculation. With introduction of cyclic boundary conditions only a half of the propeller had to be simulated, which simplified the problem even further.

The mesh, prepared by the snappyHexMesh utility, consists of 1.6 million cells in a cylindrical volume domain (Fig. 8 left). MRF zone occupies only a small portion of the complete volume and with its cylindrical shape it tightly embraces the propeller and the spinner (Fig. 8 right). The propeller has been simulated and later optimized in a take-off and cruise flight regimes. During take-off the propeller spins at 2300RPM and the inflow velocity is 30m/s, which results in a tip Mach number of 0.57. During cruise on the other hand the propeller spins at 2550RPM. The inflow velocity is 51.4m/s which results in a tip Mach number of 0.65. Even though the Mach numbers are quite high the effects of compressibility were not taken into consideration in order to simplify the CFD simulation. Instead an incompressible RANS solver simpleFoam was used employing high-turbulence Spalart-Allmaras turbulence model. Densities of 1.19kg/m^3 and 1.11kg/m^3, and kinematic viscosities of $1.5 \cdot 10^{-5}$ m^2/s and $1.58 \cdot 10^{-5}$ m^2/s were used as airflow properties at the take-off and cruise simulations, respectively. y^+ and pressure surface distributions of a converged baseline simulation at take-off condition can be seen in Fig. 9. y^+ranges on average from a value of 10 to a satisfactory maximum of 130.

According to the CFD simulation the propeller baseline produces approximately 1070 N of thrust force and needs 53511 W of power in order to overcome the aerodynamic drag.

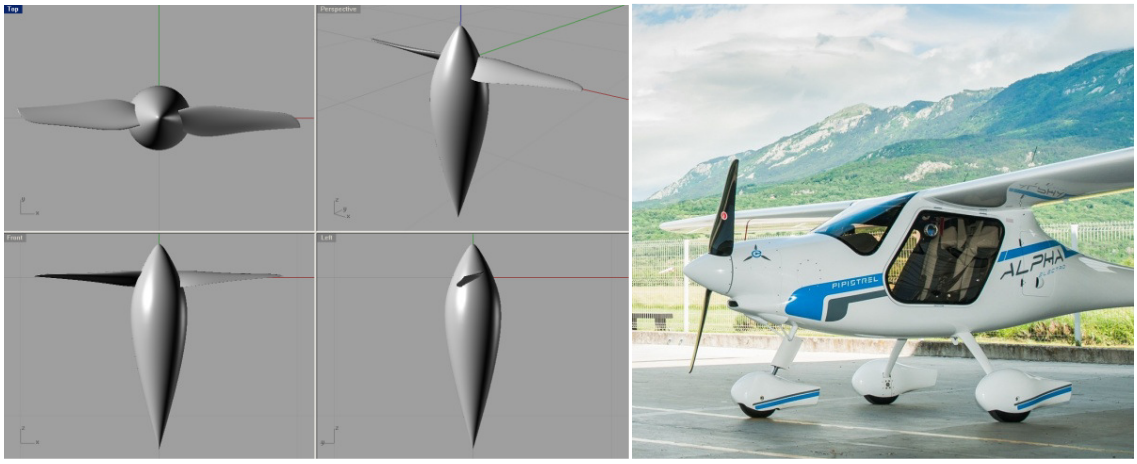


Figure 7: CAD model of the propeller geometry (left), physical propeller mounted on Pipistrel's electric aircraft – WATTsUP (right).

6.1 Mode shapes and RBF solutions

The propeller with its internal structure was analyzed with the Abaqus™ FEA software where its mode shapes and natural frequencies were calculated in a rotated frame of reference. A subset of first five physically most prominent modes has been chosen for the mode-superposition FSI approach. Emphasized mode shapes together with the corresponding frequencies are presented in Fig. 10, starting with Mode 1 at the top left and ending with Mode 5 at the bottom right. First four modes represent first four bending deformation modes, whereas the fifth mode represents the first order of twist deformation.

Individual eigenvectors are then used as inputs for the RBF Morph™ tool where RBF solutions for the CFD mesh morphing are computed. In order to reduce the morphing time, a morphing domain is set in a form of a cylinder around the propeller blade (Fig. 11). Mesh

morphing is therefore limited only to this domain volume thus leaving the major portion of the CFD mesh undeformed. Since the propeller is mechanically not connected to the spinner, a small portion of the spinner around the propeller root is allowed to modify, leaving most of the spinner undeformed.

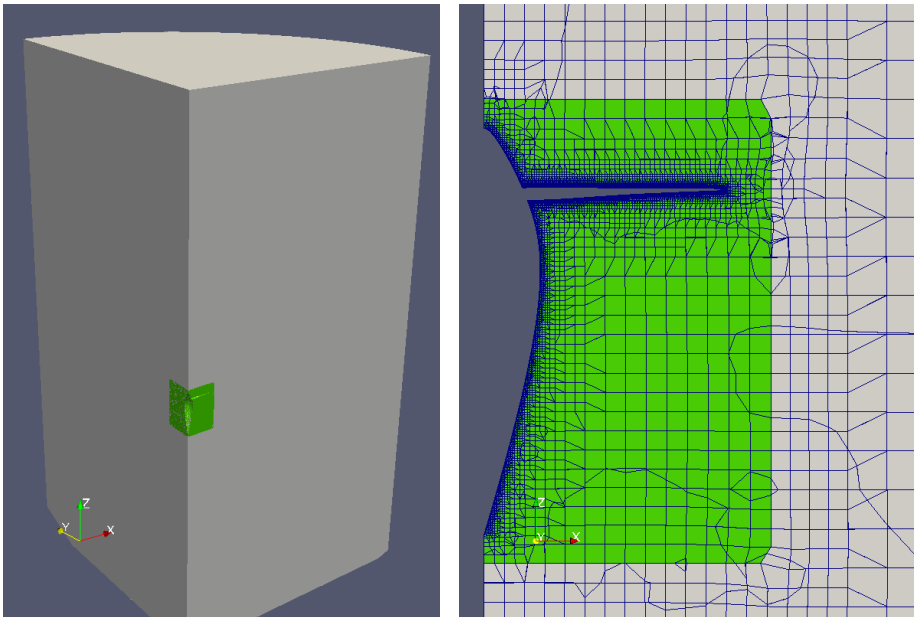


Figure 8: A half of the total CFD volume mesh (left) and a slice through a mesh (right). MRF zone is denoted in green colour.

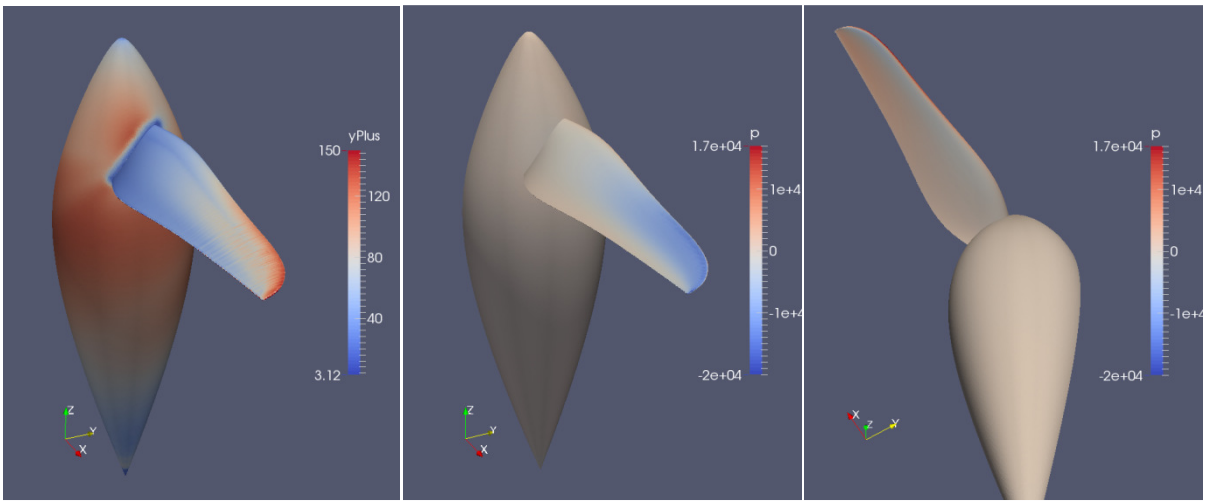


Figure 9: y^+ (left) and pressure surface distribution on the upper (middle) and bottom (right) side of a fully converged baseline simulation at take-off flight conditions.

6.2 FSI baseline results

First an FSI analysis was run on propeller baseline geometry at take-off conditions. An absolute value of a relative difference of each modal coordinate with respect to its converged value (fifth iteration) over first four FSI cycle iterations is depicted in Fig. 12 left. All modal coordinates vary for less than 0.2% already after the second iteration. Other representatives of fast convergence of proposed FSI strategy are the characteristic properties of the propeller, thrust and power, collected at each iteration (Fig. 12 right). Here, the left and the right ordi-

nate axes represent propeller thrust in Newtons and its power in Watts, respectively. It can be seen that both propeller characteristics practically don't change after the second iteration.

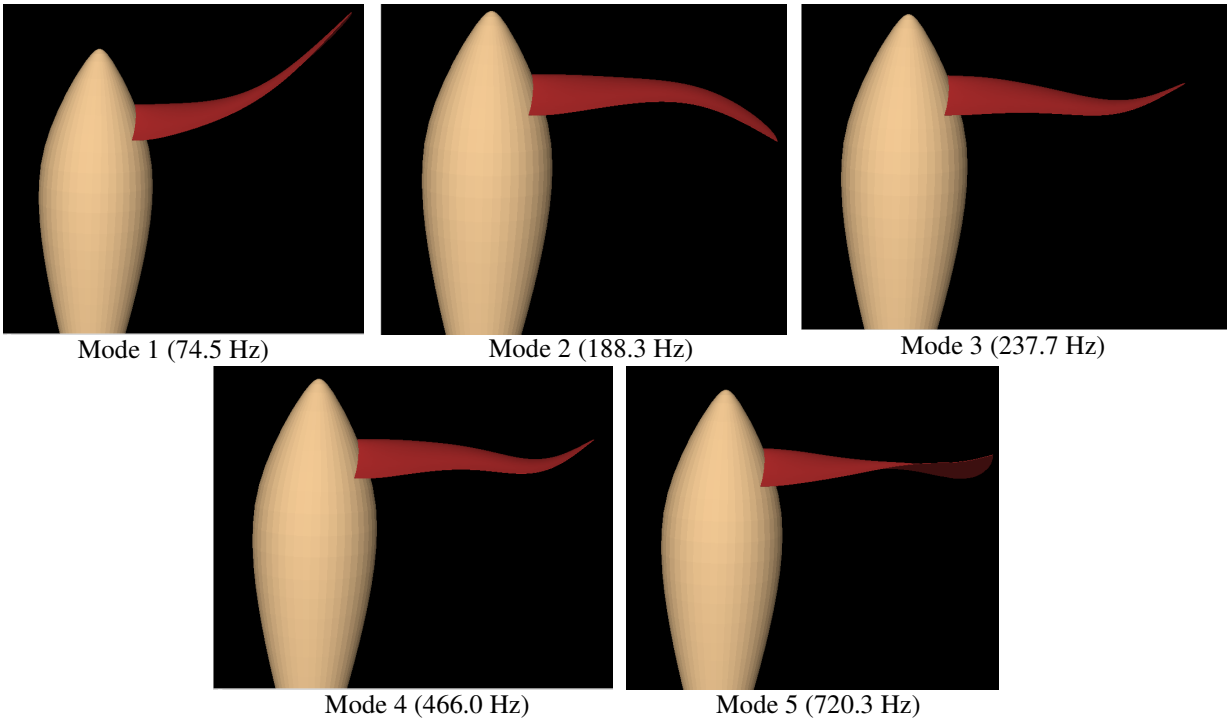


Figure 10: First five propeller blade mode shapes with corresponding natural frequencies. Amplitudes are emphasized for better presentation.

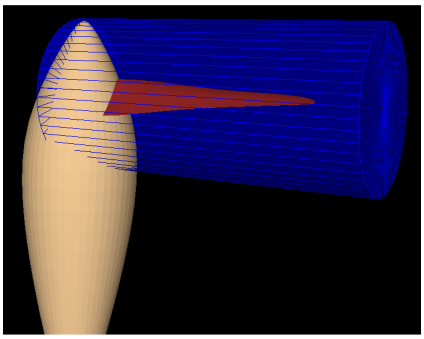


Figure 11: Morphing domain around the propeller blade depicted in blue colour.

Deformation of the propeller geometry during the FSI analysis can be observed in Fig. 13. Geometry comparison was done in software ParaView. The baseline geometry is depicted in grey colour, the first FSI iteration in green colour and the fifth FSI iteration (final and converged geometry) in red colour. The most obvious deformation is blade root bending being also the first and also the most dominant eigenmode. Such deformation results as a consequence of the large suction region (low pressure) on the outer portion of upper side of the blade seen in the middle Fig. 9. Slight blade tip twist can also be noticed in Fig. 13. The source of such deformation can be assigned to the high pressure region located on the outer and aft portion of lower side of the blade seen on the right hand side of Fig. 9.

Results of the FSI analysis on the baseline configuration at the take-off conditions and at cruise conditions are presented in numbers in Table 2 and Table 3, respectively. From both tables it can be seen that the thrust and the power do not change significantly in spite of a

quite perceivable geometry change between the baseline and the deformed geometry. Convergence of all modal weights, propeller thrust and power (Fig. 12) as well as the geometry deformation (Fig. 13) indicate a very rapid FSI procedure converge which drastically accelerated the optimization presented in the following section.

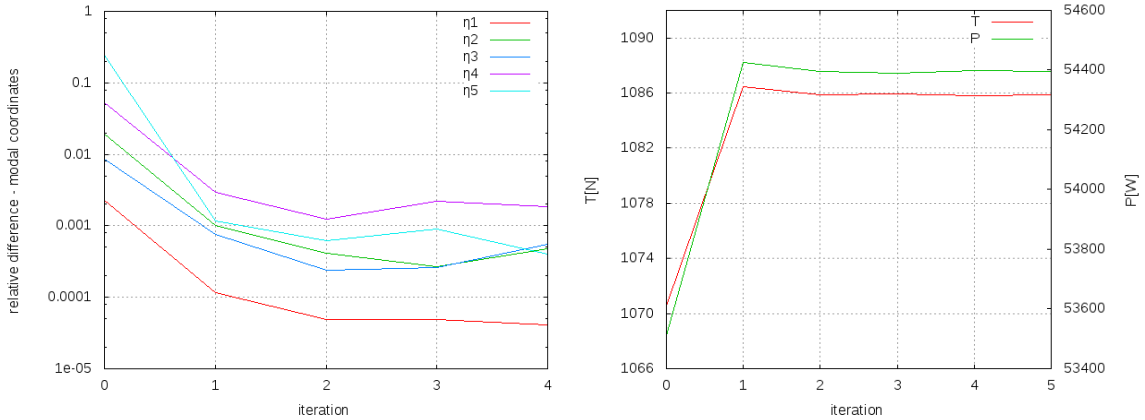


Figure 12: Convergence of a relative difference (absolute value) of each modal coordinate (with respect to its converged value) (left) and propeller thrust and power (right) during FSI analysis for the propeller baseline at take-off conditions.

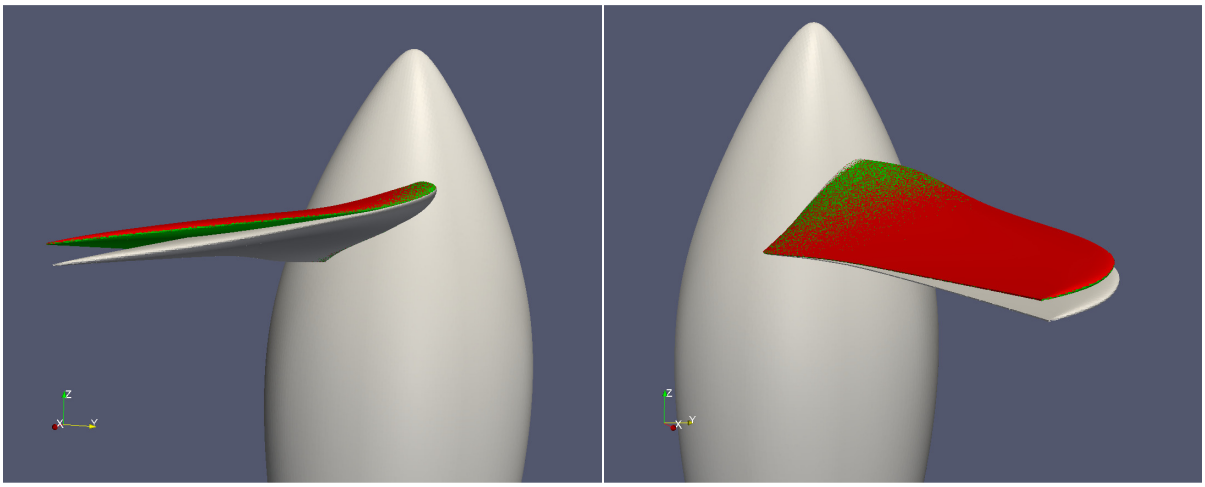


Figure 13: Comparison of propeller blade geometry at different FSI iterations at take-off flight conditions. Baseline (grey), the first iteration (green) and the fifth and final iteration (red).

Parameter	0. iteration/baseline	5. iteration	relative difference
Thrust force (N)	1070	1086	+1.5%
Power (W)	53511	54395	+1.6%

Table 2: Results of an FSI analysis on baseline configuration at take-off conditions.

Parameter	0. iteration/baseline	5. iteration	relative difference
Thrust force (N)	223	220	-1.4%
Power (W)	28941	28820	-0.4%

Table 3: Results of an FSI analysis on baseline configuration at cruise conditions.

6.3 Propeller optimization

The main objective of the present test case, propeller optimization, was met through DoE analysis employing proposed FSI strategy. Optimization variables, chosen to find the optimal propeller geometry in the vicinity of the baseline shape, were propeller pitch p and twist angle t. Both deformations were applied using RBF Morph™ tool. As a result, no additional meshing was needed in order to run CFD simulations. Propeller was optimized with respect to two flight regimes (take-off and cruise conditions) for as large propeller efficiency as possible. The optimization objective function used was a weighted sum of propeller efficiencies at cruise and take-off flight condition

$$F(t,p) = \frac{1}{2}\nu_{CRUISE} + \frac{1}{2}\nu_{TAKE-OFF}, \tag{18}$$

with the propeller efficiency defined as

$$\nu = \frac{TV}{P}, \tag{19}$$

where V denotes the velocity of the aircraft (cruise or take-off). As it can be seen from Eq. (18), both contributions were equally weighted. By varying optimization variables, propeller pitch and twist angle, a maximal combined efficiency and therefore an optimal propeller geometry was sought. A positive pitch angle change is defined as an increase of the angle of attack of the airflow incidenting on the propeller blade (Fig. 14 left). Positive twist angle on the other hand reduces the local pitch angle of attack (Fig. 14 middle).

Pitch and twist variations were introduced to the geometry via morphing procedure by applying additional artificial morphing RBF solutions at the morphing step of the FSI cycle. Since the fifth propeller mode shape represents a twist-like deformation by itself, the said mode was used to apply the additional twist to the propeller shape. On the other hand, the pitch variation had to be applied by calculating additional RBF solution. For the latter, a new moving domain in a shape of a cylinder and a bigger morphing domain were set (Fig. 14 right). All mesh cells, located inside the moving domain (red), rotated around the blade spanwise axis for a prescribed pitch angle. The cells embraced in the morphing domain (blue), but located outside of the moving domain, were morphed in such a way that a smooth transition from rotated to stationary cells (outside of the morphing domain) was made. At this point an assumption has been made that the change of mode shapes and natural frequencies caused by the deformation (of small extend) of the geometry (pitch and twist angle) is negligible.

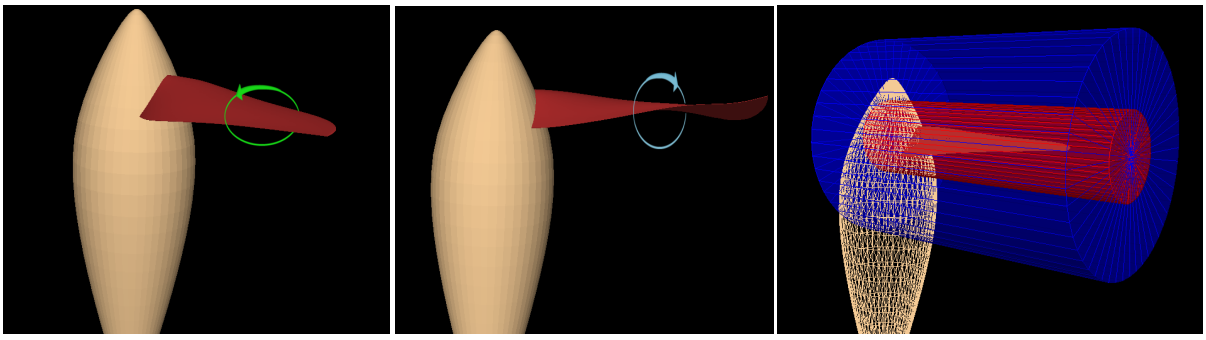


Figure 14: The pitch (left) and twist (middle) positive direction deformation indicated by an arrow. The morphing domain (blue cylinder) and the moving domain (red cylinder) (right).

At each of two flight conditions a separate DoE table was filled with 8 times 9 different combinations of pitch and twist deformations around the baseline configuration. Each point in

the DoE table is calculated using a proposed FSI cycle strategy, combining CFD simulations and mesh morphing. The whole procedure was performed automatically using bash script. The script chose a combination of the optimization parameters, started FSI cycle and stored the results for each set of parameters. The results of both DoE procedures are presented in Fig. 15. Each blue dot represents a single DoE point, where the efficiency is calculated from the thrust and power of the deformed propeller geometry.

DoE results were in a postprocessing phase interpolated using a third degree polynomial surface (coloured surfaces in Fig 15). A summation of both surfaces according to Eq. (18) formed a surrogate model that was later used in order to perform a constrained optimization. Constraints used were take-off and cruise engine power i.e. 55 kW and 41.3 kW, respectively. The output of the optimization was a pitch and twist pair (p, t), that produced the maximal objective function i.e. the maximal weighted sum of both efficiencies. Surrogate model is together with the baseline configuration (green point) and optimal configuration (blue point) presented in Fig. 16.

Pitch and twist angles in the optimal point are 1.045° and 1.482°, respectively. The weighted sum of both efficiencies increased by 4.0%. Taking into account the fact that only two global degrees of freedom were used in order to optimize the propeller and that the baseline geometry already presents an optimized design (done in the past by other Pipistrel in-house software), such an increase of efficiency can be considered as a satisfactory result.

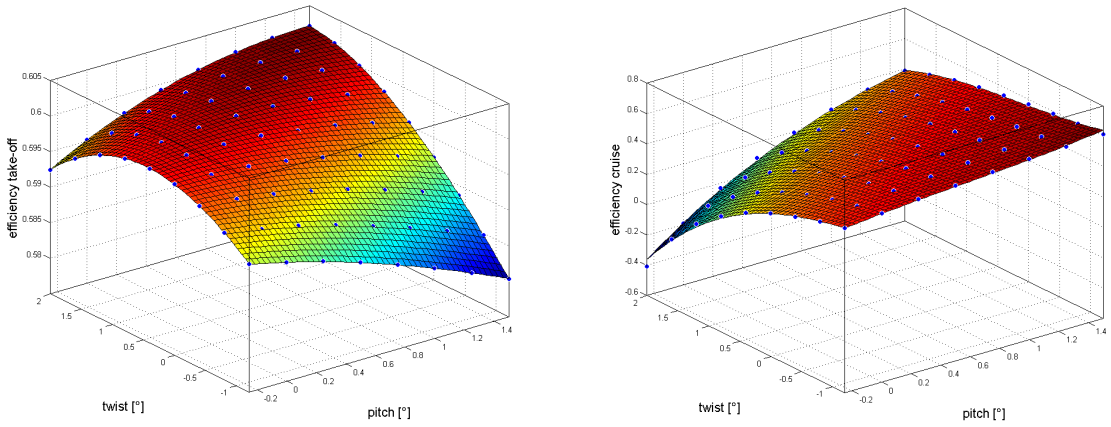


Figure 15: DoE results at 72 different pitch and twist pairs for take-off (left) and cruise (right) flight conditions.

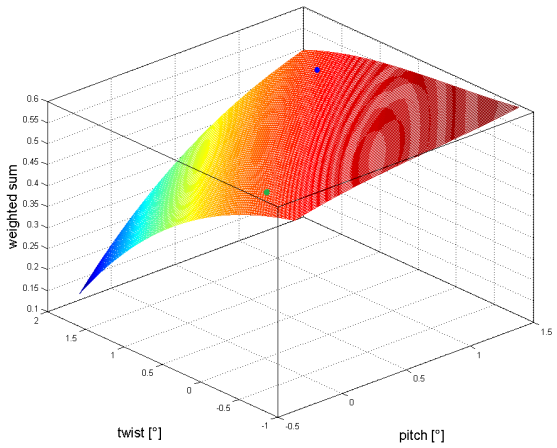


Figure 16: Surrogate model with the baseline configuration denoted with a green point and optimal configuration with blue point.

A comparison of the deformed optimized (red) and the deformed baseline geometry (grey) can be seen in Fig. 17. From the figure it can be seen how did the local angle of attack at the blade inner section increase by the positive pitch angle yet decrease at the blade outer section by the positive twist angle. Thrust force and power of the optimized propeller are for the baseline (0th iteration) and deformed configuration (5^{th} iteration), together with their relative difference, presented in Table 4 for the take-off flight condition and in Table 5 for the cruise condition. The order of magnitude of relative differences didn't change from the baseline analysis (Tables 1 and 2).

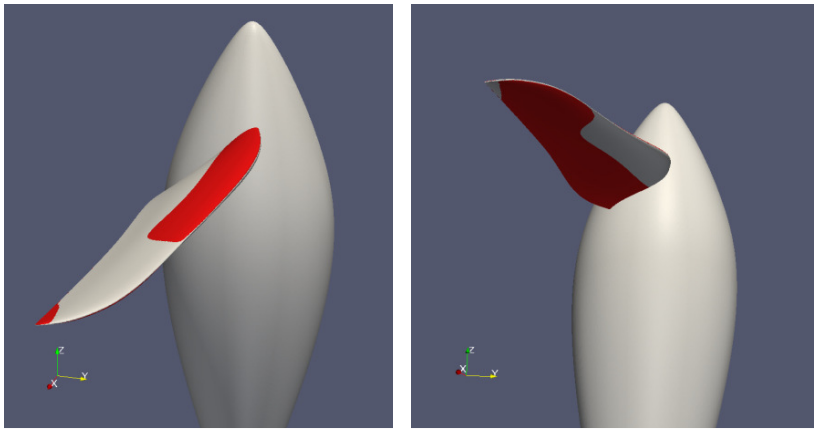


Figure 17: Geometry comparison of deformed optimized propeller blade (red) and deformed baseline configuration (grey).

Parameter	0. iteration/baseline	5. iteration	relative difference
Thrust force (N)	1089	1102	+1.2%
Power (W)	54140	55004	+1.6%

Table 4: Main results of the FSI analysis at optimal pitch and twist at take-off conditions.

Parameter	0. iteration/baseline	5. iteration	relative difference
Thrust force (N)	251.2	247.4	-1.5%
Power (W)	30049	29901	-0.5%

Table 5: Main results of the FSI analysis at optimal pitch and twist at cruise conditions.

7 CONCLUSIONS

A mesh morphing based FSI strategy was proposed that, connecting the modal analysis of the structure of interest, transformation of calculated mode shapes directly to the CFD environment using radial basis functions, and iterative calculation of aerodynamic loads and CFD mesh morphing, presents an efficient and reliable analysis of engineering problems.

FSI strategy was first validated by analyzing a well-tested wind tunnel case in transonic flow conditions. The results obtained have been favourably compared to those measured during the referenced experimental campaign. The strategy was then used in a real world engineering application as the main component of the overall propeller optimization problem, where a surrogate method based on the DoE method was used in order to increase the weighted sum of efficiencies at two different flight regimes by 4.0%.

The main advantage of the proposed technique with respect to other standard approaches is that the modal parameterization has to be built only once. Embedding the structural modes of deformable parts directly into the computational model enables an immediate application of deformation to the CFD grid during the calculation stage.

Finally, the proposed FSI approach represents a universal procedure that can be exploited regardless of the CFD and FEM solver choice, it can be extended to very large models where RBF Morph™ has already proven its computational time efficiency and it is applicable to both steady and unsteady aeroelastic studies.

ACKNOWLEDGMENT

This work was developed in the framework of the RBF4AERO "Innovative benchmark technology for aircraft engineering design and efficient design phase optimisation" project funded in the Aeronautics and Air Transport (AAT) research thematic area of the EUs 7th Framework Programme (FP7- AAT, 2007-2013) under Grant Agreement no. 605396.

REFERENCES

[1] H.J. Bungartz, M. Schäfer, *Fluid-structure interaction: Modeling, simulation, optimization*, Springer-Verlag, 2006.

[2] H. Gene, W. Jin, L. Anita, Numerical methods for fluid-structure interaction – A review, *Communications in Computational Physics*, 12(2):337-377, 2012.

[3] Y. Bazilevs, K. Takizawa, T.E. Tezduyar, *Computational fluid-structure interaction: Methods and applications*, Wiley, 2013.

[4] S. Keye, Fluid-structure coupled analysis of a transport aircraft and comparison to flight data. In *AIAA Paper 2009-4198*, *39th AIAA Fluid Dynamics Conference*, San Antonio, Texas, USA, June 22-25, 2009.

[5] M.E. Biancolini, C. Groth, E. Costa, F. Lagasco, A mesh morphing based technique to efficiently perform FSI analyses for aeroelastic design applications, *AIDAA Italian Association of Aeronautics and Astronautics XXII Conference*, Napoli, Italy, September 9-12, 2013.

[6] S, Jakobsson, O. Amoignon, Mesh deformation using radial basis function for gradient based aerodynamic shape optimization, *Computers and Fluids*, 36(6):1119-1136, 2007.

[7] A. de Boer, M.S. van der Schoot, H. Bijl, Mesh deformation based on radial basis function interpolation, *Computers and Structures*, 85(11-14):784-795, 2007.

[8] A.H. van Zuijlen, A. de Boer, H. Bijl, Higher-order time integration rhrough smooth mesh deformation for 3D fluid-structure interaction simulations, *Journal of Computational Physics*, 224:414-430, 2007.

[9] M.E. Biancolini, U. Cella, An advanced RBF Morph application: coupled CFD-CSM aeroelastic analysis of a full aircraft model and comparison to experimental data, *8th MIRA International Vehicle Aerodynamics Conference,* Grove, UK, October 13-14, 2010.

[10] M.E. Biancolini, C. Groth, An efficient approach to simulating ice accretion on 2D and 3D airfoils, *Advanced Aero Concepts, Design and Operations Conferenc,* Bristol, UK, July 22, 2014.

[11] E.M. Papoutsis-Kiachagias, M. Andrejašič, S. Porziani, C. Groth, D. Eržen, M.E. Biancolini, E. Costa, K.C. Giannakoglou, Combining an rbf-based morpher with continuous adjoint for low-speed aeronautical optimization applications. In *ECCOMAS Congress 2016, VII European Congress on Computational Methods in Applied Sciences and Engineering*, Crete, Greece, June 5-10, 2016.

[12] D.H. Kapsoulis, V.G. Asouti, K.C. Giannakoglou, E. Costa, S. Porziani, C. Groth, U. Cella, M.E. Biancolini, Evolutionary aerodynamic shape optimization through the RBF4AERO platform. In *ECCOMAS Congress 2016, VII European Congress on Computational Methods in Applied Sciences and Engineering*, Crete, Greece, June 5-10, 2016.

[13] M.E. Biancolini, Fluid structure interaction (FSI) with RBF Morph: a generic Formula 1 front end, *EnginSoft International Conference 2011 - CAE Technologies for Industry,* Verona, Italy, October 20-21, 2011.

[14] M.E. Biancolini, I.M. Viola, M. Riotte, Sails trim optimisation using CFD and RBF mesh morphing, *Computers and Fluids,* 93:46-60, April 10, 2014. doi:10.1016/j.compfluid.2014.01.007

[15] S.G. Kelly, *Mechanical vibrations: theory and applications*, Cengage Learning, 2012.

[16] R.D. Cook, D.S. Malkus, M.E. Plesha, R.J. Witt, *Concepts and applications of finite element analysis*, *4th Edition*, John Wiley and Sons, 2002.

[17] L. Meirovitch, *Fundamentals of vibrations, Reissue edition*, Waveland Press, Inc., 2010.

[18] C. Micchelli, Interpolation of scattered data: Distance matrices and conditionally positive definite functions, *Constructive Approximation*, 2(1):11–22, 1986.

[19] A. Beckert, H. Wendland, Multivariate interpolation for fluid-structure interaction problems using radial basis functions, *Aerospace Science and Technology*, 5(2):125–134, 2011.

[20] U. Cella, M.E. Biancolini, Aeroelastic analysis of aircraft wind-tunnel model coupling structural and fluid dynamic codes, *Journal of Aircraft*, 49(2), 2012.

[21] AIAA Aeroelastic Prediction Workshop at https://c3.nasa.gov/dashlink/static/media/other/AEPW_legacy.htm, (25.2.2016)

[22] J. Ballmann, A. Boucke, L. Reimer, C. Dickopp, Results of dynamic experiments in the HIRENASD project and analysis of observed unsteady processes, In *IFASD Paper 2009-103*, *International Forum on Aeroelasticity and Structural Dynamics*, Seattle, USA, 2009.

[23] P. Chwalowski, J.P. Florance, J. Heeg, C.D.Wieseman, B.P. Perry, Preliminary computational analysis of the HIRENASD configuration in preparation for the aeroelastic prediction workshop, In *IFASD Paper 2011-108*, *International Forum on Aeroelasticity and Structural Dynamics*, Paris, France, 2011

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A PATH-CONSERVATIVE OSHER-TYPE SCHEME FOR AXIALLY SYMMETRIC COMPRESSIBLE FLOWS IN FLEXIBLE TUBES

Julia Leibinger[1], Michael Dumbser[2] and Uwe Iben[1]

[1]Robert Bosch GmbH
Robert-Bosch-Campus 1, 71272 Renningen, Germany
e-mail: julia.leibinger@de.bosch.com

[2] Department of Civil, Environmental and Mechanical Engineering
University of Trento, Via Mesiano 77, 38123 Trento, Italy
e-mail: michael.dumbser@unitn.it

Keywords: compressible flows in compliant tubes, flexible ducts with variable cross section, fluid-structure interaction (FSI), visco-elastic wall behavior, non-conservative hyperbolic systems, path-conservative finite volume schemes.

Abstract. *In industrial applications there exist numerous hydraulic devices, as for example the fuel injection system in modern combustion engines, or hydraulic braking systems. These devices become more and more complex, and also the requirements concerning their efficiency are continuously increasing. Therefore, it is necessary to understand these systems well in order to optimize them already at the design stage. For that purpose, it is important to have a numerical simulation tool available that is able to calculate a solution for a given configuration quickly and accurately. In hydraulic devices, the different components, like pumps, valves or vessels, are usually connected via flexible tubes. Thus, the relevant physical effects of such compliant tubes must be considered in the numerical simulation, especially at high frequencies. In typical industrial applications, pressures can reach several thousand bar, hence the flow has to be usually considered as compressible, and at the same time also the visco-elastic behavior of the tube wall needs to be properly accounted for. Then, a suitable numerical method needs to be developed in order to solve the resulting partial differential equations (PDE) of this coupled fluid-structure interaction (FSI) problem accurately and efficiently.*
In our talk, the equations of an axially symmetric flow of a compressible barotropic fluid are considered for flows through flexible elastic and visco-elastic tubes, so that the dominant physical effects can be reproduced according to the composition of the tube wall. To calculate the numerical solution of the derived PDE, a path-conservative Osher-type solver - the so-called Dumbser-Osher-Toro (DOT)- is applied. This solver can deal with hyperbolic conservation laws that may also contain non-conservative products. To validate the proposed mathematical model and the numerical scheme, the numerical solutions are compared with analytical results derived in the frequency domain, as well as with experimental measurements. In both cases, the proposed path-conservative finite volume scheme based on the DOT Riemann solver produces very good results.

1 Introduction

Many industrial applications involve complex hydraulic systems. Their behavior needs to be understood already while designing and constructing the different components. Therefore, we need accurate mathematical models for the simulation of the dynamic behavior of the individual components of such hydraulic systems, which can be, for example, the fuel injection system into the combustion engine or the braking system in a car. Very often, flexible tubes are an important part of these complex systems, in order to connect the different components with each other. Hence, they have a very important influence on the overall behavior of the entire system. That is why in this paper we introduce a new mathematical model and provide a novel numerical solver in order to simulate the dynamic behavior of flexible tubes. The mathematical model represents the combined effects of the flow of a compressible barotropic fluid inside a flexible tube of circular cross section. Thus, the model accounts for the fluid-structure interaction (FSI) problem between a compressible fluid and the elastic behavior of the surrounding structure of the tube wall.
For the fluid we use the one-dimensional cross-sectionally averaged compressible Euler equations in a duct with variable circular cross section. For the material of the tube wall there exist various models [14]. For example, the Hooke's law can be applied to reproduce the elastic behavior of the tube in a rather simple way (see [11]). Another model is the so called Maxwell model. This model is built form springs and dash pots and can therefore reproduce the visco-elastic behavior of polymer material ([13]). Finally, we derive a system of non-conservative hyperbolic partial differential equations.
For solving the non conservative PDE system we apply the successful family of path-conservative schemes developed by Castro & Parés and co-workers [3, 20, 1, 12, 2, 18, 4, 5, 6], which are based on the theory of Dal Maso, Le Floch and Murat [17] on weak solutions for hyperbolic PDE with non-conservative products. In particular, we use an extension of the Riemann solver of Osher and Solomon [19] that was recently proposed by Dumbser and Toro in order to solve *general* conservative and non-conservative hyperbolic PDE systems [10, 9] (DOT). We combine these two solvers and apply them to our PDE system that describes the coupled fluid-structure interaction problem.
To verify the solver, some one-dimensional Riemann problems are solved, and the numerical solution is compared with a quasi-exact solution provided in [7], where also the fluid-structure-interaction is taken into account. To show that we are able to reproduce the visco-elastic behavior of real tubes with the Maxwell model, we compare it to an analytical solution derived in the frequency domain and to available experimental data [16].

2 The governing PDE system of the coupled FSI problem

To simulate the dynamic behavior of compressible fluids in flexible tubes we need to model the unsteady flow of a barotopic fluid through a tube with variable cross sectional area, as well as the elasticity of the tube wall. This results in a system of PDEs with a conservative and a non-conservative part.

2.1 The cross-sectionally averaged Euler equations for the fluid

The axially symmetric flow through a tube with variable cross-sectional area is described by the one dimensional cross sectionally averaged compressible Euler equations

$$\frac{\partial}{\partial t}(A\rho) + \frac{\partial}{\partial x}(A\rho u) = 0, \tag{1}$$

$$\frac{\partial}{\partial t}(A\rho u) + \frac{\partial}{\partial x}(A\rho u^2 + Ap) - p\frac{\partial A}{\partial x} = F_R. \tag{2}$$

In these equations $A(x,t)$ represents the variable cross sectional area, ρ the cross sectionally averaged density inside the tube, p the cross sectionally averaged pressure, u is the average velocity of the fluid and the term F_R accounts for the friction between fluid and wall (see for example [21]).
To characterize the dependency between pressure and density we assume a barotropic fluid. In this model, the pressure $p(\rho)$ only depends on the density. This so-called equation of state (EOS) also considers a two phase flow caused by cavitation, which appears, when the pressure drops below the vapor pressure p_v:

$$\rho(p) = \begin{cases} \rho_0 + \frac{1}{c_0^2}(p - p_v), & \text{if } p \geq p_v, \\ \frac{1}{\frac{\mu(p)}{\rho_v(p)} + \frac{1-\mu(p)}{\rho_0}}, & \text{if } 0 < p < p_v, \end{cases}, \tag{3}$$

where ρ_0 is the density at standard conditions and μ is the mass fraction of vapor $\mu(p) = -K(p - p_v)$, with a cavitaion constant K, the speed of sound c_0 of the fluid and the vapor density $\rho_v(p)$ calculated by the ideal gas law.

So far we only described the fluid, i.e. we still need an equation for the cross sectional area $A(x,t)$. In the next chapter we describe two models for the tube wall.

2.2 Material model for the visco-elastic tube wall

In this chapter we briefly describe two material models: first, a rather simple one, namely the Laplace law, where the cross sectional area only depends on the pressure and second, a more intricate one, the Maxwell model, where the time dependent visco-elastic behavior of the tube material is also taken into account.

2.2.1 The Laplace law

With the Laplace law, a linear Hooke-like behavior is described. In the so derived model the cross sectional area is only affected by the current local pressure inside the tube,

$$A = A_0 + \frac{1}{\beta}(p - p_{ext}), \tag{4}$$

with the equilibrium cross sectional area A_0, the elasticity constant β and the external pressure p_{ext}.

2.2.2 The Maxwell model

The previously described Laplace law has only a limited validity for real materials. For polymer tubes we also need to consider the time dependent behavior of the material. Therefore a Maxwell model composed of springs and dash pots is applied. In Figure 1 a schematic representation of such a model is shown.

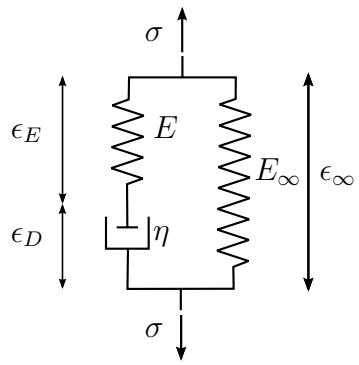


Figure 1: Schematic representation of a 3-parameter Maxwell model.

From this combination of springs and dash pots the following PDE results

$$\frac{\partial \phi}{\partial t} = -\frac{1}{\frac{E_0 \phi}{2Wc^2} + \rho} \frac{\partial (\phi \rho u)}{\partial x} + \frac{(2W(p - p_{ext}) - (\phi - 1)E_\infty)\,\phi}{\tau_1 (2Wc^2 \rho + E_0 \phi)}, \tag{5}$$

with $\phi = \frac{A}{A_0}$ the normalized cross sectional area, the elasticity constants E_∞ and $E_0 = E_\infty + E$, the relaxation parameter $\tau_1 = \frac{E}{\eta}$, the fluid density ρ, the fluid velocity u, the fluid pressure p and the speed of sound c of the fluid. W is a geometry factor that represents the wall thickness and is defined as follows:

$$W = 2\frac{\left(\frac{r_a}{r_i}\right)(1+\nu) + 1 - 2\nu}{\left(\frac{r_a}{r_i}\right)^2 - 1}, \tag{6}$$

with the inner and outer radius r_i and r_a and the Poisson ratio ν. For a more detailed description of the material models, see [13, 15]

3 The DOT Riemann solver

The equations described in the previous chapter need to be solved. We derive a system of PDEs with a conservative and non-conservative part. Such equations read as follows:

$$\frac{\partial \mathbf{Q}}{\partial t} + \frac{\partial}{\partial x}\mathbf{f}(\mathbf{Q}) + \mathbf{B}(\mathbf{Q})\frac{\partial \mathbf{Q}}{\partial x} = \mathbf{S}(\mathbf{Q}). \tag{7}$$

Here, $\mathbf{Q}(x,t) \in \mathrm{R}^n$ is the state vector, $\mathbf{f}(\mathbf{Q})$ is a nonlinear flux that contains the conservative part and $\mathbf{B}(\mathbf{Q})$ is an $n \times n$ matrix that includes the purely non-conservative part of the system. $\mathbf{S}(\mathbf{Q})$ is a nonlinear algebraic source term which results from the friction model F_R and the Maxwell model.

To solve the equations we unify two solvers in order to get a new solver that can handle both, a conservative an non-conservative part. The basic ideas of both solvers is the Riemann solver of Osher an Solomon and its extension to *general* hyperbolic systems described in [10, 9]. To get the combined solver the equation (7) needs to be reformulated as follows

$$\frac{\partial \mathbf{Q}}{\partial t} + \mathbf{A}(\mathbf{Q})\frac{\partial \mathbf{Q}}{\partial x} = \mathbf{S}(\mathbf{Q}), \tag{8}$$

with the new system matrix $\mathbf{A}(\mathbf{Q}) = \partial \mathbf{f}/\partial \mathbf{Q} + \mathbf{B}(\mathbf{Q})$. Now the finite volume scheme is written like

$$\mathbf{Q}_i^{n+1} = \mathbf{Q}_i^n - \frac{\Delta t}{\Delta x}\left(\mathbf{f}_{i+\frac{1}{2}} - \mathbf{f}_{i-\frac{1}{2}}\right) - \frac{\Delta t}{\Delta x}\left(\mathbf{D}_{i+\frac{1}{2}} + \mathbf{D}_{i-\frac{1}{2}}\right) - \Delta t \mathbf{B}(\mathbf{Q}_i^{n+\frac{1}{2}})\frac{\Delta \mathbf{Q}_i^n}{\Delta x} + \Delta t \mathbf{S}(\mathbf{Q}_i^{n+\frac{1}{2}}). \tag{9}$$

In this equation the numerical flux is defined by

$$\mathbf{f}_{i+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{f}(\mathbf{Q}^+_{i+\frac{1}{2}}) + \mathbf{f}(\mathbf{Q}^-_{i+\frac{1}{2}})\right) - \frac{1}{2}\int_0^1 \left|\mathbf{A}\left(\Psi(\mathbf{Q}^-_{i+\frac{1}{2}}, \mathbf{Q}^+_{i+\frac{1}{2}}, s)\right)\right| \frac{\partial \Psi}{\partial s} ds \tag{10}$$

and the so called jump terms by

$$\mathbf{D}_{i+\frac{1}{2}} = \frac{1}{2}\int_0^1 \mathbf{B}\left(\Psi(\mathbf{Q}^-_{i+\frac{1}{2}}, \mathbf{Q}^+_{i+\frac{1}{2}}, s)\right) \frac{\partial \Psi}{\partial s} ds. \tag{11}$$

The path Ψ is connecting the left state $\mathbf{Q}^-_{i+\frac{1}{2}}$ with the right state $\mathbf{Q}^+_{i+\frac{1}{2}}$. In our calculation we use a simple linear path and the path integrals are computed *numerically* via a simple Gauss-Legendre quadrature rule, so we get the final formulations

$$\mathbf{f}_{i+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{f}(\mathbf{Q}^+_{i+\frac{1}{2}}) + \mathbf{f}(\mathbf{Q}^-_{i+\frac{1}{2}})\right) - \frac{1}{2}\left(\sum_{j=1}^{N_G} \omega_j \left|\mathbf{A}\left(\Psi(\mathbf{Q}^-_{i+\frac{1}{2}}, \mathbf{Q}^+_{i+\frac{1}{2}}, s_j)\right)\right|\right)\left(\mathbf{Q}^+_{i+\frac{1}{2}} - \mathbf{Q}^-_{i+\frac{1}{2}}\right) \tag{12}$$

and

$$\mathbf{D}_{i+\frac{1}{2}} = \frac{1}{2}\left(\sum_{j=1}^{N_G} \omega_j \mathbf{B}\left(\Psi(\mathbf{Q}^-_{i+\frac{1}{2}}, \mathbf{Q}^+_{i+\frac{1}{2}}, s_j)\right)\right)\left(\mathbf{Q}^+_{i+\frac{1}{2}} - \mathbf{Q}^-_{i+\frac{1}{2}}\right). \tag{13}$$

For a more detailed description of the solver see [15].

4 Numerical results

We verify the solver by comparing it to a quasi-exact solution ([7, 8]) by solving the Riemann problems defined in Table 1. An exact solution can only be obtained for the Laplace law, since the Maxwell model is time dependent, which introduces a source term on the right hand side of the PDE system. To validate the physical behavior of the Maxwell model, we compare it to analytical and experimental results in the frequency domain.

4.1 Riemann problems

We present the solution of two Riemann problems. The Riemann problem one remains in the fluid domain whereas the second Riemann problem results in two phase flow. The results for the calculations with the Laplace law and the comparison with the quasi-exact solution are shown in Figs. 2 and 3 for the time $t_{end} = 4 \cdot 10^{-4}$. The agreement between both solutions is good.

Additionally, Figs. 4 and 5 illustrate the solutions of the Maxwell model for the time $t_{end} = 8 \cdot 10^{-4}$ for the same initial condition. The parameters for the Maxwell model are $E_0 = 10^9$ and $E_\infty = 0.9 \cdot 10^9$ ant $\eta = 2 \cdot 10^4$. The time dependency of the model can be clearly seen, when

we compare the results to the one obtained with the Laplace law, where the wave profiles are sharper.

case	p_L	u_L	p_R	u_R	β	number of cells
RP1	$8 \cdot 10^5$	0	$2 \cdot 10^5$	0	$4.38 \cdot 10^{13}$	200
RP2	$2 \cdot 10^4$	0	$2 \cdot 10^2$	0	$4.38 \cdot 10^{13}$	500

Table 1: Initial states for Riemann problems (p[Pa], u[m/s])

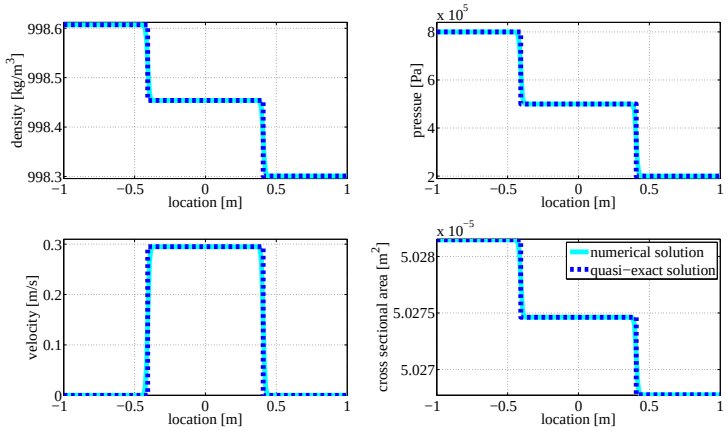


Figure 2: Riemann problem for the Maxwell model with in liquid domain and comparison to quasi-exact solution.

4.2 Comparison with experimental results

To show that the Maxwell model can actually reproduce the behavior of real tube materials, the numerical results are compared to experimental ones and to analytical solutions obtained in the frequency domain. In the experimental setup, the test object was excited with frequencies between 0 and 2000 Hz, then volumetric flow rate and the pressure were measured at the in- and outlet. The exact method is described in [16]. In Fig. 6 one can see the results. The green dotted lines are the experimental results, the dashed lines are the analytical reference solution obtained in the frequency domain and the solid lines are the solutions obtained from the numerical simulations obtained with the previously described numerical method and mathematical model. The calculations are done with and without friction model. For frequencies up to 800 Hz one can see a good agreement. Beyond 800 Hz, the agreement with experiments is worse. This could have two possible reasons: first, the quality of the experimental results seems to deteriorate, which is visible from the unphysical jumps in the experimental results beyond 800 Hz. Second, a higher order Maxwell model, which means a Maxwell model with more springs and dash pots in parallel could improve the accuracy of the numerical simulation.

5 Conclusion

We presented a nonconservative system of nonlinear hyperbolic PDEs in one space dimension for the simulation of compressible fluid flow in flexible tubes. We showed that we can reproduce experimental results with the Maxwell model. The numerical solver introduced for solving the governing PDE system (DOT) was also validated with a quasi-exact solution.

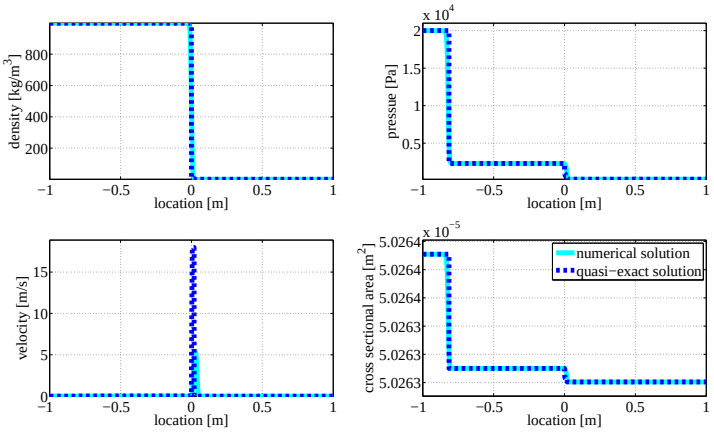


Figure 3: Riemann problem for the Laplace law with two phase flow and comparison to quasi-exact solution.

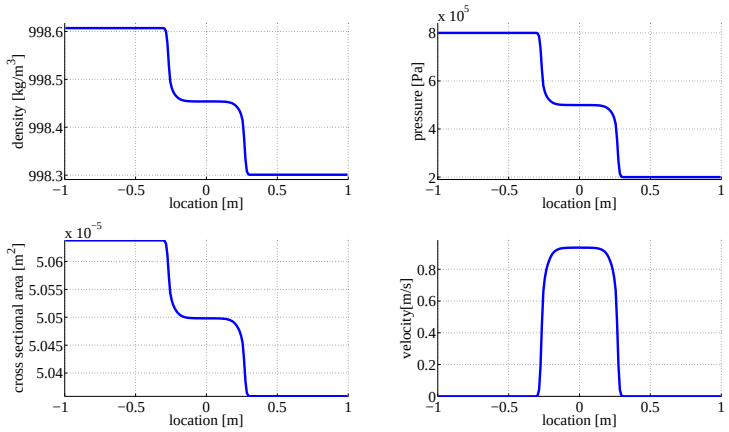


Figure 4: Riemann problem for the Maxwell model with in liquid domain.

Future work will also consider the modelling of more complex tubes, like multi-layered hydraulic tubes. These tubes have shown a different dynamic behavior in experiments and this behavior could not be well reproduced with the equations explained in this paper. Therefore the equations need to be modified and the numerical solver needs to be adapted to solve these new equations.

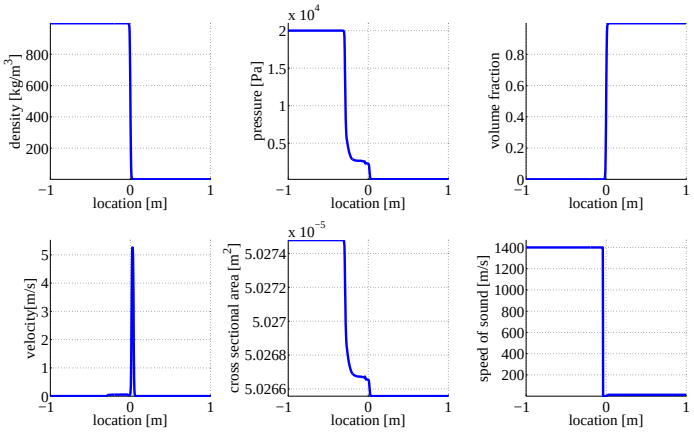


Figure 5: Riemann problem for the Maxwell model with two phase flow.

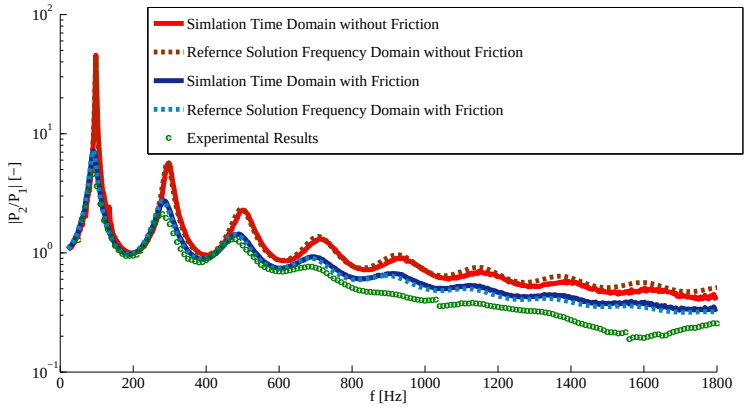


Figure 6: Comparison with experimental results in the frequency domain.

REFERENCES

[1] M. CASTRO, E. FERNÁNDEZ, A. FERRIERO, J. GARCÍA, AND C. PARÉS, *High order extensions of Roe schemes for two dimensional nonconservative hyperbolic systems*, Journal of Scientific Computing, 39 (2009), pp. 67–114.

[2] M. CASTRO, J. GALLARDO, J. LÓPEZ, AND C. PARÉS, *Well-balanced high order extensions of Godunov's method for semilinear balance laws*, SIAM Journal of Numerical Analysis, 46 (2008), pp. 1012–1039.

[3] M. CASTRO, J. GALLARDO, AND C. PARÉS, *High-order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. applications to shallow-water systems*, Mathematics of Computation, 75 (2006), pp. 1103–1134.

[4] M. CASTRO, A. PARDO, C. PARÉS, AND E. TORO, *On some fast well-balanced first order solvers for nonconservative systems*, Mathematics of Computation, 79 (2010), pp. 1427–1472.

[5] M. DUMBSER, M. CASTRO, C. PARÉS, AND E. TORO, *Ader schemes on unstructured meshes for nonconservative hyperbolic systems: Applications to geophysical flows*, Computers & Fluids, 38 (2009), pp. 1731 – 1748.

[6] M. DUMBSER, A. HIDALGO, M. CASTRO, C. PARÉS, AND E. TORO, *Force schemes on unstructured meshes ii: Non-conservative hyperbolic systems*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 625 – 647.

[7] M. DUMBSER, U. IBEN, AND M. IORIATTI, *An efficient semi-implicit finite volume method for axially symmetric compressible flows in compliant tubes*, Applied Numerical Mathematics, 89 (2015), pp. 24–44.

[8] M. DUMBSER, U. IBEN, AND C. MUNZ, *Efficient implementation of high order unstructured WENO schemes for cavitating flows*, Computers & Fluids, 86 (2013), pp. 141 – 168.

[9] M. DUMBSER AND E. F. TORO, *On universal Osher–type schemes for general nonlinear hyperbolic conservation laws*, Communications in Computational Physics, 10 (2011), pp. 635–671.

[10] ——, *A simple extension of the Osher Riemann solver to non-conservative hyperbolic systems*, Journal of Scientific Computing, 48 (2011), pp. 70–88.

[11] R. ETLENDER, *Modellierung und Simulation der Wellenausbreitung in flexiblen hydraulischen Leitungen*, PhD thesis, Universität Stuttgart, 2012.

[12] J. GALLARDO, C. PARÉS, AND M. CASTRO, *On a well-balanced high-order finite volume scheme for shallow water equations with topography and dry areas*, Journal of Computational Physics, 227 (2007), pp. 574–601.

[13] M. GOEKE, *Modellierung der Materialdmpfung bei der Wellenausbreitung in flexiblen Leitungen*, master's thesis, University Stuttgart, 2013.

[14] R. LAKES, *Viscoelastic Solids*, CRC Press Inc, 1998.

[15] J. LEIBINGER, M. DUMBSER, U. IBEN, AND I. WAYAND, *A path-conservative osher-type scheme for axially symmetric compressible flows in flexible visco-elastic tubes*, Applied Numerical Mathematics, (2016).

[16] R. LEONHARDT, *Dynamische Untersuchungen von Hydraulikkomponenten*, PhD thesis, Universitt Karlsruhe, 2008.

[17] G. D. MASO, P. LEFLOCH, AND F. MURAT, *Definition and weak stability of nonconservative products*, J. Math. Pures Appl., 74 (1995), pp. 483–548.

[18] M. MUÑOZ AND C. PARÉS, *Godunov method for nonconservative hyperbolic systems*, Mathematical Modelling and Numerical Analysis, 41 (2007), pp. 169–185.

[19] S. OSHER AND F. SOLOMON, *Upwind difference schemes for hyperbolic systems of conservation laws*, Math. Comp., 38 (1982), pp. 339–374.

[20] C. PARS, *Numerical methods for nonconservative hyperbolic systems: a theoretical framework.*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 300–321.

[21] W. ZIELKE, *Frequency-dependent friction in transient pipe flow*, Journal of Basic Engineering, 90 (1968), pp. 109–115.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
T. Lohner, K. Michaelis, K. Stahl (eds.)
Crete Island, Greece, 5–10 June 2016

LIMITING SHEAR STRESS FORMULATION FOR TEHL SIMULATION

T. Lohner[1], K. Michaelis[1], and K. Stahl[1]

[1] Gear Research Centre (FZG), Technical University of Munich (TUM), Boltzmannstraße 15, 85748 Garching, Germany
e-mail: lohner@fzg.mw.tum.de

Keywords: Elastohydrodynamic lubrication, limiting shear stress, friction, TEHL simulation

Abstract. *Flow curves of typical lubricants show a reduction of lubricant viscosity with increasing shear rate (non-Newtonian fluid behavior). Moreover, measurements of flow curves in pressurized viscometers as well as measurements of friction curves of highly loaded lubricated contacts reveal a limiting value for the shear stress that the lubricant can transmit. This is referred to as the limiting shear stress, at which point visualization of shear bands suggests that internal slip occurs in the lubricant film. It has been shown that the limiting shear stress is not a constant lubricant property, but varies with pressure and temperature. Furthermore, a dependency on the entrainment velocity of the lubricant has been suggested in the literature.*

Many TEHL (thermal elastohydrodynamic lubrication) simulation models have used the viscous Eyring model to describe the non-Newtonian fluid behavior. As the model is not based on a limiting shear stress, the Eyring shear stress and other lubricant property dependencies (e.g. pressure-viscosity) were adjusted to recalculate the measured friction data. However, since this approach appears to be lacking in physical background, visco-plastic models including a limiting shear stress have been introduced. For this, knowledge of the dependencies of the limiting shear stress is required.

This study suggests an expression for the limiting shear stress of lubricants derived from measurements of friction curves at the twin disk test rig. The contact-integral nature of the measurements is adjusted for local use in TEHL simulation models. The limiting shear stress formulation is applied to a finite element (FE) based TEHL simulation model and tested. The non-Newtonian fluid model considered is the simplified Bair/Winer model.

TEHL simulation results achieve very high levels of agreement between the measured and simulated friction curves for a large variety of operating conditions and different lubricants including mineral and synthetic oils. If the sliding velocity is increased further once the maximal friction is reached, it is also possible to quantify the portion of friction reduction that is due to decreasing lubricant viscosity and the portion that is due to the temperature dependency of the limiting shear stress.

1 INTRODUCTION

Newton's law of shear stress for fluid flow describes a linear dependency of shear stress on shear rate with the dynamic viscosity as constant of proportionality. However, flow curves of typical lubricants measured under high pressure exhibit a reduction of lubricant viscosity with increasing shear rate and hence non-Newtonian fluid behavior. Bair [2] [3] measures flow curves in high-pressure rotational viscometers at constant temperature and shows, with increasing shear rate: a linear (Newtonian) regime, a non-linear (shear thinning) regime and a regime of shear-rate-independent shear stress (interpreted as limiting shear stress). A physical interpretation of the latter regime is given by Bair and McCabe [2] [4] observing shear bands with internal slip in fluid flow under high pressure. Thereby, wall slip was not observed. Glass transition is not necessarily accompanied by an occurrence of limiting shear stress (Bair and Winer [8]).

It has been shown that the limiting shear stress is not a constant lubricant property. It is often reported to be a linear function of pressure (Bair [2], Bair and Winer [7], Beilicke et al. [10], Habchi, Bair and Vergne [17]). Bair und Winer [8] show an additional dependency on temperature. Mayer [22] proposes that the limiting shear stress is also subject to variations of entrainment velocity of highly loaded contacts.

The limiting shear stress can either be determined from high-pressure viscometers (primary data measurement) or derived from friction curves measured at tribometers. The former can be limited or very elaborate, particularly at high pressures. For pressures of up to $1000 N/mm^2$, Evans and Johnson [13] show strong correlation between values for the limiting shear stress obtained from a viscometer and values obtained from measured friction curves at the twin disk test rig. The results of a recent research project [24] confirm this for pressures of $> 1000 N/mm^2$. Due to cross-influences between viscous heating and the contact-integral nature of friction curves, the correct interpretation is crucial when deriving limiting shear stresses (Bair [2], Evans and Johnson [13]).

In order to describe the non-Newtonian fluid behavior, functional relationships have been fitted to measured flow curves. A widely used model that does not take into account limiting shear stress, but which presents good numerical practicability (Bartel [9]) has been proposed by Eyring [14]. Bair et al. [5] show that the Eyring model cannot provide a good representation of the shear thinning behavior of lubricants. Furthermore, shear stresses are not limited and can become infinitely large. Models with limiting shear stress have been proposed, e.g. by Bair and Winer [6], Wolff and Kubo [28], and Beilicke et al. [10].

Towards a physically based calculation of highly loaded lubricated contacts, nowadays detailed TEHL simulation programs are used. These programs describe lubricant properties as functions of pressure and temperature as well as of shear rate in terms of viscosity. In this context, Bair et al. [5] notes that the simulation of friction has sometimes turned into a "curve fitting contest" with little consideration of primary measured data.

Wolff and Kubo [28] investigate the influence of different non-Newtonian fluid models by means of isothermal EHL simulations. They recommend the use of limiting-shear-stress-based models and the consideration of thermal effects. Habchi [16] also emphasizes the importance of taking into account thermal effects, based on TEHL simulations with a modified Carreau-Yasuda model [1]. As the Carreau-Yasuda model does not limit the shear stress, it has been truncated numerically based on a linear pressure dependency of the limiting shear stress. Based on the Eyring model, Bobach et al. [11] obtain very strong correlation between TEHL simulation results and measured friction curves from the FZG twin disk test rig. Based on the same TEHL simulation model and a combined Eyring and simplified Bair/Winer model, Beilicke et al. [10] again achieve very strong correlation between simulated and measured

friction curves. In these studies, the limiting shear stress is described as linearly dependent on pressure and derived from measurements at a twin disk test rig.

This paper suggests an expression for the limiting shear stress of lubricants derived from measurements of friction curves at the FZG twin disk test rig. One mineral oil (MIN100), two polyalphaolefine oils (PAO10 and PAO100) and one polyglycol oil (PG100) are considered. The contact-integral nature of the measurements is adjusted for local use in TEHL simulation models. The limiting shear stress formulation is applied to a finite element (FE) based TEHL simulation model as described by Lohner et al. [21] and tested by means of a non-Newtonian fluid model based on the simplified Bair/Winer model (Wolff and Kubo [28]). TEHL simulation results show very high levels of agreement between measured and simulated friction curves for a large variety of operating conditions and for the different types of lubricants considered. If the sliding velocity is increased further once the maximal friction is reached, it is also possible to quantify the portion of friction reduction that is due to decreasing lubricant viscosity and the portion that is due to the temperature dependency of the limiting shear stress.

2 LIMITING SHEAR STRESS FORMULATION

The proposed limiting shear stress formulation for implementation in TEHL simulation models has its origin in a contact-integral approach, which is further developed to a contact-local approach. The work in this section is supported by student coursework by Paschold [23], performed with guidance from the author.

The contact-integral approach $\tau_{\lim|exp}$ is derived from measurements of friction curves at the FZG twin disk test rig and consists of the sum of three terms as given by Eq. (1). The lubricant-specific parameters $E_{\dot{\gamma}1}$, $E_{\dot{\gamma}2}$, $E_{\dot{\gamma}3}$ und $E_{\dot{\gamma}4}$ are obtained from regression analysis. $\tau_{\lim|exp}$ is then transferred to a contact-local approach $\tau_{\lim|sim}$ acc. to Eq. (2).

$$\tau_{lim|exp}(p_m, v_\Sigma, T_{ref}) = E_{\dot{\gamma}1} \cdot \mathrm{p_m} + \left[E_{\dot{\gamma}2} + E_{\dot{\gamma}3} \cdot \ln\left(\mathrm{v_\Sigma} \cdot 1\frac{s}{m}\right)\right] + E_{\dot{\gamma}4} \cdot T_{ref} \tag{1}$$

$$\tau_{lim|sim}(p, v_\Sigma, T_{ref}, T) = \begin{cases} \tau_{lim,m} & f\ddot{u}r\ \tau_{lim|sim} \leq \tau_{lim,m} \\ \frac{4}{\pi} \cdot \tau_{\lim|exp}(p_m = p, v_\Sigma, T_{ref}) \cdot sinh\left(\frac{T_{ref}}{T}\right) & f\ddot{u}r\ \tau_{lim|sim} > \tau_{lim,m} \end{cases} \tag{2}$$

The workflow for deriving $\tau_{lim|sim}$ and the lubricant-specific parameters $E_{\dot{\gamma}1}$, $E_{\dot{\gamma}2}$, $E_{\dot{\gamma}3}$ and $E_{\dot{\gamma}4}$ is schematically shown in Figure 1. The starting point features measured friction curves, whose behavior resembles that of flow curves obtained from high pressure viscometers (see section 1). They typically present a linear (Newtonian) regime, a non-linear (shear thinning) regime and a regime of maximal or rather decreasing friction (see a and b in Figure 1). For mild operating conditions, no maximum of friction may be observed for the considered range of sliding velocity v_g (see c in Figure 1). Due to the increasing bulk (T_M) and TEHL contact temperature as v_g increases, the friction curves become dominated by thermal effects.

2.1 Contact-integral approach

A maximum of the measured friction curves is interpreted as some integral value $\tau_{lim|exp}$ for the limiting shear stress. Therefore, for all friction curves featuring a maximum, the values of p_m, v_Σ, T_M, T_{Oil} and $\tau_{lim|exp}$ are evaluated (see Figure 1 left), whereby $\tau_{lim|exp}$ is obtained from the ratio of friction force F_R and Hertzian contact area A_H. Mayer [22] measured friction curves for different lubricants in the fluid film lubrication regime over a large range of operat-

ing conditions ($p_H = \{600|1200\} N/mm^2$, $\vartheta_{Oil} = \{40|100\}°C$, $v_\Sigma = \{1 \dots 16\} m/s$, $v_g = \{0 \dots 5.33\} m/s$) at the FZG twin disk test rig with cylindrical disks with polished surfaces. Four selected lubricants from Mayer [22] - MIN100, PAO100, PAO10 and PG100 - are considered in this study. The measured kinematic viscosities ν at $40°C$ und $100°C$ are shown in Table 1. Further lubricant data is documented by Mayer [22] and ITR Clausthal [19].

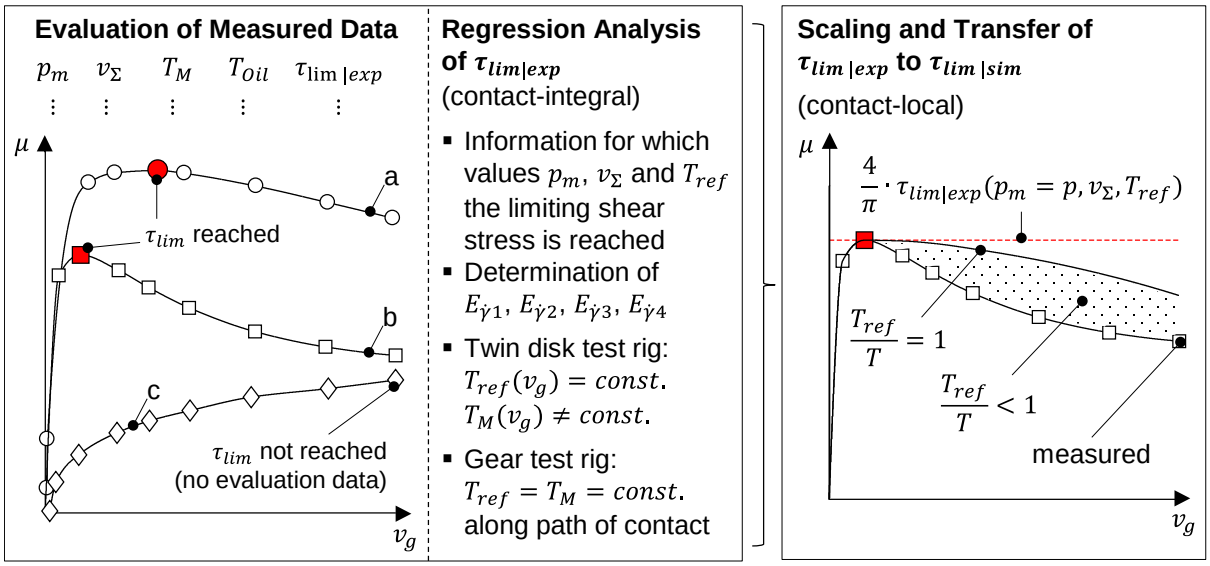


Figure 1: Workflow for determination of contact-integral limiting shear stress $\boldsymbol{\tau_{lim|exp}}$ and its scaling and transfer to a contact-local approach $\boldsymbol{\tau_{lim|sim}}$

Kinematic viscosity	**MIN 100**	**PAO100**	**PAO10**	**PG100**
$\nu(40°C)\ in\ mm^2/s$	95.0	104.6	63.7	99.4
$\nu(100°C)\ in\ mm^2/s$	10.0	15.5	9.9	18.0

Table 1: Measured kinematic viscosities of the considered lubricants (Mayer [22])

Figure 2 exemplifies the measured friction curves for MIN100 and PG100, with two operating conditions (low and high load).

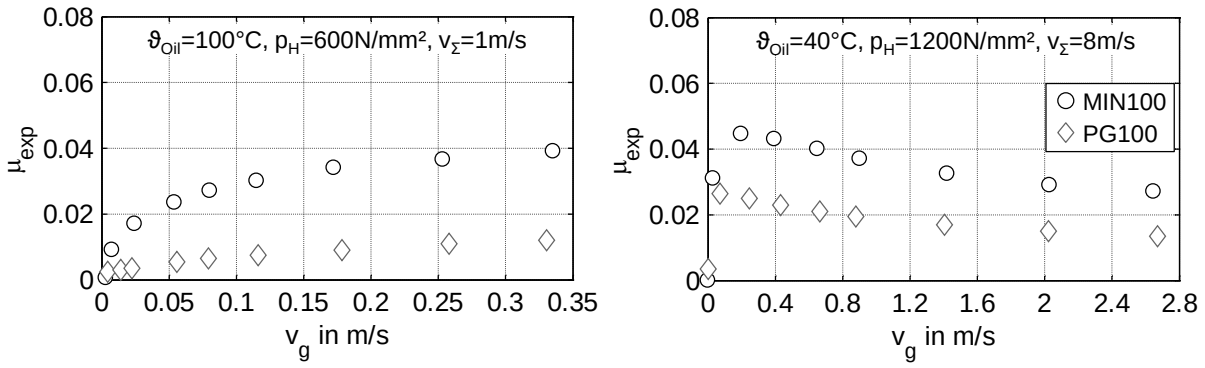


Figure 2: Measured friction curves at the FZG twin disk test rig (exemplarily, measured data acc. to Mayer [22])

Based on the measured friction curves of Mayer [22], Figure 3 shows the evaluated values of $\tau_{lim\,|exp}$ plotted over p_m, v_Σ und T_M for the four selected lubricants MIN100, PAO100, PAO10 and PG100.

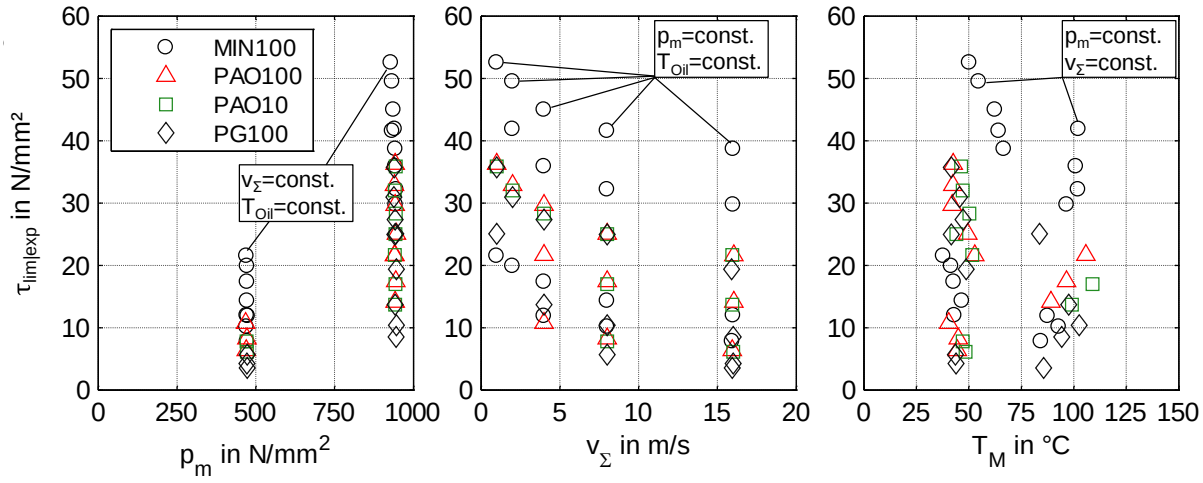


Figure 3: Contact-integral limiting shear stress values $\boldsymbol{\tau_{lim\,|exp}}$ derived from measured friction curves at the FZG twin disk test rig plotted over $\boldsymbol{p_m}$, $\boldsymbol{v_\Sigma}$ and $\boldsymbol{T_M}$ (measurement data acc. to Mayer [22])

Thereby, the following functional relationships of $\tau_{\lim\,|exp}$ with respect to p_m, v_Σ and T_M can be employed:

- $\tau_{lim\,|exp}(p_m) = E_{\dot{\gamma}1} \cdot p_m$ (see Bair [2], Bair und Winer [7], Beilicke et al. [10] and Habchi, Bair und Vergne [17])
- $\tau_{lim\,|exp}(v_\Sigma) = E_{\dot{\gamma}2} + E_{\dot{\gamma}3} \cdot ln(v_\Sigma)$ (see Mayer [22])
- $\tau_{lim\,|exp}(T_{ref} = T_M) = E_{\dot{\gamma}4} \cdot T_{ref}$

As no useable relationship for $\tau_{lim\,|exp}(T_{ref} = T_M)$ has been found in the literature and as only two measurement levels specified by T_{Oil} are available, a simplified linear function is chosen. The bulk temperature T_M is a system-related quantity of the FZG twin disk test rig and increases depending on the operating condition more or less with increasing sliding velocity v_g. Therefore, only T_M at maximal friction (see Figure 1 left) is considered and denoted as reference temperature T_{ref}. Strictly speaking, T_{ref} (and T_M) is not a contact-integral value but more a "boundary condition" of the temperature level of the TEHL contact. However, for almost all derived limiting shear stresses $\tau_{lim\,|exp}$ the amount by which the TEHL contact temperature exceeds T_M is low and comparable, so that T_{ref} can be interpreted as a contact-integral quantity. For the FZG twin disk test rig, it was found that T_{ref} can be well-approximated by a function of p_m, v_Σ and T_{Oil}. This however is not representative of other tribological systems. For gears, T_M is approximately constant over the path of contact, i.e. $T_{ref} = T_M \neq f(p_m, v_\Sigma)$.

The simple operation of adding together the single functional relationships of $\tau_{lim\,|exp}$ in Eq. (1) yields very strong correlation with measured $\tau_{lim\,|exp}$ values (Figure 4 left). Thereby, the lubricant-specific parameters $E_{\dot{\gamma}1}$, $E_{\dot{\gamma}2}$, $E_{\dot{\gamma}3}$ und $E_{\dot{\gamma}4}$ shown in Table 2 are obtained from regression analysis based on Eq. (1). A comparison of the regression quality of the proposed relationship ($\tau_{lim|exp}(p_m, v_\Sigma, T_{M,sys})$, Figure 4 left) with a frequently used linear pressure-dependent relationship ($\tau_{lim|exp}(p_m)$, Figure 4 right) illustrates that the latter is not sufficient.

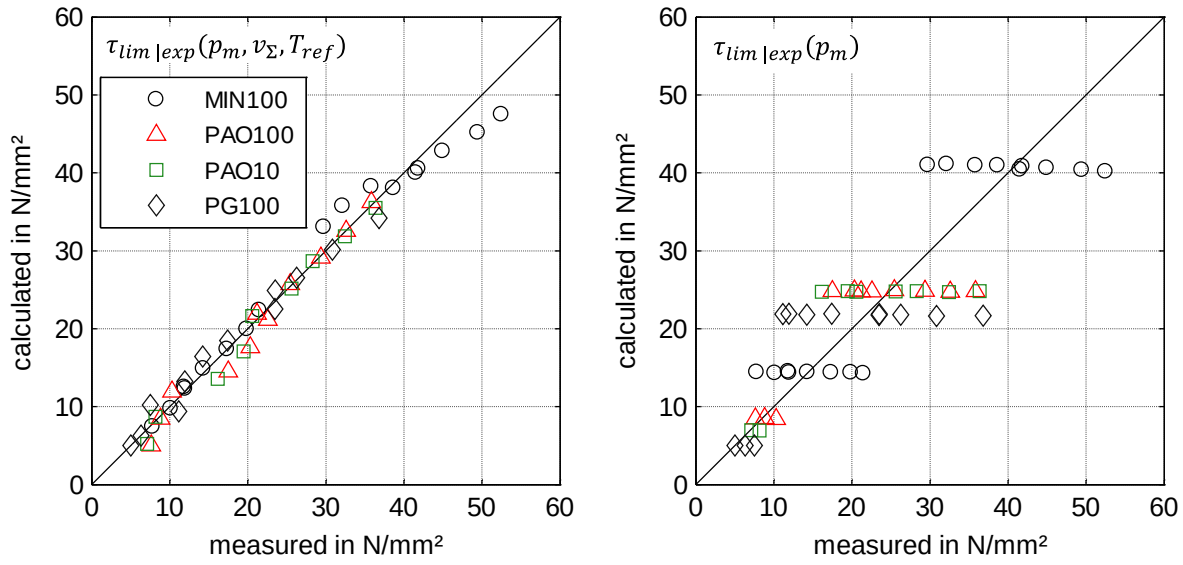


Figure 4: Regression quality of contact-integral limiting shear stress for the proposed relationship $\boldsymbol{\tau_{lim|exp}(p_m, v_\Sigma, T_{M,sys})}$ (left) and the relationship $\boldsymbol{\tau_{lim|exp}(p_m)}$ (right) frequently used in the literature

	MIN 100	**PAO100**	**PAO10**	**PG100**
$E_{\dot{\gamma}1}$	0.0588	0.0376	0.0351	0.0287
$E_{\dot{\gamma}2}\ in\ N/m^2$	32.08	50.21	48.94	70.41
$E_{\dot{\gamma}3}\ in\ N/m^2$	−3.352	−4.770	−4.876	−5.136
$E_{\dot{\gamma}4}\ in\ N/m^2/K$	−0.1195	−0.1564	−0.1455	−0.2005

Table 2: Lubricant-specific parameters of contact-integral limiting shear stress $\tau_{lim\,|exp}$ derived from friction curves

2.2 Contact-local approach

The contact-integral approach $\tau_{lim|exp}(p_m, v_\Sigma, T_{M,sys})$ presented in Eq. (1) is so far only based on evaluated friction curve measurements. In the next step, it is further developed to the contact-local approach $\tau_{lim\,|sim}(p, v_\Sigma, T_{ref}, T)$ in Eq. (2) (see Figure 1 right).

First, p_m is replaced by the local pressure p in the TEHL contact and $\tau_{lim\,|sim}$ is scaled by a factor of $4/\pi$. The latter is required due to the contact-integral nature of $\tau_{lim\,|exp}$ and may be somehow related to the relation $p_H = 4/\pi \cdot p_m$. Example TEHL simulations have shown that the scaling factor of $4/\pi$ provides a very good match for the level of measured friction coefficients for different types of lubricants. Second, a relative temperature dependency of the limiting shear stress is introduced by relating the reference temperature T_{ref} to the local temperature T in the TEHL contact. Hence, $\tau_{lim\,|sim}(p, v_\Sigma, T_{ref}, T)$ becomes a non-linear quantity that varies over the film thickness length (z) and height direction (x). It will be shown that the decrease of friction curves after the maximal friction (see a and b in Figure 1) is partly due to decreasing viscosity and partly due to decreasing limiting shear stress. The *sinh* function provides very good match for the temperature-dependent decrease of the limiting shear stress[1].

[1] The Eyring model is also based on the *sinh* function. As Eyring [14] derived his model from measurements at a capillary viscometer where a proper separation of shear thinning and thermal effects is not possible (Bair et al. [5]), some relation between the temperature dependency of $\tau_{lim\,|sim}$ and primary measured data may be assumed.

A limiting shear stress less than or equal to zero is meaningless. Hence, the minimal value of $\tau_{lim\,|sim}(p, v_\Sigma, T_{ref}, T)$ is generally set to $\tau_{lim,m} = 5N/mm^2$.

2.3 Assumptions

The proposed limiting shear stress formulation for implementation in TEHL simulation models is subject to some assumptions:

- The derived and assumed functional relationships of $\tau_{\lim|exp}$ with respect to p_m, v_Σ and T_{ref} are representative.
- The dependency of $\tau_{lim|exp}$ on the contact-integral quantity p_m is representative for the contact-local quantity p.
- T_{ref} is an appropriate reference value for the temperature dependency of $\tau_{lim|sim}$. The (inevitable) error when calculating T_{ref}/T (T_{ref} represents a bulk temperature) is small and similar for different operating conditions.
- The sum velocity v_Σ influences $\tau_{lim|exp}$ [2] directly.

3 TEHL CALCULATION

The TEHL simulation model used is based on a complete FE implementation and described in detail by Lohner et al. [21]. Only the main characteristics of the model and lubricant properties are repeated here. For simplification, a line contact model is considered.

3.1 TEHL simulation model

The fluid flow is described by the transient generalized Reynolds equation, whereas the elastic deformation of the equivalent body is obtained by solving the linear elasticity equation. For constant temperature, Reynolds, elasticity and load balance equations are solved within a FEM-model(P,H) based on the full-system approach of Habchi [16] achieving very high convergence rates without requiring underrelaxation. For constant pressure and film thickness, the temperature distribution in lubricant and solids is calculated within a FEM-model($\bar{T}$) by solving the corresponding energy equations, which are also based on FE discretization. An iterative procedure is established between the FEM-model(P,H) and FEM-model($\bar{T}$) until the maximum absolute difference of two consecutive solutions of pressure and temperature is smaller than 10^{-3}. The coefficient of friction $\mu_{f|sim}$ of the TEHL contact is evaluated by integrating the shear stress in the middle of the lubricant film:

$$\mu_{f|sim} = \frac{\int_{x_{in}}^{x_{ex}} \tau_{fzx}\big|_{z=\frac{h}{2}} dx}{F_N / l_{eff}} \tag{3}$$

3.2 Temperature and pressure dependency of lubricant properties

The temperature and pressure dependency of the viscosity is defined as suggested by Hepermann et al. [18]. The temperature dependency of the dynamic viscosity at ambient pressure $\eta(T)$ is described acc. to Vogel [27], Fulcher [15] and Tammann and Hesse [26]:

[2] It is also conceivable that v_Σ is only a quantity representing the influence of relaxation time (contact time) or the influence of lubricant film formation on the formation of shear bands.

$$\eta(T) = A_\eta \cdot exp\left(\frac{B_\eta}{C_\eta + (T - 273.15K)}\right) \tag{4}$$

The pressure dependency of the viscosity at a given temperature $\eta(T,p)$ is modelled acc. to Roelands [25]

$$\eta(T,p) = \eta(T) \cdot exp\left\{(ln(\eta(T)) + 9.67) \cdot \left[-1 + \left(1 + \frac{p}{p_{\eta 0}}\right)^{z_\eta(T)}\right]\right\} \tag{5}$$

with a temperature-dependent pressure exponent $z_\eta(T)$ acc. to

$$z_\eta(T) = \frac{\alpha_p(T) \cdot p_{\eta 0}}{\ln(\eta(T)) + 9.67} \tag{6}$$

and a temperature-dependent pressure-viscosity coefficient $\alpha_p(T)$ acc. to

$$\alpha_p(T) \ = E_{\alpha_p 1} \cdot exp\left(E_{\alpha_p 2} \cdot T\right). \tag{7}$$

The temperature and pressure dependency of lubricant density $\rho_f(T,p)$ is modelled following the Bode model [12]:

$$\rho_f(T,p) = \frac{\rho_s \cdot (1 - \alpha_s \cdot T)}{1 - D_{\rho 0} \cdot ln\left(\frac{D_{\rho 1} + D_{\rho 2} \cdot T + D_{\rho 3} \cdot T^2 + p}{D_{\rho 1} + D_{\rho 2} \cdot T + D_{\rho 3} \cdot T^2}\right)} \tag{8}$$

All model constants of the lubricant models above have been obtained from regression of measurements provided by ITR Clausthal [19]. The temperature and pressure dependency of the thermal conductivity $\lambda_f(p)$ and heat capacity per volume $(c_{p,f} \cdot \rho_f)(T,p)$ are based on the models of Larsson and Andersson [20]. Thereby, different parameters are available for different types of lubricant.

3.3 Non-Newtonian fluid behavior

The non-Newtonian fluid behavior is based on the simplified Bair/Winer model (Wolff and Kubo [28]), which can generally be written as:

$$\dot{\gamma}_{\mathrm{fzx}} = \frac{1}{G} \cdot \frac{d\tau_{fzx}}{dt} + \frac{\tau_{fzx}}{\eta(T,p)} \cdot \frac{1}{\left(1 - \left|\frac{\tau_{fzx}}{\tau_{\lim|sim}}\right|\right)} \ \ with \ \tau_{fzx}(T,p,\dot{\gamma}_{fzx}) = \eta(T,p,\dot{\gamma}_{fzx}) \cdot \dot{\gamma}_{fzx} \tag{9}$$

Wolff and Kubo [28] show that the visco-elastic term mainly influences friction for very small slip ratios. Therefore and due to unknown values of the shear modulus, the visco-elastic term is neglected ($G = \infty$) for the sake of simplicity.

4 RESULTS AND DISCUSSION

In the following, simulated values for the coefficients of friction $\mu_{f|sim}$ are compared with measured values μ_{exp} given by Mayer [22] (see section 2.1) for a large variety of operating conditions and for the lubricants MIN100, PAO100, PAO10 and PG100.

Figure 5 exemplifies the simulated and measured friction curves for MIN100 and PG100, with two different operating conditions (see also Figure 2). The measured coefficients of friction very strongly agree with the full-potential TEHL simulation results (thermal, $\sinh(T_{ref}/$

T) acc. to Eq. (2), solid lines)[3]. Deviations can in particular be recognized at the beginning of the friction curves at the non-linear (shear thinning) transition regime. Even though these zones of the friction curves require a higher measurement resolution for systematic comparisons, these deviations could possibly be due to neglected visco-elastic effects ($G = \infty$ in Eq. (9)) and an inaccurate representation of shear thinning by Eq. (9). Furthermore, the overestimation of the coefficient of friction at beginning of the friction curves for MIN100 appears to be correlated with the strong increase of the viscosity of MIN100 with pressure measured by ITR Clausthal [19]. In this context, stationary viscosity measurements at high pressure viscometers may be not simply transferable to TEHL contacts featuring very low contact times.

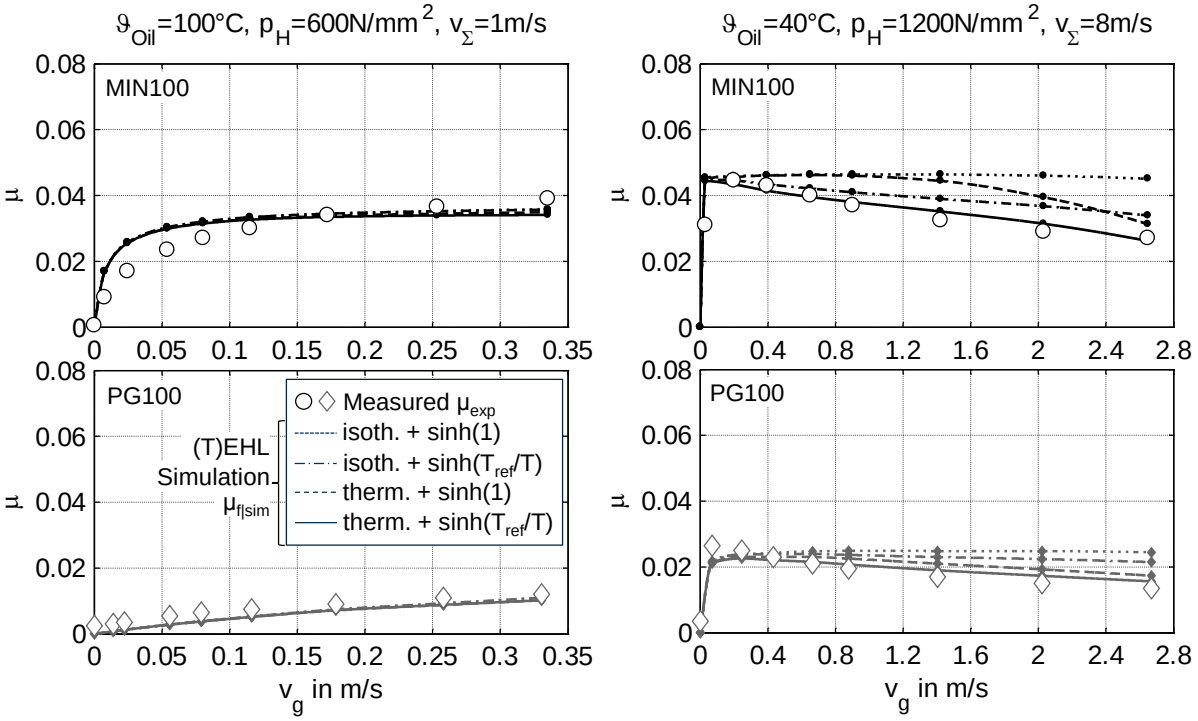


Figure 5: Comparison of simulated ($\boldsymbol{\mu_{f|sim}}$) and measured ($\boldsymbol{\mu_{exp}}$) coefficients of friction for two different operating conditions and for lubricants MIN100 and PG100 (measured data acc. to Mayer **[22]**)

In order to quantify the complex relationship of the influences on the simulated coefficients of friction, simulation results with decreasing level of detail are provided in Figure 5. As its effects are best recognizable for the operating condition with $p_H = 1200N/mm^2$, $\vartheta_{Öl} = 40°C$ and $v_\Sigma = 8m/s$, the following explanations are focused on that.

When thermal effects and the relative temperature dependency of $\tau_{\lim|sim}$ (isothermal, $sinh(1)$ in Eq. (2), dotted lines) are deactivated, the friction curves are almost constant as the sliding velocity v_g increases. Despite $\tau_{\lim|sim}(v_g) = const$, there is a small decrease at high values of v_g caused by shear thinning effects (Habchi, Bair und Vergne [17]). When thermal effects (thermal, sinh(1) in Eq. (2), dashed lines) are activated, a comparatively large decrease in $\mu_{f|sim}$ can be observed. This is mainly due to viscous heating and the corresponding reduction of integral viscosity. The remaining difference in μ_{exp} is a result of the temperature de-

[3] The simulations are exclusively conducted for the measurement points of Mayer [22] (nine measurement points for each friction curve, in-between interpolated for the sake of presentation). The measured bulk temperature of each measurement point serves as boundary condition of the TEHL simulation model.

pendency of $\tau_{lim|sim}$ (isothermal, $\sinh(T_{ref}/T)$ in Eq. (2), dot-dashed lines). The effects of the individual influences depend significantly on the bulk and TEHL contact temperatures, interact mutually and cannot simply be summed together.

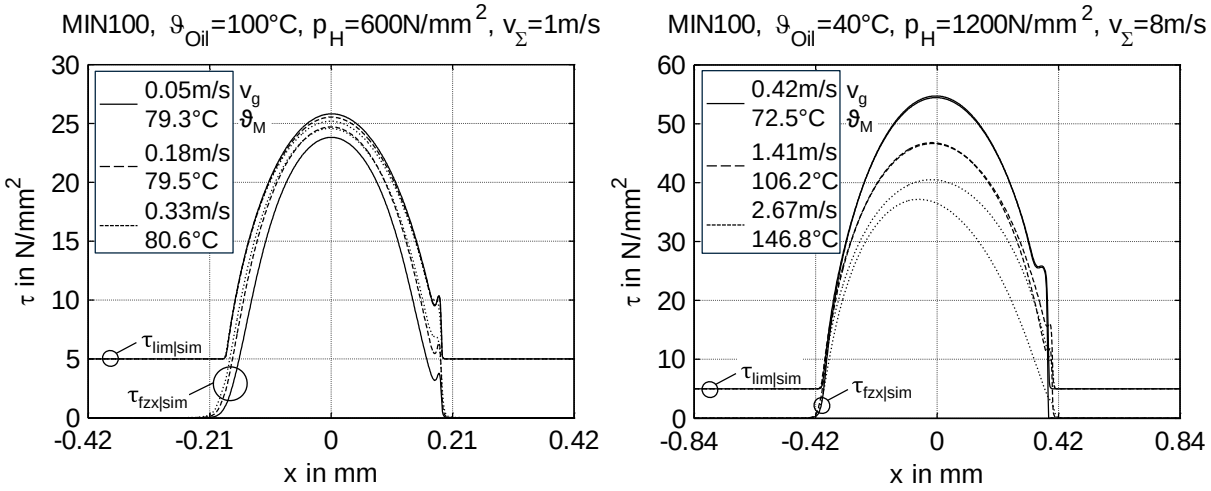


Figure 6: Contact-local shear stress ($\boldsymbol{\tau_{fzx|sim}}$) and limiting shear stress ($\boldsymbol{\tau_{lim\,|sim}}$) distributions in the middle of the lubricant film thickness ($\boldsymbol{z = h/2}$) for MIN100 and selected operating conditions in **Figure 5**

Figure 6 exemplifies the contact-local distributions of the shear stress $\tau_{fzx|sim}$ and limiting shear stress $\tau_{lim\,|sim}$ at the midpoint of the lubricant film thickness ($z = h/2$) for MIN100, with the operating conditions selected in Figure 5 (full-potential TEHL simulation results). The contact-local behavior of $\tau_{lim\,|sim}$ is acc. to Eq. (2) determined by its dependency on the local quantities p and T. This also means that $\tau_{lim\,|sim}$ may only be achieved at specific locations in the TEHL contact. Figure 6 (right) demonstrates again on a contact-local basis that the decrease of the friction curves after a maximal friction value is partly due to decreasing limiting shear stress and partly due to decreasing viscosity with increasing TEHL contact temperature: For the low and medium sliding velocity of $v_g = \{0.42|1.41\} m/s$, $\tau_{lim\,|sim}$ is reached for almost the whole contact area, whereas for the high sliding velocity of $v_g = 2.67 m/s$, $\tau_{fzx|sim}$ is significantly lower than $\tau_{lim\,|sim}$. For the low load operating condition in Figure 6 (left), $\tau_{lim\,|sim}$ is not reached at all for the whole contact area. Moreover, $\tau_{lim\,|sim}$ is almost equal for all sliding velocities, because its relative temperature dependency has almost no influence, due to the low TEHL contact temperatures. Nevertheless, $\tau_{lim\,|sim}$ determines the level of the coefficient of friction.

In summary, Figure 7 shows a comparison of the simulated ($\mu_{f|sim}$) and measured (μ_{exp}) coefficients of friction for all investigated operating conditions of Mayer [22]. It can be seen that the simulated values for the coefficients of friction based on the introduced limiting shear stress formula in section 2 and the non-Newtonian fluid behavior in section 3.3 correlate very strongly with the measured values over a large spread of operating conditions and different types of lubricants. In order to assign the symbols in Figure 7 to the "position" on the friction curves, the symbols are colored proportionally to the slip ratio. Most of the symbols with poor correlation between $\mu_{f|sim}$ and μ_{exp} show very low slip ratios and are therefore located at the beginning of the friction curves. Reasons for this have already been discussed above. It should be mentioned that for some friction curves of PAO10 small solid load portions and therefore

mixed lubrication regimes are not precluded. This has however so far not been considered in the presented TEHL simulation model.

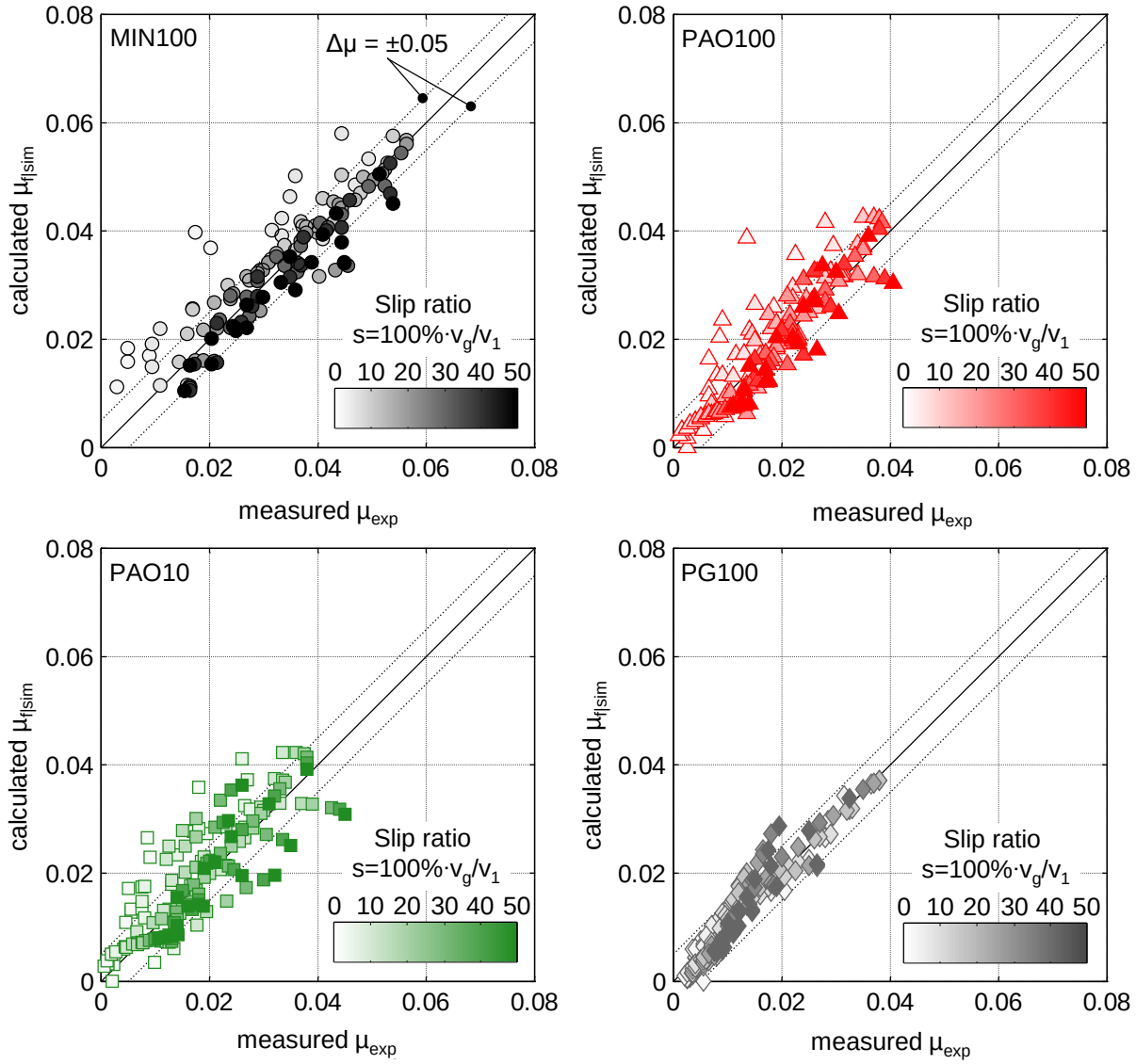


Figure 7: Comparison of simulated ($\boldsymbol{\mu_{f|sim}}$) and measured ($\boldsymbol{\mu_{exp}}$) coefficients of friction for all operating conditions and for lubricants MIN100, PAO10, PAO100 and PG100 (measured data acc. to Mayer [22])

5 CONCLUSION

This paper focuses on a limiting shear stress formulation for use in TEHL simulation models. The main conclusions drawn from this work are:

- Non-Newtonian fluid models based on limiting shear stress are currently probably the most physical and suitable way to calculate the friction of TEHL contacts.
- The limiting shear stress is not a constant lubricant property but depends on pressure, temperature, entrainment velocity and type of lubricant.
- The contact-integral nature of limiting shear stresses derived from friction measurements at tribometers has been successfully adjusted for local use in TEHL simulation models.

- The presented contact-local limiting shear stress formulation is transferable to other tribological systems apart from the twin disk test rig e.g. gear contacts.
- TEHL simulation results show very good agreement between measured and simulated coefficients of friction for a large variety of operating conditions and lubricants.
- When the sliding velocity is increased further once the maximal friction is reached, the portion of friction reduction due to decreasing lubricant viscosity and the portion due to temperature dependency of the limiting shear stress have been quantified.

Further work might focus (i) on even extending the data basis for the contact-integral limiting shear stress equation, (ii) on accurate modelling of the beginning of the friction curves at the non-linear (shear thinning) transition regime, and (iii) on the transition of the calculated contact-local limiting shear stress $\tau_{lim\,|sim}$ to the minimum limiting shear stress $\tau_{lim,m}$.

NOMENCLATURE

A_H	Hertzian flattening area	x	Film thickness length direction
$c_{p,f}$	Specific heat capacity of lubricant	z	Film thickness height direction
$E_{\dot{\gamma}1}$, $E_{\dot{\gamma}2}$, $E_{\dot{\gamma}3}$, $E_{\dot{\gamma}4}$	Lubricant specific parameters of $\tau_{lim\,|exp}$	z_η	Pressure exponent of Roelands equation
F_N	Normal force	α_p	Pressure-viscosity exponent
F_R	Friction force	$\dot{\gamma}_{fzx}$	Simulated shear rate
G	Shear modulus of lubricant	η	Dynamic viscosity
h	Lubricant film thickness	ϑ_M	Bulk temperature in °C
l_{eff}	Effective contact width	ϑ_{Oil}	Oil inlet temperature in °C
p	Pressure	λ_f	Thermal conductivity of lubricant
p_H	Hertzian pressure	μ_{exp}	Measured coefficient of friction
p_m	Mean Hertzian pressure ($= \pi/4 \cdot p_H$)	$\mu_{f\,|sim}$	Simulated coefficient of friction
s	Slip ratio ($= 100\% \cdot v_g/v_1$)	ν	Kinematic viscosity
T	Temperature in K	ρ_f	Density of lubricant
T_M	Bulk temperature in K	τ_{fzx}	Simulated shear stress
T_{Oil}	Oil inlet temperature in K	$\tau_{lim\,|exp}$	Contact-integral limiting shear stress
T_{ref}	Reference temperature in K	$\tau_{lim\,|sim}$	Contact-local limiting shear stress
v_g	Sliding velocity ($= v_1 - v_2$)	$\tau_{lim\,|m}$	Minimum limiting shear stress
v_Σ	Sum velocity ($= v_1 + v_2$)		

REFERENCES

[1] Bair, S.: A Rough Shear-Thinning Correction for EHD Film Thickness. Tribology Transactions 47(3): pp. 361-365 (2004).

[2] Bair, S.: High Pressure Rheology for Quantitative Elastohydrodynamics. 1st Edition, Amsterdam: Elsevier (2007).

[3] Bair, S.: The Nature of the Logarithmic Traction Gradient. Tribology International 35: pp. 591-597 (2002).

[4] Bair, S.; McCabe, C.: A Study of Mechanical Shear Bands in Liquids at High Pressure. Tribology International 37: pp. 783-789 (2004).

[5] Bair, S.; Vergne, P.; Kumar, P.; Poll, G.; Krupka, I.; Hartl, M.; Habchi, W.; Larrson, R.: Comment on „History, Origins and Prediction of Elastohydrodynamic Friction“ by Spikes and Jie. Tribology Letters 58(16): pp. 1-8 (2015).

[6] Bair, S.; Winer, W. O.: A Rheological Model for Elastohydrodynamic Contacts based on Primary Laboratory Data. ASME Journal of Lubrication Technology 101: pp. 258-265 (1979).

[7] Bair, S.; Winer, W. O.: Some Observations in High Pressure Rheology of Lubricants. Journal of Lubrication Technology 104: pp. 357-364 (1982).

[8] Bair, S.; Winer, W. O.: The High Pressure High Shear Stress Rheology of Liquid Lubricants. Journal of Tribology 114: pp. 1-9 (1992).

[9] Bartel, D.: Simulation von Tribosystemen: Grundlagen und Anwendungen. Habilitationsschrift. 1st Edition, Wiesbaden: Vieweg+Teubner (2009).

[10] Beilicke, R.; Bobach, L.; Bartel, D.; Bader, N.; Poll, G.; Brouwer, L.; Schwarze, H.: Tribologische Fluidmodelle zur Simulation der Reibung in geschmierten konzentrierten Kontakten. 56. GfT-Tribologie-Fachtagung, Göttingen, pp. 5/1-5/14 (2015).

[11] Bobach, L.; Bartel, D.; Beilicke, R.; Mayer, J.; Michaelis, K.; Stahl, K.; Bachmann, S.; Schnagl, J.; Ziegele, H.: Reduction in EHL Friction by a DLC Coating. Tribology Letters 60(17): pp. 1-13 (2015).

[12] Bode, B.: Modell zur Beschreibung des Fließverhaltens von Flüssigkeiten unter hohem Druck. Tribologie und Schmierungstechnik 36: pp. 182-189 (1989).

[13] Evans, C. R.; Johnson, K. L.: The Rheological Properties of Elastohydrodynamic Lubricants. Journal of Mechanical Engineering Science 200(5): pp. 303-312 (1986).

[14] Eyring, H.: Viscosity, Plasticity, and Diffusion as Examples of Absolute Reaction Rates. Journal of Chemical Physics 4(4): pp. 283-291 (1936).

[15] Fulcher, G. S.: Analysis of Recent Measurements of the Viscosity of Glasses II. Journal of the American Ceramic Society 8(12): pp. 789-794 (1925).

[16] Habchi, W.: A Full-System Finite Element Approach to Elastohydrodynamic Lubrication Problems: Application to Ultra-Low-Viscosity Fluids. Dissertation. L’Institut National des Sciences Appliquées, Lyon (2008).

[17] Habchi, W.; Bair, S.; Vergne, P.: On Friction Regimes in Quantitative Elastohydrodynamics. Tribology International 58: pp. 107-117 (2013).

[18] Hepermann, P.; Beilicke, R.; Bartel, D.; Tenberge, P.; Deters, L.: AiF-Nr. 16057, FVA-Nr. 598/I - Heft 1024 - Örtliche Fresstragfähigkeit Abschlussbericht - Bestimmung der örtlichen Fresstragfähigkeit: Einfluss von Schräg- und Hochverzahnungen. Forschungsvereinigung Antriebstechnik e.V., Frankfurt/Main (2012).

[19] ITR Clausthal: IGF-Nr. 15786 N, FVA-Nr. 583/I - Heft 1002 - Thermophysikalische Eigenschaften (Schmierstoffe) - Bestimmung und Modellierung der thermophysikalischen Eigenschaften von Schmier- und Kraftstoffen unter hohen Drücken. Forschungsvereinigung Antriebstechnik e.V., Frankfurt/Main (2013).

[20] Larsson, R.; Andersson, O.: Lubricant Thermal Conductivity and Heat Capacity under High Pressure. Journal of Engineering Tribology 214 (4): pp. 337-342 (2000).

[21] Lohner, T.; Ziegltrum, A.; Stemplinger, J.-P.; Stahl, K.: Engineering Software Solution for Thermal Elastohydrodynamic Lubrication (TEHL) Using Multiphysics Software. Advances in Tribology. Article ID 6507203: doi: 10.1155/2016/6507203, 13 pages (2016).

[22] Mayer, J.: Einfluss der Oberfläche und des Schmierstoffs auf das Reibungsverhalten im EHD-Kontakt. Dissertation. Technical University of Munich (2013).

[23] Paschold, C.: Ableitung eines Grenzschubspannungsansatzes für Schmierstoffe auf Basis experimenteller Untersuchungen am Zweischeibenprüfstand. Semester Thesis, Institute of Machine Elements, Technical University of Munich (2015).

[24] Research Project FVV 1138: Tribologische Fluidmodelle. Research Association for Combustion Engines e. V. (FVV e. V.). Funded via the German Federation of Industrial Research Associations e. V. (AiF e. V.) by the German Federal Ministry for Economic Affairs and Energy (BMWi).

[25] Roelands, C. J. A.: Correlation Aspects of the Viscosity-Temperature Relationship of Lubricating Oil. Dissertation. Delft University of Technology (1966).

[26] Tammann, G.; Hesse, W.: Die Abhängigkeit der Viskosität von der Temperatur bei unterkühlten Flüssigkeiten. Z. f. anorg. u. allg. Chemie 156: pp. 245-257 (1926).

[27] Vogel, H.: Das Temperaturabhängigkeitsgesetz der Viskosität von Flüssigkeiten. Physik. Zeitschrift 22: pp. 645-647 (1921).

[28] Wolff, R.; Kubo, A.: A Generalized Non-Newtonian Fluid Model Incorporated Into Elastohydrodynamic Lubrication. Journal of Tribology 118(1): pp. 74-82 (1996).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

THE XFEM FOR A SIMPLIFIED MODEL IN HYDRAULIC FRACTURING

Markus Schätzer[1], **Thomas-Peter Fries**[1]

[1]Institute of Structural Analysis, Graz University of Technology
Lessingstr. 25/II, 8010 Graz, Austria
e-mail: {schaetzer, fries}@tugraz.at

Keywords: energy release rate, stress intensity factors, superposition, crack opening displacements.

Abstract. *Hydraulic fracturing is an important branch in geomechanics, where a fracture initiates and propagates in the host rock due to the induced hydraulic loading. The pressure exerted by the fracking fluid onto the surrounding solid, typically is obtained by a Reynolds equation which relates the crack width with the pressure. This poses severe difficulties for the numerical treatment, also in terms of robustness. In the presented work, the Reynolds equation is replaced by pre-defined pressure distributions which leads to a much simpler and robust coupled problem. For each assumed distribution, the critical pressure is determined where the critical energy release rate G_c is exceeded. The influence of the different distributions is investigated in a separate example. The energy release rates are determined based on stress intensity factors with the XFEM by using crack opening displacements. This happens by a comparison of an approximated state which represents the computed displacements in the solid, and a reference state which represents the expected openings for a pure mode I, II and III. This method is intuitive, computationally cheap and has the advantage that only displacements are fitted, wherefore no additional consideration of pressurized crack surfaces is required.*

The critical pressure is determined based on the superposition principle in linear elastic fracture mechanics. Herein, a separate observation of the external loadings by means of volume forces and tractions at the boundary, and the internal loadings by means of the pressure exerted by the fracking fluid is done. Based on a scaling factor of the internal state, the critical pressure is extracted from the energy release rate.

1 Introduction

Hydraulic fracturing is a standard procedure for the stimulation of petroleum and geothermal reservoirs. A fluid is pumped into the solid until a fracture initiates and propagates. For the modelling, three important factors have to be considered: (*i*) the pressure exerted by the fracking fluid onto the surrounding solid, (*ii*) the deformation of the solid and the associated fracture volume, and (*iii*) the propagation of crack surfaces.

In general, the pressure induced by the fracking fluid is obtained by a Reynolds equation [1] which relates the pressure with the crack width. That is, several iterations between field (*i*) and (*ii*) are necessary, which is not always robust and leads to a complex behaviour at the crack tip/front (toughness or viscous regimes [2]). This contribution deals with a simplified model where the Reynolds equation is replaced by pre-defined pressure distributions which are scaled by only few parameters. This assumption leads to a much simpler and robust coupled problem. It is noted that only polynomial pressure distributions, depending on the distance r to the crack tip/front and an initial pressure value p_0 are considered in this contribution as shown in Section 4.2. The introduced method is not limited on this restricted group of functions and treats all pressure distributions in the same manner.

Many numerical methods exist for solving the linear elastic fracture problem, however, most of them such as the standard finite element method (FEM) and the boundary element method (BEM) have a limited efficiency due to the $\sqrt{r}$ behaviour of the displacements in the vicinity of the crack tip/front. The use of special elements, like 'quarter-point' elements [3, 4, 5], improves the approach, but a suitable mesh has to be provided for each crack geometry during propagation which frequently requires a remeshing. An optimal accuracy on *fixed* meshes is possible with the extended finite element method (XFEM) [6] which treats non-smooth solution features within elements by additional enrichment functions. In this method, the localisation of the non-smooth features or in the context of linear elastic fracture mechanics the localisation of the crack path/surface is often done by the use of level-set functions [7, 8]. However, the update of these functions after a propagation step can be cumbersome [8], wherefore in [9] a hybrid explicit-implicit crack description is used which combines the advantages of the level-set functions with those of an explicit crack description, e.g. by means of straight line segments or flat triangles. This crack representation allows a simple update of the crack geometry during the crack propagation, wherefore it is used in this work.

In the third field (*iii*), it is determined whether the crack propagates and in which direction. The maximum circumferential stress criterion is often employed, where the crack propagates in direction of the maximum circumferential stresses when a critical energy release rate G_c is reached [6]. In linear elastic fracture mechanics, the current energy release rate G is often evaluated by the J-integral [10] as they are equivalent [11]. The direction of the propagation is often determined by stress intensity factors (SIFs) which are also related to G [12].The interaction integral is one of the most important technique for the computation of SIFs, however the evaluation of this integral can be cumbersome for complex three-dimensional crack configurations. In this work, SIFs are determined by observing the displacement field, particularly the crack opening displacements (CODs), in the vicinity of the crack tip/front and their comparison with the expected openings for a pure mode *I*, *II* and *III* crack.

The focus of this contribution is the evaluation of critical pressure values p_c for assumed pressure distributions on mixed-mode loaded crack configurations. This examination is the first step towards a simplified hydraulic fracturing model.

2 Computation of SIFs by using CODs in XFEM

In this section, the XFEM approach in linear elastic fracture mechanics with a hybrid explicit-implicit crack description and the computation of stress intensity factors (SIFs) by crack opening displacements is shortly recalled.

The XFEM extends the standard finite element method (FEM) approach by additional enrichment functions which involve non-smooth features within elements with optimal accuracy. A *priori* knowledge about the non-smooth solution characteristics is essential for the enrichment functions. In linear elastic fracture mechanics, a domain Ω is cracked by the (curved) crack path/surface Γ_c. The displacements are discontinuous along Γ_c and the stresses are singular at the crack tip/surface and the following enriched approximation has proven useful [6]

$$u^h(\mathbf{x}) = \sum_I N_i(\mathbf{x}) u_i + \sum_{I^*} N_i(\mathbf{x}) \psi_{step}(\mathbf{x}) a_i + \sum_{j=1}^{4} \sum_{J^*} N_i(\mathbf{x}) \psi_{tip}^j(r,\theta) b_i^j. \tag{1}$$

Herein, ψ_{step} considers discontinuous displacements along the crack path/surface and ψ_{tip} the singular stresses at the crack tip/front. N_i are the finite element shape functions, a_i and b_i^j additional degrees of freedom at the enriched nodes I^* and J^*. It is standard to use the Heaviside function for the step enrichment ψ_{step} [6] and

$$\psi_{tip}^j(r,\theta) = \left[\sqrt{r} sin\frac{\theta}{2},\ \sqrt{r} cos\frac{\theta}{2},\ \sqrt{r} sin\frac{\theta}{2} sin\theta,\ \sqrt{r} cos\frac{\theta}{2} sin\theta \right] \tag{2}$$

for the crack tip/front enrichment. These enrichment functions are based on a polar coordinate system (r,θ) which has its origin at the crack tip/front and is aligned with the tangent at the crack tip/front. Typically, the localisation of the crack geometry and the determination of the polar coordinate system is based on level-set functions. In [9] a hybrid explicit-implicit crack description has been introduced which allows a simple update of the level-sets during propagation. Herein, three level-set functions are defined as follows:

- $\phi_1(\mathbf{x})$ is the (unsigned) distance function to the crack path/surface.
- $\phi_2(\mathbf{x})$ is the (unsigned) distance function to the crack tip/front.
- $\phi_3(\mathbf{x})$ is a signed distance function to the extended crack path/surface.

These definitions extend straightforward also to three dimensions and imply the polar coordinate system (r,θ) which is used for the enrichment functions, and a local coordinate system (a,b) which is used for the computation of SIFs. Further details of the different coordinate systems are given in [9]. In the following, the evaluation of SIFs by crack opening displacements (CODs) is discussed.

2.1 Computation of SIFs

The proposed method offers an intuitive and computationally cheap technique for the evaluation of SIFs which treats three dimensional crack configurations in the same manner as two dimensional ones. Furthermore, no modification is necessary for curved and loaded crack surfaces. The use of CODs has the advantage that translations by means of rigid body motions are considered automatically. SIFs are computed by a comparison of CODs of an approximated

state by means of the XFEM solution and a reference state which represents the expected openings for a pure mode *I*, *II* and *III* crack. The following paragraph shows the main idea of the proposed method.

The approximated CODs are determined by Eq. 1 as a post-process step. The difficulty here is that the enrichment and shape functions have to be evaluated in a special setting so that CODs are obtained. For that, in a first step a point has to be found on the implicitly defined crack path/surface. As level-set values are only given in the nodes and are interpolated within the elements by the corresponding shape functions, curved zero level-sets are present in general. This complicates the finding of points with zero level-set values within the elements. Therefore, each potentially cut element, identified by the condition

$$max[\phi_3(\mathbf{x}_i)] \cdot min[\phi_3(\mathbf{x}_i)] < 0, \tag{3}$$

is subdivided into two linear triangular elements in two dimensions or six linear tetrahedral elements in three dimensions, which enable a simple detection of zero level-sets as they are piecewise linear/planar within these simplex elements. A similar setting is used for linear elements in [13, 14] and for higher order elements in [15]. As the enrichment functions and displacements are discontinuous along the zero level-sets the point has to be *splitted* into two opposite but infinitesimally close points of the crack path/surface. This splitting is done in the simplex elements based on the normal vector of the zero level-set. Then the global crack opening displacement $\Delta\mathbf{u}^h_{\mathbf{x}}(S)$ by means of the global coordinate system (x,y,z) of the point S is evaluated by Eq. 1.

The expected crack opening displacement $\Delta\mathbf{u}^m_{\mathbf{a}}(S)$ of the point S and mode m is evaluated in the (a,b,c)-coordinate system by the governing equations of linear elastic fracture mechanics based on stress intensity factors [16] where, successively, one SIF k_m is set to 1 and the others to 0.

A comparison of these two states leads to the SIFs. As both CODs have to be in the same coordinate system, the approximated CODs are transformed into the (a,b,c)-coordinate system based on the Jacobi-matrix $\mathbf{J}$ as shown in Eq. 4. Herein, the partial derivatives of the third function c are determined by the cross product of ∇a and ∇b.

$$\Delta\mathbf{u}^h_{\mathbf{a}} = \mathbf{J} \cdot \Delta\mathbf{u}^h_{\mathbf{x}} \qquad \text{with} \quad \mathbf{J} = \begin{bmatrix} a_{,x} & a_{,y} & a_{,z} \\ b_{,x} & b_{,y} & b_{,z} \\ c_{,x} & c_{,y} & c_{,z} \end{bmatrix}, \quad i = 1,2,3. \tag{4}$$

It has been shown, for example, in [16] that each stress intensity factor k_m has a preferred direction with the highest impact on the associated displacements. Taking these directions into account, the SIFs can be directly computed by

$$k_I = \Delta u^h_b / \Delta u^I_b, \qquad k_{II} = \Delta u^h_a / \Delta u^{II}_a, \qquad k_{III} = \Delta u^h_c / \Delta u^{III}_c. \tag{5}$$

This fitting method is the basis for the further work, as all needed SIFs are evaluated by it. In the following section, the evaluation of critical pressure values for a general mixed-mode loaded crack configuration is discussed.

3 Evaluation of critical pressure values

In hydraulic fracturing, a fluid is pumped into the solid until a fracture initiates and propagates. The fracture propagates when a critical value of the energy release rate G is reached. That is, the crack propagates if the following equation is fulfilled

$$G \geq G_c. \tag{6}$$

The current energy release rate G is obtained by the associated SIFs which are evaluated by the proposed method presented in Section 2.1 and their relation to G. For a general three dimensional mixed-mode loading and the assumption that mode *II* and *III* are small compared to mode *I* which leads to a *quasi planar* crack propagation, the energy release rate is given by [12] as follows:

$$G(k_m) = \frac{\left(1-\nu^2\right)}{E}\left(k_I^2 + k_{II}^2\right) + \frac{1}{2\mu}k_{III}^2. \tag{7}$$

Herein, k_m are the current SIFs, ν the Poisson's ratio and μ the second Lamé-Constant. It is noted that this representation of G is used for simplicity. The proposed method is not limited to it, wherefore also formulations of infinitesimally kinked crack extensions [16, 17] are possible which will be considered in future works.

In general, the pressure exerted by the fracking fluid onto the surrounding solid is obtained by a Reynolds equation which relates the crack width with the pressure. In this contribution, the model is simplified as the Reynolds equation is replaced by pre-defined pressure distributions which are scaled by the initial value p_0 and the crack length l_c as shown in Section 4.2. One may distinguish two different types of loadings in hydraulic fracturing. The first is called *'external'* and contains body forces, prescribed tractions at the boundary Γ_t and prescribed displacements at the boundary Γ_u. The second type of forces is called *'internal'* and contains the pressure exerted by the fracking fluid onto the surrounding solid. For the second type it is also demanded that the displacements at Γ_u are zero. An illustration of a cracked domain, where both loading types are shown is given in Fig. 1(a).

In most practical applications of hydraulic fracturing, the interest is based on the computation of a critical pressure p_c which causes a propagation of the crack. Hence, this contribution deals with the evaluation of critical pressures for pre-defined distributions. In linear elastic fracture mechanics, there holds for a unit prescribed pressure distribution p_0 without external loadings that their associated SIFs scale proportionally with

$$k_m^p(\lambda \cdot p_0) = \lambda \cdot k_m^p(p_0). \tag{8}$$

Herein, the index p indicates that the SIFs are obtained due to pure internal loadings. A differentiation of the SIFs due to internal and external loadings is needed at a later point of this work. Eq. 8 plugged into Eq. 7 and Eq. 6 leads to the critical pressure p_c as

$$p_c = \lambda \cdot p_0 \qquad \text{with}: \quad \lambda = \sqrt{\frac{G_c}{G\left(k_m^p(p_0)\right)}}. \tag{9}$$

Eq. 9 is only valid for pure internal loadings, as an increase of the pressure level does not influence the crack behaviour or energy release rate G caused by the external loadings. Therefore, a separate observation of the crack behaviour caused by the external and internal loadings is necessary. In linear elastic fracture mechanics, this is achieved by using the superposition principle as shown in Fig. 1. Herein, the crack configuration is decoupled into a part with only external loadings and a part with only internal loadings. The associated SIFs are denoted by k_m^{ext} for the external case, k_m^p for the internal case and k_m^{total} for the combined case. If a λ-scaled unit pressure distribution p_0 is assumed, Eq. 8 holds for the internal part. The total SIFs k_m^{total} are given by the sum of the external and internal case.

$$k_m^{total} = k_m^{ext} + \lambda k_m^p. \tag{10}$$

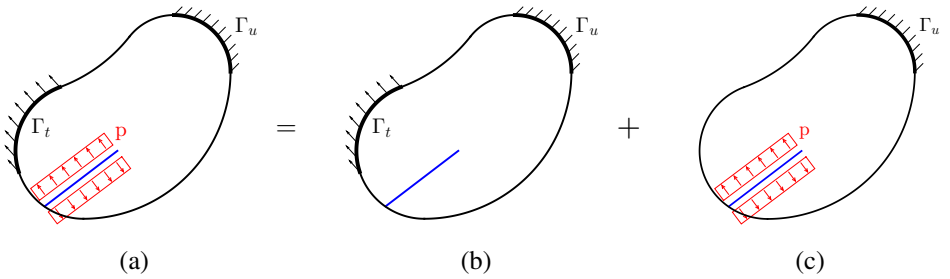


Figure 1: Superposition of an externally and internally loaded crack by means of the principle of superposition. (a) Totally loaded, (b) externally loaded and (c) internally loaded.

Eq. 10 plugged into Eq. 7 and Eq. 6 leads to a quadratic equation of λ as it holds for the principle of superposition applied to the energy release rate G_m of mode m

$$G_m = \left(\sqrt{G_m^{ext}} + \sqrt{G_m^{p}}\right)^2. \tag{11}$$

That is, for the computation of λ, coupled terms of the external and internal state are involved. The quadratic equation for λ can be expressed as follows:

$$\begin{aligned} 0 = a\lambda^2 + b\lambda + c \qquad \text{with} \quad & a = G(k_m^p), \\ & b = \frac{2(1-\nu^2)}{E}\left(k_I^{ext}k_I^p + k_{II}^{ext}k_{II}^p\right) + \frac{1}{\mu}k_{III}^{ext}k_{III}^p, \\ & c = G(k_m^{ext}) - G_c. \end{aligned} \tag{12}$$

It is noted that only one value of λ ($\lambda \geq 0$) leads to a physically reasonable critical pressure. This is checked by the criterion

$$\lambda_1\lambda_2 > 0 \qquad \text{for} \quad \lambda_1 \neq \lambda_2. \tag{13}$$

The evaluated λ leads to the critical pressure p_c by Eq. 9. In the following section, first results of the proposed simplified model for hydraulic fracturing are presented in two dimensions.

4 Numerical examples

In this section two different things are investigated in two dimensions. Firstly, how the critical pressure value due to a constant pressure distribution is changed for a given mixed-loaded crack configuration relating to the crack length. Furthermore, how the crack propagates by the presented critical pressure and external loading. Secondly, how the pressure distribution influences the critical pressure. For the investigation of these two points an edge cracked rectangular plate with the extent of $h = 7\,\text{m}$, $l = 16\,\text{m}$ and an initial crack length $l_c = 3.5\,\text{m}$ is used as shown in Fig. 2(a). The plate is clamped on the left side and is described by 4-node bilinear quadrilateral elements with different mesh sizes. A brittle and isotropic material is used with a Youngs modulus $E = 35\,\text{GPa}$, a Poissons ratio $\nu = 0.3$ and a critical energy release rate $G_c = 1\,\text{J}/\text{m}^2$. Furthermore, plane stress conditions are assumed. Obviously, the chosen setup is not really related to a problem in geomechanics. Nevertheless, the approach is general enough that it is also working in the context of plates where we find more general loading conditions such as those dominated by shear. The interest is only based on the computation of critical pressure values for a given pressure distribution, wherefore no dynamic effectes [18], cohesive models [19] or viscosity-dominated regimes [20] are considered.

4.1 Mixed-mode loaded crack configuration

At first the change of the critical pressure during a crack propagation of a mixed-mode loaded crack configuration is investigated. For that, the plate is loaded by a shear traction $\tau = 1\,\mathrm{kN/m}$ on the right side and a constant pressure within the fracture as shown in Fig. 2(a). Herein, the

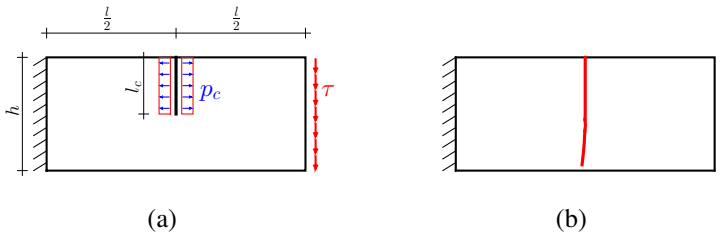


Figure 2: Mixed-loaded crack configuration: (a) Situation and (b) crack propagation.

external loading τ leads to a mixed-mode behaviour at the crack tip, whereas the internal loading only represents a mode I crack. For each propagation step the critical pressure is evaluated by the introduced method, which leads to a propagation of the crack. Based on the maximum circumferential stress criterion, the propagation angle θ_c can be expressed by the SIFs as follows [6]:

$$\theta_c = 2arctan\frac{1}{4}\left(\frac{k_I}{k_{II}} \pm \sqrt{\left(\frac{k_I}{k_{II}}\right)^2 + 8}\right), \tag{14}$$

where the sign depends on the sign of k_{II}. For the considered example, it is assumed that the crack propagates with the increment $\Delta a = 0.35\,\mathrm{m}$ in direction of θ_c. The required pressure for the propagation is illustrated in Fig. 3, where the critical pressure is plotted over the current crack length. Herein, seven different meshes are used with an element number from 9,401 to 46,025.

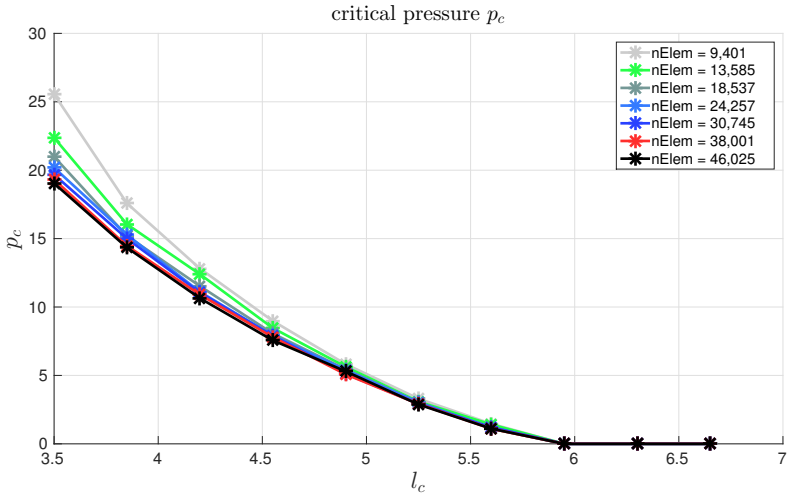


Figure 3: Critical pressures during the crack propagations.

As expected the pressure which is needed for the propagation decreases with an increasing crack length and becomes zero if a critical crack length is reached, which means that the crack propagates only by the external loading. The results also show a convergence to a fixed value.

The resulting crack path is illustrated in Fig. 2(b). At the beginning the crack propagates almost straight as the internal loading is dominant. However, at the end very low or zero pressures already leads to a propagation, wherefore the influence of the external loading increases which leads to a deviation of the straight crack path.

4.2 Different pressure distributions

In this example the influence of different pressure distributions are investigated. Herein, the plate of Section 4.1 is used again. The focus is based on the influence of the pressure distribution, wherefore only a *static* crack configuration without any external loadings is investigated. Three different distributions are considered which only depend on the distance r to the crack tip and the initial pressure p_0 as shown in Fig. 4.

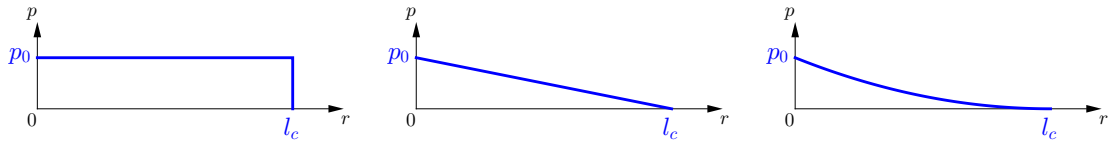


Figure 4: Different pressure distributions: (a) Constant, (b) linear and (c) quadratic.

It is noted that the introduced method is not limited on these distributions, however they allow an easy introduction into the complex topic of hydraulic fracturing. These distributions are used as internal loading of the rectangular plate which was introduced in Section each distribution on different meshes. The results are illustrated in Fig. 5 where the relation of the resulting pressure force R_c and the crack length l_c is plotted over the element number. R_c is illustrated instead of p_0 to take account of the fact that linear or quadratic distributions have a lower resulting force. The results show as expected that concentrations of the pressure at the crack tip lead to a higher

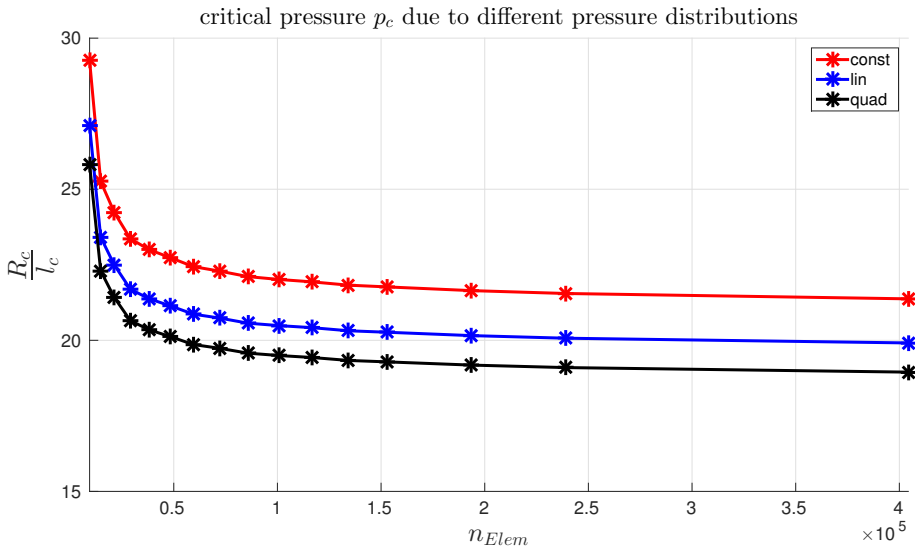


Figure 5: Critical pressures for different pressure distributions.

loading of the fracture, wherefore a lower pressure is needed for the propagation. All results of the considered pressure distributions converge similar.

5 Conclusion

In the presented work, a simplified hydraulic fracturing model by means of pre-defined pressure distributions instead of using the Reynolds equation is investigated. These distributions are scaled by only a few parameters, which results in a much simpler and robust coupled problem. The influence of the distributions is investigated in a numerical study. Critical pressure values are evaluated based on the superposition principle of linear elastic fracture mechanics, where the influence of the external and internal loadings is observed separately and scaled by a factor. This critical pressure leads to a propagation of the crack in direction of the maximum circumferential stresses.

The deformations of the solid are solved by an XFEM simulation with a hybrid explicit-implicit crack description which allows a simple update of the crack geometry during a propagation. Based on the deformations, particularly crack opening displacements (CODs), SIFs and energy release rates are computed by a comparison of the approximated and reference states. This method provides an intuitive, computationally cheap and robust method, which works in two dimensions as well as in three dimensions in a consistent manner.

Numerical results show the relation of the critical pressure value and the crack length. The influence of the pressure distributions is investigated in an own example. As all considered examples are in good agreement with the expected solutions and converge to a fixed value it is concluded that the proposed simplified hydraulic fracturing model provides a simple and robust basis for further investigations.

REFERENCES

[1] J. Adachi, E. Siebrits, A. Peirce, and J. Desroches, Computer simulation of hydraulic fractures. *Rock Mechanics & Mining Sciences*, **44**, 739–757, 2007.

[2] A. Peirce, Modeling multi-scale processes in hydraulic fracture propagation using the implicit level set algorithm. *Comput. Methods Appl. Mech. Engrg.*, **283**, 881–908, 2015.

[3] L.J. Gray, A.-V. Phan, G.H. Paulino, and T. Kaplan, Improved quarter-point crack tip element. *Eng. Fract. Mech.*, **70**, 269–283, 2003.

[4] G.P. Nikishkov, Accuracy of Quarter-point Element in Modeling Crack-tip Fields. *CMES-Computer Modeling In Engineering & Sciences*, **93**, 335–361, 2013.

[5] M. Nejati, A. Paluszny, and R.W. Zimmerman, On the use of quarter-point tetrahedral finite elements in linear elastic fracture mechanics. *Eng. Fract. Mech.*, **144**, 194–221, 2015.

[6] N. Moës, J. Dolbow, and T. Belytschko, A finite element method for crack growth without remeshing. *Internat. J. Numer. Methods Engrg.*, **46**, 131–150, 1999.

[7] M. Stolarska, D.L. Chopp, N. Moës, and T. Belytschko, Modelling crack growth by level sets in the extended finite element method. *Internat. J. Numer. Methods Engrg.*, **51**, 943–960, 2001.

[8] A. Gravouil, N. Moës, and T. Belytschko, Non-planar 3D crack growth by the extended finite element and level sets - Part II: Level set update. *Internat. J. Numer. Methods Engrg.*, **53**, 2569–2586, 2002.

[9] T.P. Fries, M. Baydoun, Crack propagation with the extended finite element method and a hybrid explicit-implicit crack description. *Internat. J. Numer. Methods Engrg.*, **89**, 1527–1558, 2012.

[10] J.R. Rice, A Path Independent Integral and the Approximate Analysis of Strain Concentration by Notches and Cracks. *J. Appl. Mech.*, **35**, 379–386, 1968.

[11] M.F. Kanninen, C.H. Popelar, Advanced fracture mechanics. *Oxford University Press*, 1985.

[12] N. Moës, A. Gravouil, and T. Belytschko, Non-planar 3D crack growth by the extended finite element and level sets - Part I: Mechanical model. *Internat. J. Numer. Methods Engrg.*, **53**, 2549–2568, 2002.

[13] T.P. Fries, A corrected XFEM approximation without problems in blending elements. *Internat. J. Numer. Methods Engrg.*, **75**, 503–532, 2008.

[14] T.P. Fries, T. Belytschko, The extended/generalized finite element method:An overview of the method and its applications. *Internat. J. Numer. Methods Engrg.*, **84**, 253–304, 2010.

[15] T.P. Fries, S. Omerović, Higher-order accurate integration of implicit geometries. *Internat. J. Numer. Methods Engrg.*, early view.

[16] T.L Anderson, Fracture mechanics: fundamentals and applications. *CRC Press*, 2005.

[17] H. Awaji, T. Kato, S. Honda, and T. Nishikawa, Criterion for combined mode I-II of brittle fracture. *J. Ceram. Soc. Japan*, **107**, 918–924, 1999.

[18] D. Haboussa, D. Grégoire, T. Elguedj, H. Maigre, and A. Combescure, X-FEM analysis of the effects of holes or other cracks on dynamic crack propagations. *Internat. J. Numer. Methods Engrg.*, **86**, 618–636, 2011.

[19] N. Moës, T. Belytschko, Extended finite element method for cohesive crack growth. *Eng. Fract. Mech.*, **69**, 813–833, 2002.

[20] A. Peirce, E. Detournay, An implicit level set method for modeling hydraulically driven fractures. *Comp. Methods Appl. Mech. Engrg.*, **197**, 2858–2885, 2008.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SENSITIVITY ANALYSIS OF HYDRAULIC FRACTURING USING AN EXTENDED FINITE ELEMENT METHOD FOR THE PKN MODEL

Hasini Garikapati[1], **Clemens V. Verhoosel**[1], **E. Harald van Brummelen**[1], **Pedro Diez** [2]

[1]Multi-scale Engineering Fluid Dynamics
Eindhoven University of Technology,Eindhoven, The Netherlands
e-mail: {h.garikapati, c.v.verhoosel, e.h.v.brummelen}@tue.nl

[2] Laboratori de Calcul Numric
Universitat Politecnica de Catalunya, Barcelona, Spain
e-mail: pedro.diez@upc.edu

Keywords: Hydraulic fracturing, PKN model, Extended finite element method, Sensitivity analysis, Uncertainty quantification

Abstract. *Hydraulic fracturing is a process that is surrounded by uncertainty, as available data on* e.g. *rock formations is scant and available models are still rudimentary. In this contribution sensitivity analysis is carried out as first step in studying the uncertainties in the model. This is done to assess which of the parameters are key drivers to the model's results. As a baseline model for hydraulic fracturing, the classical PKN model of hydraulic fracturing by Perkins and Kern (J. Pet. Tech. Trans. AIME, 222:937949 (1961)) and Nordgren (J. Pet. Tech. 253:306314 (1972)) which is widely used in the Oil and Gas industry to assist in the design of the hydraulic fracturing treatment is considered. The problem under consideration is characterized by a moving boundary, strong non-linearities and a singularity at the moving tip due to vanishing of the fracture aperture. Sufficient accuracy for such problems is required to obtain a high quality parametric sensitivity analysis and uncertainty quantification. In order to achieve this, various advanced numerical methods are studied. An XFEM approach is adopted to simulate the fracture initiation, propagation and opening of the fracture profile. Using this method, the sensitivity analysis of parameters such as fracturing fluid viscosity, fracture geometry characteristics, soil or rock formations properties via plain strain modulus and leak off coefficient (of the soil) are presented. The impact of these input parameters with respect to the output parameters such as the fracture length and width are discussed.*

1 INTRODUCTION

Hydraulic fracturing is the process of fracturing rock formations by a pressurized liquid to release hydrocarbon resources. This technique is essential for accessing vast new unconventional natural hydrocarbon reserves, which are otherwise inaccessible. Hydraulic fracturing processes are surrounded by uncertainty, as available data on *e.g.* rock formations is limited. The development of models and simulation tools to quantify the risks associated with hydraulic fracturing processes is indispensable. In this contribution, we study the influence of various sources of uncertainty on the hydraulic fracturing process, including characteristics of the fracturing fluid, rock properties, and leakoff parameters. The parameteric sensitivity analysis presented herein is an important first step in the development of an uncertainty quantification framework for hydraulic fracturing.

In order to perform a sensitivity analysis, it is preferable to use a deterministic model which is capable of mimicking the essential aspects of the process, yet remains sufficiently simple to allow for efficient numerical evaluation. In general, the ability to quickly evaluate the deterministic model is critical for many stochastic methods (*e.g.* Monte-Carlo simulations). Herein we consider the PKN model (Perkins-Kern-Nordgren) [1], [4] for hydraulically-stimulated vertical fractures, which computes the fracture aperture and length – two key quantities of interest in hydraulic fracturing– for a fixed fracture height (Figure 1). The PKN model is applicable when the vertical hydraulic fracture is bound within a horizontal permeable layer, and when its length is much larger than the height [2].

Although studied extensively, consideration of the PKN model in an uncertainty quantification framework remains an open research challenge. Attaining sufficiently accurate numerical results to perform a reliable sensitivity analysis (or, more generally, uncertainty quantification) is challenging due to the nature of the problem, which involves strong non-linearities and a moving boundary that degenerates at the tip [3]. The prevailing numerical techniques for the PKN model, which are based on explicit finite difference formulations [6], do not meet the accuracy requirements at acceptable computational costs. For this reason we employ the eXtended Finite Element Method (XFEM) to discretize the model. This numerical technique enables us to conduct a quantitative sensitivity study of the impact of various sources of uncertainty in hydraulic fracturing process.

This paper is organized as follows. In Section 2 we present the governing equations for the PKN model. In Section 3 we describe the XFEM discretization. The sensitivity analysis performed using the PKN model is presented in Section 4, where the effect of the various input parameters on the output of the process is studied. Finally, in Section 5 we draw conclusions and discuss future research directions.

2 FORMULATION

The problem setup for the PKN model [4] considered herein is schematically shown in Figure 1, where the coordinate system is chosen such that x-axis coincides with the (horizontal) direction of crack propagation. The shape of the fracture in the yz-plane is assumed to be elliptical, with the axis in vertical direction fixed at the fracture height h_f, and the axis in the horizontal direction equal to the fracture aperture, $w(x,t)$, which is defined on the one-dimensional time-dependent domain $\Omega(t) = [0, l(t)]$. The evolution in time of the fracture tip position, $l(t)$, and fracture aperture, $w(x,t)$, is driven by injection (at $x = 0$ with flow rate i) of an incompressible Newtonian fluid with dynamic viscosity μ. The considered model includes fluid loss effects, governed by the leakoff coefficient c_l, and fracture volume changes [4]. Plane strain behavior

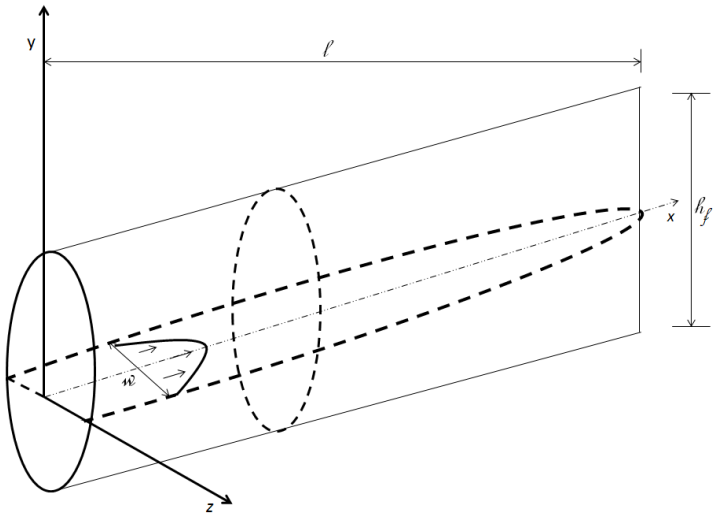


Figure 1: Schematic diagram of PKN fracture

in the vertical direction is assumed with plane strain modulus $E' = E/(1-\nu^2)$, where E and ν denote the Young's modulus and Poisson's ratio of the rock formation, respectively.

The PKN model is based on the conservation of mass of the fracturing fluid,

$$-\frac{\partial q}{\partial x} = C_L + \frac{\partial A_c}{\partial t}, \tag{1}$$

where $q(x,t)$ is the volume rate of flow through the fracture cross-section, C_L is the rate of fluid loss per unit length of the fracture, and $A_c = \frac{1}{4}\pi h_f w$ is the cross-sectionional area of the fracture. The fracture is assumed to be in an ellipitical shape [1]. And for the laminar flow of a Newtonian (fracturing) fluid, the flow rate is related to the pressure gradient by

$$\frac{\partial p}{\partial x} = \frac{64\mu q}{\pi w^3 h_f} \tag{2}$$

Under the plane strain assumption the fracture aperture can be related to the pressure by

$$p = \frac{E'}{2h_f} w \tag{3}$$

The fluid leakoff is governed by the empirical Carter model [10],

$$C_L = \frac{2h_f c_l}{\sqrt{t-\tau}} \tag{4}$$

where c_l is the leakoff coefficient and τ is the arrival time of the fracture tip at location x, *i.e.* $\tau(x) = l^{-1}(x)$. Substituting of equation (2), (3) and (4) in the balance equation (1) then yields the nonlinear partial differential equation

$$\frac{\pi E\prime}{4 \times 128\mu}\left(\frac{\partial^2 w^4}{\partial x^2}\right) = \frac{2h_f c_l}{\sqrt{t-\tau}} + \frac{\pi h_f}{4}\left(\frac{\partial w}{\partial t}\right) \tag{5}$$

subject to the following intitial and boundary conditions:

- Initial condition:
$$w(x,0) = q(x,0) = 0 \tag{6}$$

- The inlet condition – The fluid flow rate i at well-bore ($x = 0$) is constant in time and can, in view of (2) and (3), be expressed as
$$\frac{\partial w^4}{\partial x}(0,t) = -\frac{512\mu}{\pi E\prime} i. \tag{7}$$

- Tip condition – The conditions at the crack tip at $x = l(t)$ is
$$w(l(t),t) = 0. \tag{8}$$

The evolution of the domain $\Omega(t) = [x, l(t)]$ follows from [5]

$$\frac{\pi E\prime}{512\mu}\frac{\partial w^4}{\partial x} + \frac{\pi}{4}u_f h_f w + 2u_f h_f S_p = 0 \quad \text{at} \quad x = l, \tag{9}$$

where the tip propagation velocity, u_f, is the growth rate of the fracture length, i.e. $u_f = \dot{l}$. This equation conveys that the flow rate at the tip equals the sum of volume growth rate at the tip plus the rate of the spurt due to creating new surface. After rearrangement the tip velocity can be written as:

$$u_f = -\frac{\pi E\prime w^3}{32\mu h_f(\pi w + 8S_p)}\frac{\partial w}{\partial x}. \tag{10}$$

In our analysis we consider the special case formulated by Nordgren [4] fow which there are no spurt losses ($S_p = 0$) and the net pressure is zero at the tip (hence the width at the moving fluid front is zero).

We note that the nonlinearity of Equation (9) results in a weak singularity of the fracture aperture at the tip. It is important to note, however, that although the derivative $\frac{\partial w}{\partial x}$ is infinite, the tip velocity is finite by virtue of the fact that w itself goes to zero at the tip. From the perspective of numerical approximation it is essential to appropriately capture this tip singularity, as it governs the evolution of the fracture front. In the next Section 3 we discuss how we employ the eXtended Finite Element Method to adequately capture this singular behavior.

3 NUMERICAL APPROXIMATE METHOD

3.1 XFEM Discretization

Since the introduction of XFEM by Belytschko and Black, it has been employed in various settings to model fracture and propagation of cracks. XFEM can help to overcome imposed challenges and simplify the problems posed by traditional finite elements [7] in material modeling. In our present PKN model, the idea behind using XFEM is to enrich the usual finite element spaces with additional degrees of freedom, which incorporate the physics of the tip behaviour explained in Section 2. To include the singularity at the tip, we define a special enrichment function. The approximation of the fracture width w^h is given by

$$w^h = \sum_{i=1}^{n} N_i(x) w_i + \psi(x)\hat{w} \tag{11}$$

where the first part of the right hand side are the standard FEM shape functions with w_i are the nodal degrees of freedom and the second part represents the enrichment function, most often

based on asymptotic solutions. In case of the PKN model, the asymptotic expansion for crack opening is given by [12]

$$w(t,x) \approx w_0(t)(1-x)^{1/3} \tag{12}$$

The enrichment function used in our case is of the form $\psi = \hat{a}(l-x)^{1/3}$, where l is the fracture length and $\hat{a}$ is a constant. This special function is chosen because it satisfies the properties at the tip. The function becomes zero at the tip, $x = l$ and its derivative goes to minus infinity at $x = l$. When the spurt losses are zero $S_p = 0$, the propagation velocity becomes

$$u_f = -\frac{E\prime}{96\mu h_f}\left(\frac{\partial w^3}{\partial x}\right) \tag{13}$$

Thus, the derivative of the function $\psi^3(x)$ is a finite value. To compute an approximate solution to the strong form of the equations discussed in Section 2, the weak form of the problem with a test function v is given by

$$\int_0^L -\frac{\partial q}{\partial x} v dx = \int_0^L \frac{2h_f c_l}{\sqrt{t-\tau}} v dx + \int_0^L \frac{\pi h_f}{4}\left(\frac{\partial w_0}{\partial t}\right) v dx \qquad \forall v \tag{14}$$

Integrating by parts, results in the weak formulation:

$$iv(0) - L(w,v) = M(w,v)\frac{\partial w}{\partial t} + K(w,v)w \tag{15}$$

where $i\ (= q(0))$ is the inflow rate.

$$\left\{\begin{array}{l} K(w,v) = \int_0^L \frac{\pi w^3 h_f}{128\mu}\left(\frac{\partial w}{\partial x}\right)\left(\frac{\partial v}{\partial x}\right) dx \\ \\ M(w,v) = \frac{\pi h_f}{4}\int_0^L wv dx \\ \\ L(w,v) = \int_0^L \frac{2h_f c_l}{\sqrt{t-\tau}} v dx \end{array}\right. . \tag{16}$$

Note that the function $\tau(x)$ in the leakoff term is interpolated at the corresponding nodes of the mesh at time t for each step. The weak form (14) is discretized using the XFEM discretization (11). This leads to the non-linear system of equations:

$$iv(0) - L(w^h,v^h) = M(w^h,v^h)\frac{\partial w^h}{\partial t} + K(w^h,v^h)w^h \qquad \forall v^h \tag{17}$$

To solve this system of equations, we either use Picard or Newton iterations. For the sake of simplicity, we restrict our discussion to Picard iterations in the next section.

3.2 The incremental-iterative solution procedure

```
# Parameters
Input  : μ, i, h, E', DX, DT,nsteps
# Fracture geometry along with the pressure
Output : (w^1, l^1, p^1),(w^2, l^2, p^2)...(w^nstep, l^nstep, p^nstep)

# (1)Initialization
l^0, w^0

# Time loop starts
for k in range(nsteps):
    # (2)Propogation velocity and fracture length at current time
    u^tip
    l^k = l^{k-1} + u^tip * DT
    # (3) Define new domain, mesh
    domain,   nodes(1, 2, ...j, j + 1, ..)
    funcsp = FEMfuncsp + XFEMfuncsp
    # (4) Interpolate τ, w_j^{k-1} for every point on current mesh
    τ(x) = (l^k)^(-1)
    τ_j^k = Intepolate(τ(x))
    w_j^old = Interpolate(w^{k-1}(x))
    # solve for the width at the current step
    p^k, w^k = nonlinearsolver(w^old, points, geometry)
end

# To solve the resulting non linear equations
def nonlinearsolver(w^old, points, geometry):
    # Define solver parameters, relaxation factor for picard as α
    tol, maxiter, α
    while tol < toliter   and   iter < maxiter:
        iter  +=  1
        n = iter # n subscript denote the values at current iteration
        Kfunc , Mfunc, Lfunc                                  # see equation (15)
        f                                                      # inflow condition
        (w_j)_n = (Kfunc * DT + M).solve(f * DT − L * DT + (M.(w_j^old)))
        # Picard Iteration
        (w_j^k)_n = (w_j^k)_{n−1} + α((w_j^k)_n − (w_j^k)_{n−1})
        # tolerance of iteration
        toliter = f * DT − L * DT + (M.(wold)) − (K * DT + M).(w_n^k)
    end
    p^k = E' w^k/2h
    return p^k, w^k
end
```

Algorithm 1: Psuedo code of XFEM PKN model

The incremental-iterative solution procedure with which we compute the solution to the model presented above is described in Algorithm 1. With this algorithm it has to be noted that:

1. To avoid degeneracy as $t \to 0$, an initial approximation of l^0 and w^0 is considered. For instance, we consider a wedge shaped profile $w^0 = ww(1 - \frac{x}{l^0})$, where ww can be a very

small number.

2. With this initial approximation, we advance to next time steps (k). We compute tip velocity, u^k using equation (13).
3. With the new fracture length (l^k), we define the new domain, nodes and the basis functions as explained in equation (11).
4. We need to compute the values of τ_j, w_j^{k-1} (j representing node number). For this, we interpolate the function $\tau(x) = l^{-1}(x)$ to each node, τ_j. By projecting w^{k-1}, the width from the previous time step, onto the current mesh, we obtain w_i^{k-1}
5. The new w_j^k, p_j^k obtained by solving the nonlinear system of equations (17).

4 NUMERICAL SIMULATIONS

In this section we present the results obtained based on the XFEM formulation outlined in the previous section. We investigate the results using the data used in the comparative study by Warpinski *et al.* [8]. We study mesh convergence of the method. In Section 4.2, we present the sensitivity analysis for the various input parameters of the model.

4.1 Benchmark Result

The input data for our benchmark simulation is from [8] and all parameters are given in Table 1. The Figure 2(a), shows the width profile of the crack at different time steps. It can be observed from the width profile that the crack is widened after every time step. We validate the results with the test cases in [8] and observe that the results to be in good agreement.

c_l	$9.84 * 10^{-6}$ m/$s^{1/2}$
S_p	0 m
h_f	51.8 m
E'	6.13 x 10^{10} Pa
μ	0.2 Pa.s
i	0.0662 m^3/s

Table 1: Input data

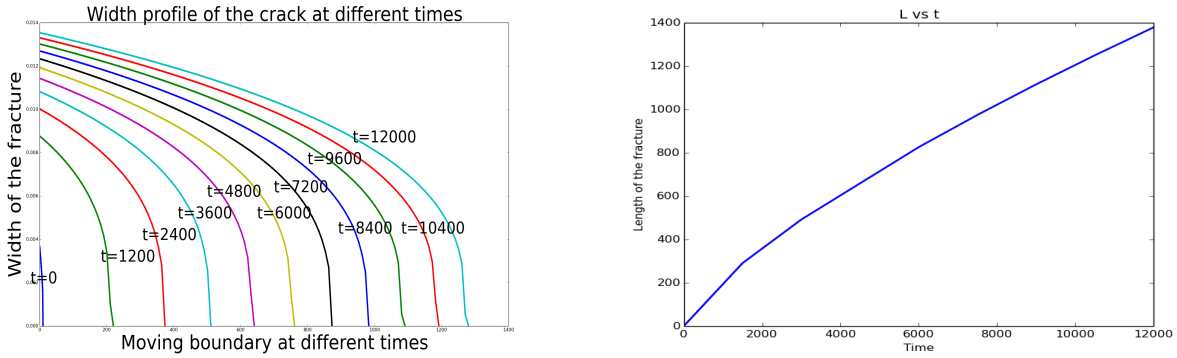


Figure 2: (a) Moving boundary at different times (b) Length of the fracture vs time

Studying mesh independence is important to assess the convergence of the method and to know the mesh size to obtain accurate results. We present the discretization aspects for both space and time steps in Table 2. We perform the simulations for time $T = 100$ seconds. These results are taken into consideration to perform sensitivity studies. Table 2 shows the length and maximum width of the fracture obtained after 100 seconds for corresponding space and time step. We take into consideration the Courant-Friedrichs-Levy condition (in our case, $\frac{u^{tip}\Delta t}{\Delta x}$) to set the limit for Δt. In order to keep the discretization error low, we take $\Delta x \leq \Delta t$. We therefore have, $0.1\Delta x \leq 0.1\Delta t \leq \Delta x$ in the Table 2. We compute the relative error for the length taking the mesh independent value as reference.

Time step (Δt)	Space step (Δx)	Fracture length (m)	Relative error for the length (%)	Maximum width (m)
0.5	0.4	32.805	3.61	0.0054
0.5	0.2	33.151	2.59	0.0054
0.5	0.1	33.264	2.26	0.0054
0.25	0.2	33.303	2.14	0.0054
0.25	0.1	33.626	1.20	0.0054
0.25	0.05	33.712	0.95	0.0054
0.125	0.05	33.034	0.00	0.0054
0.125	0.025	34.034	0.00	0.0054
0.125	0.0125	34.034	0.00	0.0054

Table 2: Step size dependence

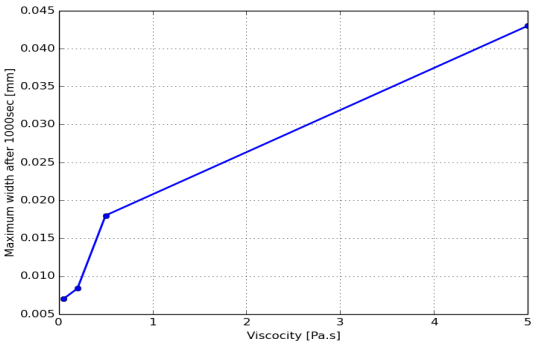
It can be seen that after $\Delta x = 0.05$ and $\Delta t = 0.125$, the results become mesh independent for the length. For the meshes smaller than this the relative error is less than 0.001%. So we consider the step size $\Delta x = 0.05$ and $\Delta t = 0.125$ for the simulations in the next Section 4.2. In case of the width, we observe a change in the value between different step size is less than 10^{-3}. Note that these results are obtained using an adaptive mesh refinement startegy which refines the mesh as we get closer to the tip. These convergence results may vary on a different mesh.

4.2 Sensitivity analysis

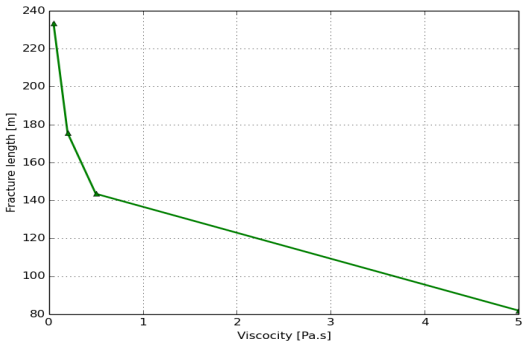
The idea behind this section is to find the potentially important factors that influence a model. This is to identify the input factors that drive most of the variation in the output [9]. We choose independent input factors and perform numerical experiments to identify the impact on the output parameters. This screening exercise is often the first step to apply uncertainty quantification techniques on a model. It is to be noted that this method provides the qualitative sensitivity measure of the input factors, but does not quantify the importance of the factor on the output. So, in this section, the sensitivity of the process with respect to the properties of the fluid and the rock to the fracture geometry and fracturing pressure is analyzed for the PKN model with the formulation discussed in the previous section 3. The effects of fluid viscosity, Young's modulus of the rock, and the fluid leak-off coefficient on fracture geometry are studied. We do not consider the effect on the pressure in this case, as the width of the fracture and pressure are linearly related (3).

4.2.1 Viscosity of the fluid

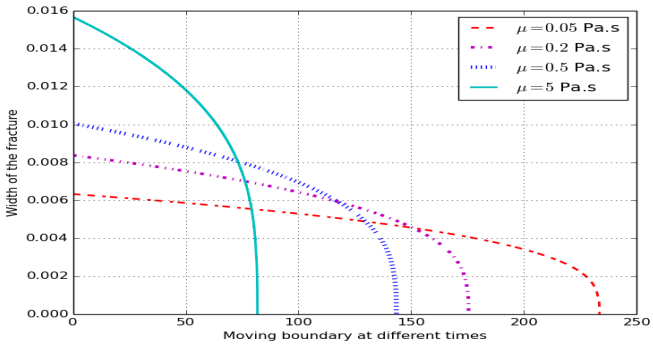
For the numerical experiments, we consider the variation of the viscosity in the range of 0.05 Pa.s to 5 Pa.s and the values of the other parameters are the same as in Table 1. Figure 3 (a) shows that crack width increases with the increase in the fracturing fluid viscosity and Figure 3 (b) shows that the fracture length decreases with the increase in fluid viscosity. We can see from Figure 3 (c) that with a higher viscous fluid, a wider and shorter fracture is produced when compared to a lower viscous fluid which produces narrower and longer fractures.



(a) Fracture maximum width changes with different fluid viscosities



(b) Fracture length changes with different fluid viscosities

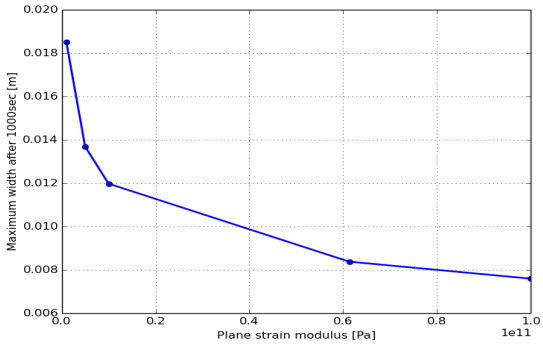


(c) Fracture geometry changes with different fluid viscosities

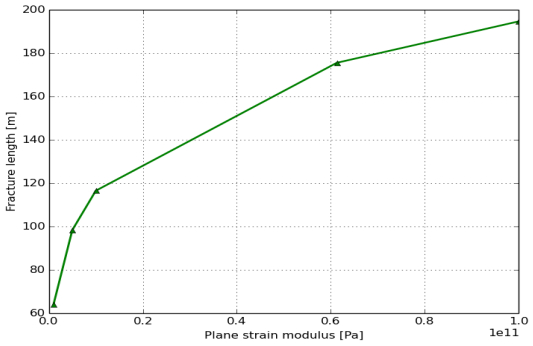
Figure 3: Effects of viscosity on the fracturing geometry

4.2.2 Plane strain modulus

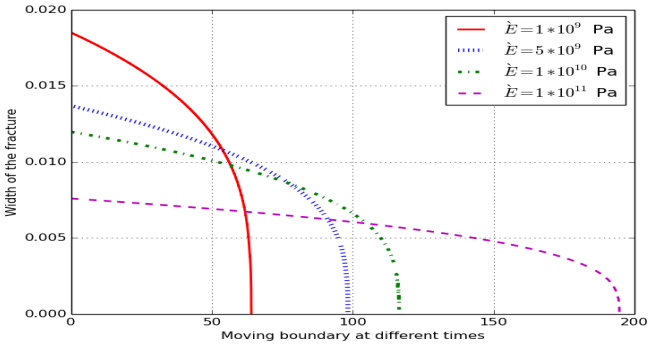
We now consider the effect of plane strain modulus on the fracture geometry, while keeping the fluid and operation conditions unchanged. Inputting the plane strain modulus with values between 10^3 MPa and 10^4 MPa and running these cases, we get the following results. Figure 4 (a), (b), (c) express that a shorter and wider fracture will be generated when the formation is soft with a low plane strain modulus, while a longer and narrower fracture will be produced when the formation is hard with a high plane strain modulus, using the same fracturing fluid and operation conditions. When interpolated to pressure, the results indicate that higher fracturing pressure is required to fracture harder reservoirs which is inline with the experimental results [8].



(a) Fracture maximum width changes with different plane strain moduli



(b) Fracture length changes with different plane strain moduli

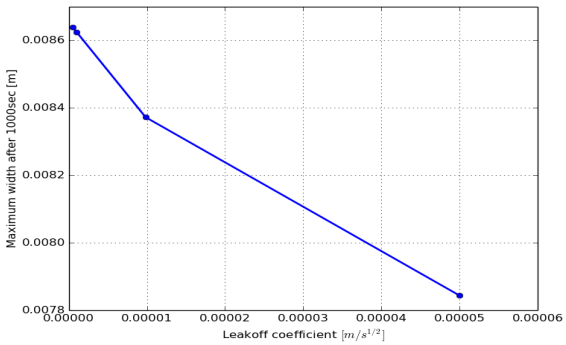


(c) Fracture geometry changes with different plane strain moduli

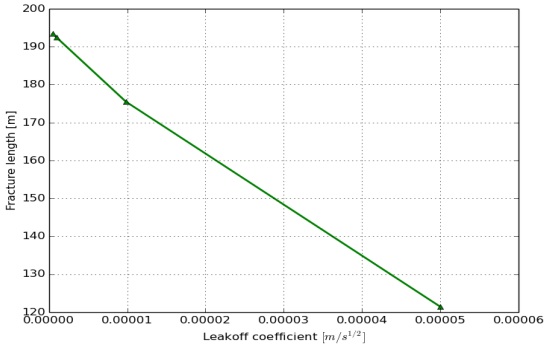
Figure 4: Effects of plane strain modulus on the fracturing geometry

4.2.3 Leakoff coefficient

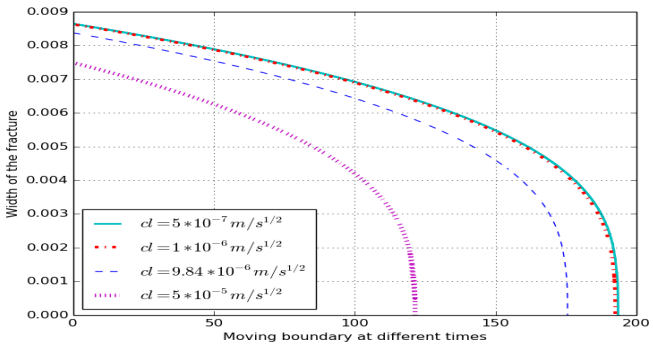
The leakoff coefficient is a parameter which is dependent on both fluid and formation properties. By keeping the other parameters unchanged, the effect of the leakoff coefficient is studied. Figures 5 (a), (b), (c) show that the fracture length and width decrease when the leakoff coefficient is decreased.The leakoff coefficient has a different effect when compared to fluid viscosity and the plane strain modulus. When the leakoff coefficient is higher, smaller and narrower fractures are obtained.



(a) Fracture maximum width changes with different leakoff coefficients



(b) Fracture length changes with different leakoff coefficients

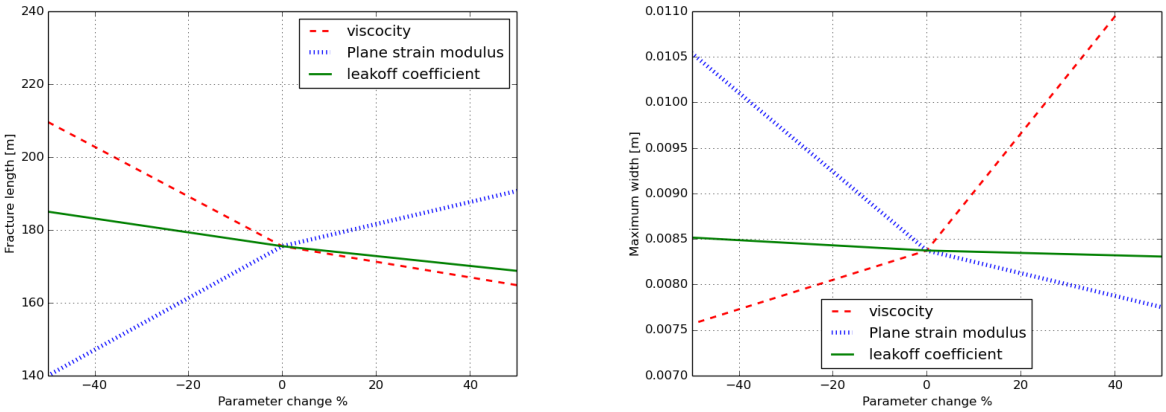


(c) Fracture geometry changes with different leakoff coefficients

Figure 5: Effects of leakoff coefficient on the fracturing geometry

5 CONCLUSIONS AND FUTURE WORK

In the PKN model we considered the three independent input parameters of the model and varied these independently to observe the change in the fracture geometry. The results obtained by the parameter change can be summarized in Figure 6. The parameter change is taken from the base values of Table 1.



(a) length vs parameter change

(b) width vs parameter change

Figure 6: Effects of multiple parameters on fracture geometry

The following conclusions are drawn from the numerical experiments of the sensitivity analysis:

- **Fracture length**: Both viscosity and the leakoff coefficient have the same effect on fracture length when they are changed, yet we notice that fracture length is more sensitive to the viscosity than to the leakoff coefficient. The plane strain modulus has a different effect, as the fracture length increases with an increase of its value.
- **Fracture width**: Width decreases with an increase of leakoff coefficient or plane strain modulus. However, fracture width is less sensitive to the leakoff coefficient than the plane strain modulus. Although the effect of viscosity on the width is opposite when compared to the other parameters, it can be seen that the width is more sensitive to it.

Note that similar conclusions can be drawn from analytical solution or from dimensional analysis. However, our idea is to apply stochastic methods in the framework of hydraulic fracturing which requires high accuracy. To accomplish this, we applied XFEM to the PKN model.

The sensitivity analysis presented herein serves as the basis for the development of an uncertainty quantification framework based on Bayesian inference.

REFERENCES

[1] T.K. Perkins, L.R. Kern, Widths of hydraulic fractures, *Society of Petroleum Engineers*, **13**(9), 937-949, 1961.

[2] P. Valko, M.J. Economides, *Hydraulic fracturing,* 2^{nd} *Edition*. Wiley, 1995.

[3] Y. Kovalyshen, E. Detournay, A Reexamination of the Classical PKN Model of Hydraulic Fracture. *Transport in Porous Media*, **81**, 317-339, 2010.

[4] R. Nordgren, Propagation of Vertical Hydraulic Fractures. *Society of Petroleum Engineers*, **12**(4), 306-314, 1972.

[5] L.F. Kemp, Study of Nordgren's equation of Hydraulic Fracturing. *SPE Production Engineering*, **5**(3), 311-314, 1990.

[6] E. Detournay, A.H.D. Cheng, J.D. McLennan, A Poroelastic PKN Hydraulic Fracture Model Based on an Explicit Moving Mesh Algorithm. *ASME. J. Energy Resour. Technol.*, **112**(4),224-230, 1990.

[7] T. Belytschko, R. Gracie, G. Ventura, A review of extended/generalized finite element methods for material modeling. *Modelling and Simulation in Materials Science and Engineering*, **17**(4), 043001, 2009.

[8] N.R. Warpinki, Z.A. Moschovidis, C.D. Parker and Abou-Sayed, IS.: Comparison Study of Hydraulic Fracturing models: Test Case - GRI-Staged Experiment No. 3. *SPE Production Engineering and Facilities*, **9**(1), 7-16, 1994.

[9] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, 2004.

[10] G.C. Howard, C.R. Fast, Optimum Fluid Characteristics from Fracture Extension, *API Drilling and Production Practice*, **24**, 261-270, 1957.

[11] E. Detournay, Mechanics of Hydraulic Fractures. *Annual Review of Fluid Mechanics*, **48**, 311-339, 2016.

[12] M. Economides, K. Nolte, *Reservoir Stimulation,* 3^{rd} *Edition*. Wiley, 2000.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL MODELLING OF COUPLED HYDRO-THERMAL PROCESSES OF THE SOULTZ HETEROGENEOUS GEOTHERMAL SYSTEM

Musa D. Aliyu[1] and Hua-Peng Chen[2*]

[1,2]Department of Engineering Science, University of Greenwich, Chatham Maritime, Kent, ME4 4TB, UK.

[1]M.D.Aliyu@greenwich.ac.uk; *Corresponding author: H.Chen@greenwich.ac.uk

Keywords: Hot Dry Rock, Geothermal System, Supercritical Carbon Dioxide, Heterogeneous, Simulation.

Abstract. *This paper presents an application of the finite element method (FEM) in modelling coupled, hydro-thermal (HT) processes of the Soultz heterogeneous geothermal system, using different working fluids. A two-dimensional (2-D) model of the geothermal system was developed in the Multiphysics FEM application solver (COMSOL), and the geometry comprises of three wells with five geological layers. Also, two discrete fractures were incorporated in the geometry, each with a 2mm thickness. The model mesh consists of various types of elements, including triangular elements, edge elements and vertex elements. The working fluids used in the study were water (H_2O) and supercritical carbon dioxide ($SCCO_2$), using constant density and viscosity. Moreover, other field variables employed in the work were those available in the literature of the geothermal site considered. A long-term simulation of 30 years was applied, and the temperature, pressure, and enthalpy distribution in the production well were studied. The results obtained showed the superiority of water, concerning energy production, when constant density was used. On the other hand, the pressure of the supercritical carbon dioxide in the production wellbore is far superior to that of water. Therefore, supercritical carbon dioxide can serve as a surrogate to water in the future because of its lower pumping cost in comparison to water. Further studies of other parameters are required to affirm the practicability of the scheme.*

1. INTRODUCTION

The thermal energy generated and stored in the earth is called geothermal; the earth's volume has temperatures greater than (>) 1000°C, with only 0.1% at temperatures less than (<) 100°C [1]. The internal structure of our planet and physical processes occurring there are linked with the origin of this heat [2] and also the decay of naturally occurring radioactive isotopes [3]. In the globe, it is estimated that there is 1.45×10^{26} J of geothermal energy, which is equal to 4.95×10^{6} billion tons of standardised coal [4]. The concept of extracting geothermal energy from hot dry rock (HDR) or engineered geothermal system (EGS) is achieved by force circulating fluid between injection and production wells through a naturally fractured rock mass to create a reservoir by hydraulic fracturing [5, 6]. The in situ stress state in the reservoir will perturb as a result of injection/extraction stimulation, which in turn can lead to fracture initiation/propagation and activation of joints and faults [7]. Also, during reservoir stimulation, the interaction between the working fluids and the host rock may result in mineral dissolution/precipitation in the fractures, faults and the wells. Therefore, engineering design is vital to understanding the response of the fractures, faults, and bedding planes to external stresses in HDR reservoirs [8].

Furthermore, projects on this system to date use either water (H_2O) or brine (NaCl) for the reservoir creation and extraction. These substances when dissolved transports mineral species from the system to some parameters such as permeability, porosity, and the wells. For example, calcite and amorphous silica precipitation have caused obstacles in current operating geothermal system at extraction and injection wells, respectively [9]. Water losses in a geothermal reservoir were also part of the critical issues experienced in Rosemanowes site (UK), European geothermal site (France), and many other projects in the Europe and Asia.

Further to the problems mentioned above experienced using those working fluid in the systems, a novel approach was proposed by Brown [10] to replace H_2O with supercritical carbon dioxide ($SCCO_2$) for both reservoir creation and heat extraction. The study, reported by Pruess [11], reveals that $SCCO_2$ offers larger compressibility and expansivity compared to water. Also using $SCCO_2$ in a closed-loop geothermal reservoir as a substitute for water offers three significant advantages as mentioned by Brown [9] . If $SCCO_2$ is utilised as the circulating fluid in HDR reservoirs, the fluid losses experienced will serve as geological storage of the CO_2. Previous research conducted on this topic were on homogenous hot dry rock (HDR) system, but, however, the systems were heterogeneous in nature. This paper compared the use of $SCCO_2$ and H_2O as working fluid in the heterogeneous geothermal system located at Soultz (France) using the finite element method (FEM).

1.1 Overview of the Soultz Geothermal System

The Soultz site in France was initiated by the European Commission after a detailed surveyed conducted on the three most foremost geothermal projects in Europe that include Rosemanowes, Bad Urach and Soultz to get a commercialised geothermal site within the Europe. A decision was made in 1987 to locate the site in the Soultz, France, due to volcanic activities experienced in the vicinity of a petroleum reservoir site. The project received its initial funding from the European Commission and the relevant energy ministries of France, Germany, and the United Kingdom. Scientist and engineers of these countries established a permanent base on the site to coordinate the activities of various research teams from the participating countries and also to plan the work [12]. Other nations also joined the project later; these countries include Italy, Switzerland and most recently Norway. The United States and Japan

have also made their contributions from the research. The project has undergone four different stages, but only the last phase will be considered here.

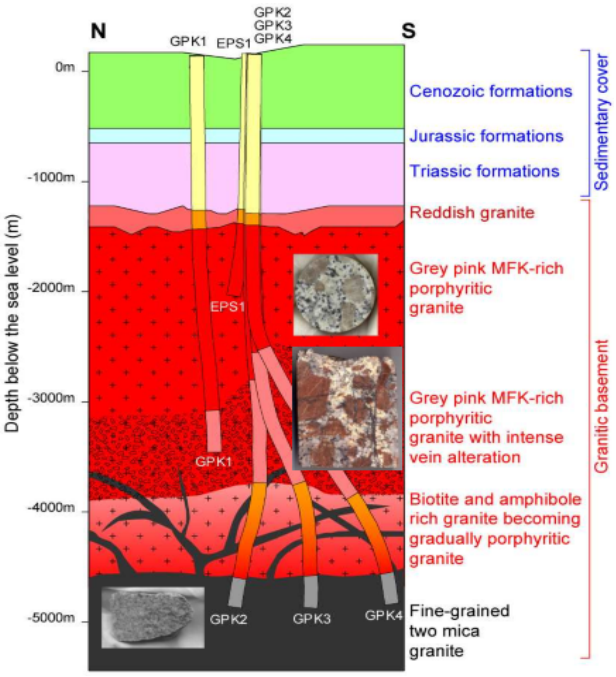


Figure 1: Geological Formation of Geothermal Reservoir at Soultz-Site (France) [13].

Figure 1 represents the geological formation of the entire system and Figure 2 characterises the temperature gradient of the system. The temperature profile is divided into three core zones. The upper zone has a geothermal gradient of 110ºC/Km. The intermediate and the lower zone attribute geothermal gradient of 5ºC/Km and 30ºC/Km, respectively.

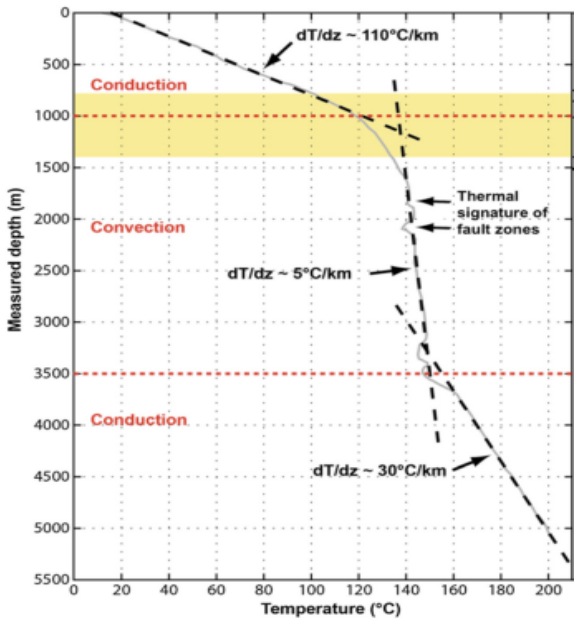


Figure 2: Equilibrium Temperature Profile Obtained from GPK-2 [14].

Moreover, the system wells arrangements are also depicted in Figure 3. GPK3 is the injection well, GPK2 and GPK4 are the production wells. EPS1 and GPK1 are the previous wells used in the earlier stage of the project.

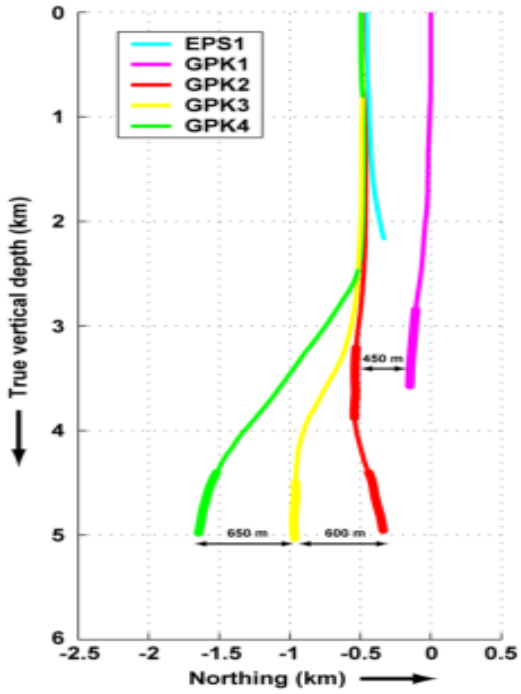


Figure 3: Well Arrangements at Soultz Site (South-North Vertical Cross-Section) [15].

2. MATHEMATICAL AND FINITE ELEMENT FORMULATION OF THE PROBLEM

2.1 Mathematical Formulations

This section gives an exposition of the equations applied in the study. The first set of the equation is based on the law of conservation of mass and is used for the hydraulic process with the assumption that the flow obeys Darcy's law for free movement and is given here as,

$$\rho S \frac{\partial P}{\partial t} + \nabla \cdot \rho \left[\frac{-\kappa}{\mu} (\nabla P + \rho g) \right] = Q_M \tag{1}$$

where ρ is the density, S represents the storativity, P is the pressure, ∇ is the shorthand for the first derivatives with respect to the dimensions of the problem, κ denotes the permeability, μ is the dynamic viscosity, g represents the acceleration of gravity, t is the time and Q_M represent the source term. The second set of the equation is the energy balance equation for heat transport in porous media by considering the Fourier's law of heat flux is expressed as,

$$\rho C_P \frac{\partial T}{\partial t} + \rho C_p v \cdot \nabla T = \nabla \cdot (\lambda \nabla T) + Q \tag{2}$$

where C_p is the specific heat capacity, T represents the temperature, v is the convective term for the velocity, λ denotes the thermal conductivity and Q is the heat source term. The permeability model in the equation (1) can also be replaced by hydraulic conductivity model, and the relationship between the two are:

$$K = \frac{\kappa \rho g}{\mu} \qquad or \qquad \kappa = \frac{K\mu}{\rho g} \tag{3}$$

in which K represents the hydraulic conductivity. The third set of the equation deals with the fracture properties. Equation (4) is the fracture permeability equation govern by the cubic law with the assumption of laminar flow between two parallel plates which is given as,

$$\kappa_{fr} = \frac{a^2}{12} \tag{4}$$

where K_{fr} denotes the fracture permeability, a represents the fracture aperture. Blöcher [16] relates the fracture transmissibility to the aperture and permeability, expressed as,

$$TR_{fr} = \kappa_{fr} a \qquad or \qquad a = \sqrt[3]{12\,TR_{fr}} \tag{5}$$

where TR_{fr} denotes the fracture transmissibility, and by substituting left-hand side of (5) into (4) yields the right-hand side of (5), which is referred as the fracture aperture. The Darcy's and Fourier's flux in equations (1) and (2) can also be expressed as,

$$q_h = -\frac{k}{\mu}\left[\nabla P + \rho g\right] \qquad and \qquad q_T = -\lambda \nabla T \tag{6}$$

2.2 Finite Element Method

This section presents the application of finite element method (FEM) to coupled thermo-hydraulic (TH) problems in discretely fractured porous media. The partial differential equations (PDE) used were already described in the previous section.

2.2.1 Weak Formulation

The Green's theorem and the Method of Weighted Residuals (MWR) were applied to the governing equations provided in the previous section to derive weak formulation of the problem. The weak forms for the mass conservation equation (1) and the energy balance equation (2) are expressed as,

$$\int_{\Omega} w\rho S \frac{\partial P}{\partial t} d\Omega - \int_{\Omega} \rho \nabla w^T \cdot q_h d\Omega + \int_{\Gamma_h^q} w\rho (q_h \cdot n) d\Gamma - \int_{\Omega} w Q_m d\Omega = 0 \tag{7}$$

$$\int_{\Omega} w\rho C_P \frac{\partial T}{\partial t} d\Omega + \int_{\Omega} w\rho C_p v \cdot \nabla T d\Omega - \int_{\Gamma_T^q} w(q_T \cdot n) d\Gamma - \int_{\Omega} w^T Q d\Omega = 0 \tag{8}$$

where Ω and ɾ represents the model domain and boundary, w is the weighting function, n is the normal to the boundary, subscript and superscript T denotes the thermal and transpose, respectively. The boundary conditions are specified for all field functions P and T.

2.2.2 Galerkin FEM Formulations

The weak forms (7) and (8) of the TH balance equations were spatially discretised using the standard Galerkin Method. Primary variables were fluid flow pressure P and temperature T, which were approximated by interpolation functions

$$P^y = N_P P \tag{9}$$

$$T^y = N_T T \tag{10}$$

where P and T were the scalars of the nodal values of the unknowns. N_P and N_T were the shape functions for P and T, respectively. The finite element formulation of the governing equations were given in a matrix form as,

$$M_h^m P'^m + K_h^m P^m - q_h^m = 0 \tag{11}$$

$$M_T^m T'^m + K_T^m T^m - q_T^m = 0 \tag{12}$$

where M and K were process-specific mass and stiffness, respectively. The term q contained the contributions of the coupled processes. The process-specific matrices for the fluid flow and thermal are given in equations (13), (14), (15), (16), (17), and (18) and same can be written as,

$$M_h^m = \int_\Omega N_P^T \rho S N_P d\Omega \tag{13}$$

$$K_h^m = \int_\Omega \nabla N_P^T \frac{\kappa}{\mu} \nabla N_P d\Omega \tag{14}$$

$$q_h^m = -K_h^m \rho g - \int_{\Gamma_q} N_P^T q d\Gamma \tag{15}$$

$$M_T^m = \int_\Omega N_T^T C_P \rho N_T d\Omega \tag{16}$$

$$K_T^m = \int_\Omega N_T^T C_P \rho v \cdot \nabla N_T d\Omega + \int_\Omega \nabla N_T^T \lambda \nabla N_T d\Omega \tag{17}$$

$$q_T^m = \int_\Gamma N_T^T q_T \cdot n d\Gamma + \int_\Omega N_T^T Q d\Omega \tag{18}$$

3. PROBLEM DESCRIPTION AND IMPLEMENTATION

3.1 Geometry, Mesh and System Properties

A two-dimensional model of the geothermal system is developed based on the above FEM procedures. The model geometry is depicted in Figure 4; it is comprised of three wells with five geological layers as discussed earlier in the previous section. Also, two discrete fractures are inputted in the geometry with 2 mm thickness each. The model mesh is shown in Figure 5, consisting of 1,973 triangular elements, 409 edge elements, and 48 vertex elements. Furthermore, geological properties and densities of the system are also presented in Table 1 [17]. Other field variables applied in the model are provided in Table 2 as determined by [18]. Other parameters used are given in the literature of Brown [10] and Pruess [11].

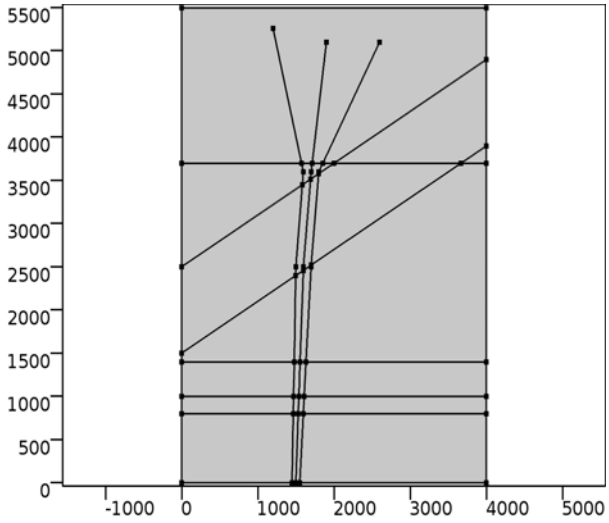


Figure 4: Model Geometry

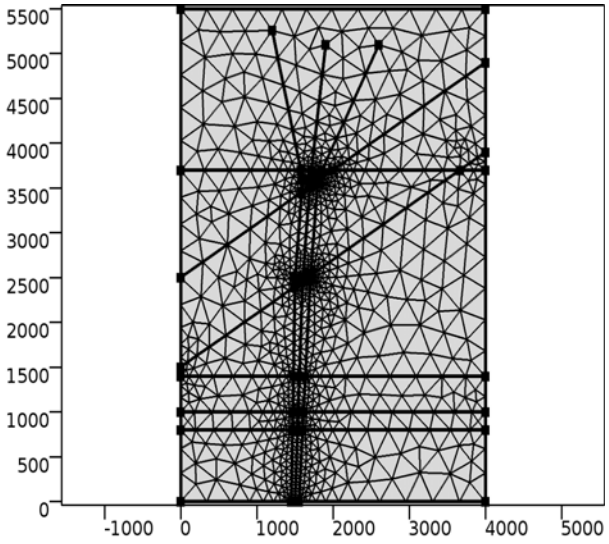


Figure 5: Model Mesh

Formation	Depth (km)	Density (kg/m^3)
Tertiary	0 - 0.75	2350
Jurassic	0 - 0.75	2550
Keuper	0.75 - 0.85	2700
Muschelkalk	0.85 - 1.0	2700
Bunstsandstein	1.0 - 1.4	2500
Basement	1.4 - 5.5	2600

Table 1: Geological Properties and Densities at Soultz (France)

Depth (km)	Permeability (m^2)	Porosity (%)	Thermal Conductivity (W/m/K)
0 - 0.8	10^{-17}	15	1.4
0.8 - 1.0	10^{-16}	15	2.1
1.0 - 1.4	5×10^{-15} - 10^{-14}	15	2.5
1.4 - 3.7	3×10^{-15}	9	3.0
3.7 - 5.0	10^{-18}	1	3.0
Faults	10^{-17} - 3×10^{-14}	15	2.5

Table 2: The Petro-Physical Properties of Main Lithological Formations at Soultz EGS.

3.2 Initial and Boundary Conditions

The initial temperature of the system is given as,

$$T_0 = T_{surf} - b(-z) \tag{19}$$

where T_0 is the initial temperature, T_{surf} denotes the surface temperature, and a value of 10ºC is applied in this study, b is the geothermal gradient, an average gradient of 30ºC/km is used, and z is the depth of the system from the surface. The initial pressure is taken as a constant hydrostatic pressure of 51 MPa. On the other hand, the boundary condition applied is the Dirichlet type (i.e. fixed boundary condition) on well GPK3, in other words, injection well and it corresponds to the temperature and the pressure of 50ºC and 45 MPa respectively.

The simulation was for 30 years, because of the relatively long simulation time and the stability afforded by the constant pressure and temperature conditions. A time-dependent solver BDF (Backward Difference Formula) was employed in COMSOL, and the number of degrees of freedom (DOF) adopted was 5050 (plus 870 internal DOFs). The scheme has an advantage of limiting time step. In this case, it took only 34-time steps to model the 30-year production. The physical computer memory used for the solution is 963 MB and a virtual memory of 1042 MB.

4. RESULTS AND DISCUSSIONS

In this study, a finite element model is developed for the Soultz heterogeneous geothermal system using different working fluids. The first sets of the results presented are the pressure distribution along the geothermal system as shown in Figure 6. Figure 6a presents the pressure distribution in the system after one year of simulation. As can be seen from the illustration, part of the sedimentary layers has a significant drop in the pressure due to the nature of binding between the grains, which leads to an earlier formation of hydraulic fractures within the layers. After additional pumping for 10 years, a total drop in the pressure of the sedimentary layers is observed, as shown in Figure 6b.

Furthermore, Figures 6c and 6d presents the distribution of the pressure for 20 and 30 years of simulation. As seen from both figures the pressure drawdown has less effect on the sedimentary layers due to complete formation of hydraulic fractures in the entire system.

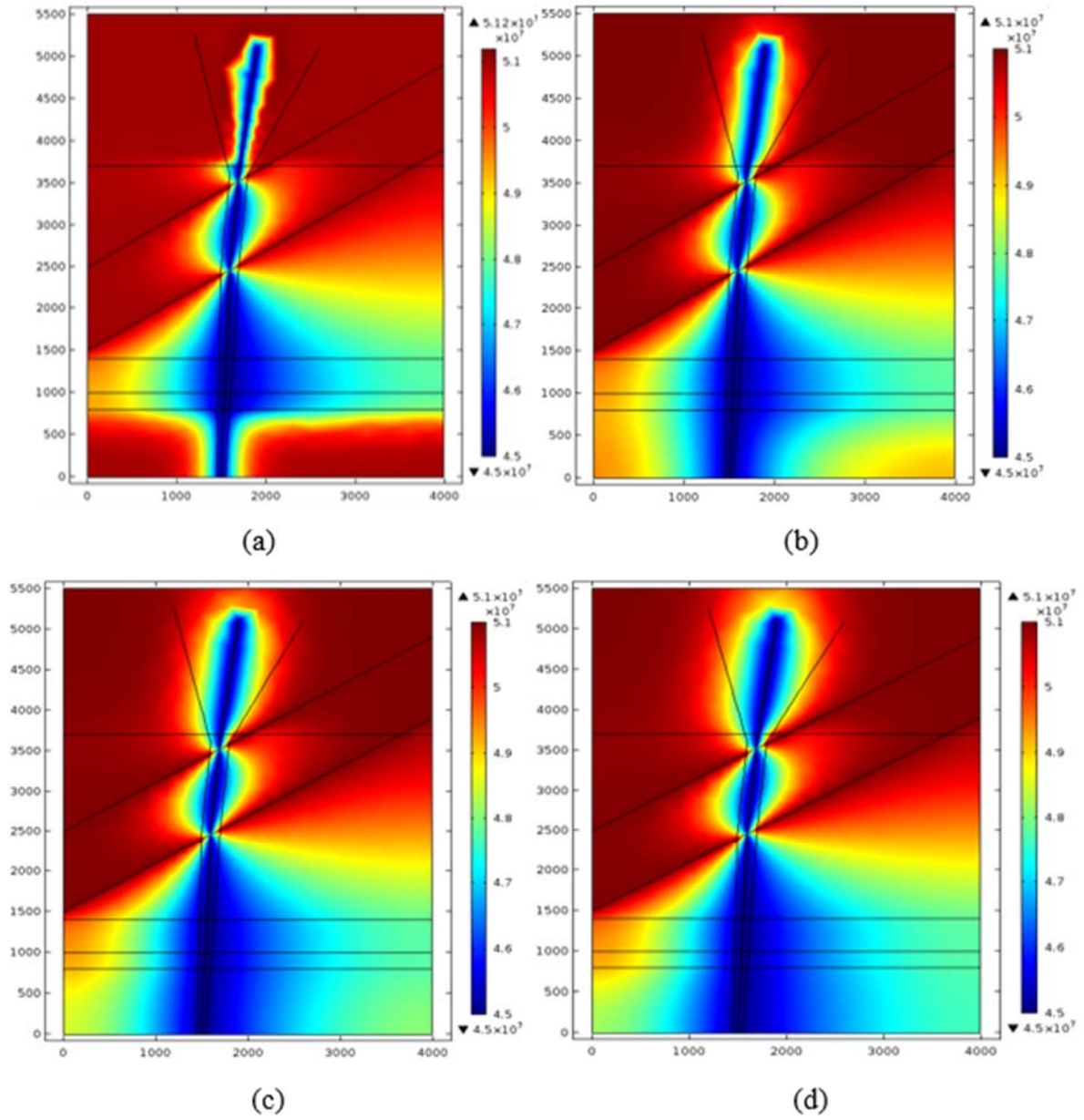


Figure 6: (a) Pressure (MPa) Distribution after 1 Year, (b) Pressure (MPa) Distribution after 10 Years, (c) Pressure (MPa) Distribution after 20 Years, and (d) Pressure (MPa) Distribution after 30 Years

The second set of results analysed in this study includes the production temperature, enthalpy, and pressures at the wellhead of the system. Figure 7 presents the temperature produced at the wellhead, and as can be observed, a drawdown from the initial system temperature of 200°C to 198.8°C is observed after six years of simulation for the $SCCO_2$. On the other hand, a drawdown from 200°C to 198.9°C is observed after seven years of simulation for the H_2O. Also, after 30 years of simulation, the wellhead temperatures for the H_2O at wells, GPK2 and GPK4, are 168.6°C and 165.8°C, respectively. However, in the case of the $SCCO_2$, the wellhead temperatures are 162.9°C and 160°C, respectively.

Figure 8 presents the production pressure of the system under 30 years of simulation. As it can be observed, the pressure produced at the wellhead after one year of simulation is 46

MPa for the H_2O and 50.8 MPa for the $SCCO_2$. Also, the least produced pressure by the $SCCO_2$, at 30 years, is 46.7 MPa, which is greater than the highest produced pressure by the H_2O.

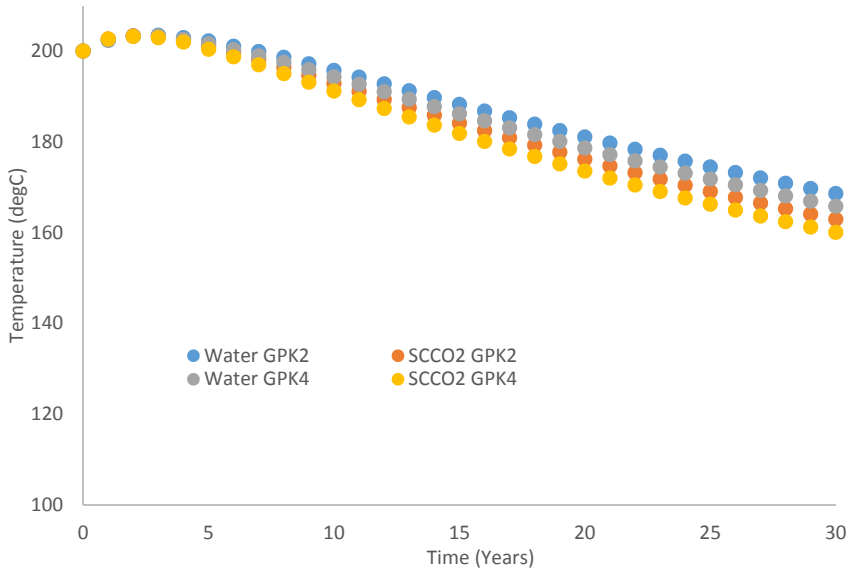


Figure 7: Temperature at the Production Wells

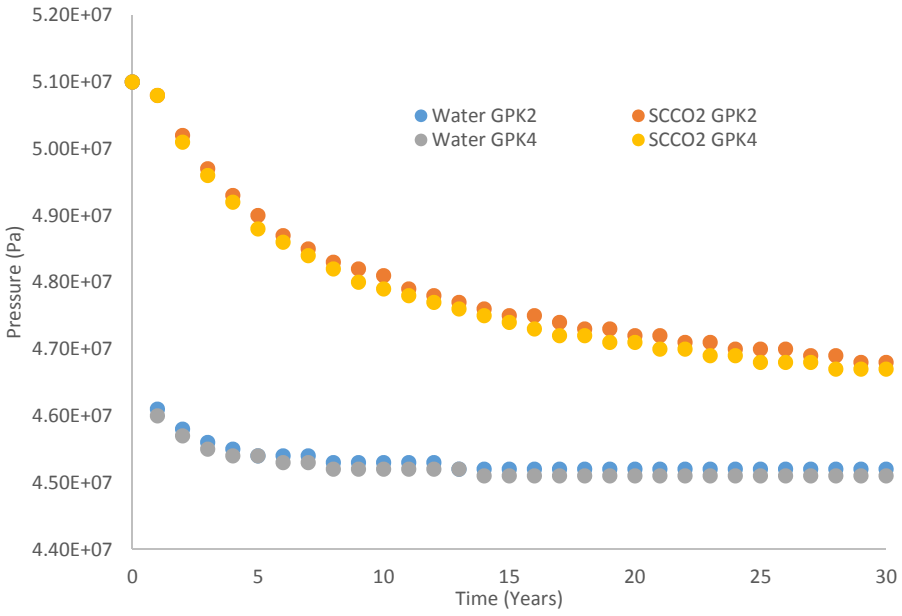


Figure 8: Pressure at the Production Wells

Figure 9 presents the enthalpy produced at the wellhead. As seen, the drawdown pattern is similar to that of the temperature scenario, but the difference in the values obtained for both the working fluids is higher. The maximum produced enthalpy for the H_2O is 748.4 kJ/kg at three years of simulation while that of the $SCCO_2$ is 570.3 kJ/kg at two years of simulation. The least produced enthalpy by the H_2O, which is 590.5 kJ/kg at 30 years, is greater than the maximum produced by the $SCCO_2$. However, the same is observed in the pressure results but

in the reverse order. This occurs because of the relationship between the two parameters, which is expressed as

$$H = U + pV \tag{19}$$

where H, U, p, and V are the total enthalpy, energy of the work done in the system, pressure, and volume of the system, respectively. The above expression, from system thermodynamics, confirms the analysis performed on these parameters to be true.

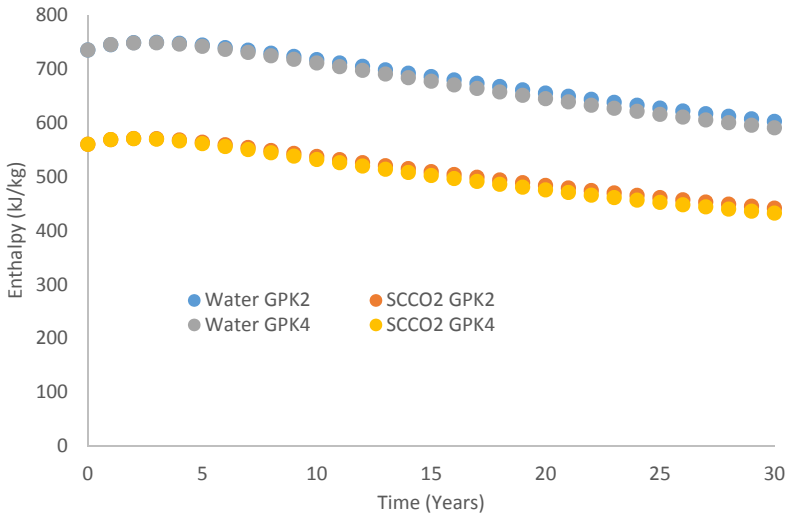


Figure 9: Enthalpy at the Production Wells

5. CONCLUSIONS

A coupled, thermo-hydraulic solution of a heterogeneous, geothermal system is presented, using water and supercritical carbon dioxide as the working fluids. The model takes into account non-isothermal, single-phase flow in porous media connected to discrete fractures. The results of present study show that H_2O is superior to $SCCO_2$, regarding energy extraction (i.e., in cases of temperature and enthalpy), but $SCCO_2$ is produced at a higher pressure at the production wells and reaches larger overall flow rates than H_2O. The difference concerning temperature is of less significance than that of enthalpy, which is quite substantial. On the other hand, the pressure of $SCCO_2$ at the production wells is higher than water because the buoyancy properties of $SCCO_2$ is greater than that of H_2O. The study also affirms the possibility of replacing the $SCCO_2$ with H_2O in the future.

REFERENCES

[1] P. Bayer, L. Rybach, P. Blum, and R. Brauchler, "Review on life cycle environmental effects of geothermal power generation," *Renew. Sustain. Energy Rev.*, vol. 26, pp. 446–463, Oct. 2013.

[2] E. Barbier, "Geothermal energy technology and current status: an overview," *Renew. Sustain. Energy Rev.*, vol. 6, no. 1–2, pp. 3–65, Jan. 2002.

[3] M. Brunner, H. L. Gorhan, and L. Rybach, "GEOTHERMAL ENERGY : Plentiful and Environmentally Friendly," Germany, 2015.

[4] Y. Huang, S. Ren, and D. Zhang, "Current Status and Developing Trend of Direct Use in Global Geothermal Resource," in *Fortieth Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, California, January 26-28, 2015 SGP-TR-204*, 2015, pp. 1–8.

[5] D. W. Brown and D. V Duchane, "Scienti progress on the Fenton Hill HDR project since 1983," *Geothermics*, vol. 28, no. 4, pp. 591–601, 1999.

[6] J. Willis-Richards and T. Wallroth, "Approaches to the modelling of hdr reservoirs: A review," *Geothermics*, vol. 24, no. 3, pp. 307–332, 1995.

[7] A. Ghassemi, S. Tarasovs, and A. H.-D. Cheng, "A 3-D study of the effects of thermomechanical loads on fracture slip in enhanced geothermal reservoirs," *Int. J. Rock Mech. Min. Sci.*, vol. 44, no. 8, pp. 1132–1148, Dec. 2007.

[8] J. Taron and D. Elsworth, "Coupled mechanical and chemical processes in engineered geothermal reservoirs with dynamic permeability," *Int. J. Rock Mech. Min. Sci.*, vol. 47, no. 8, pp. 1339–1348, Dec. 2010.

[9] J. Taron and D. Elsworth, "Thermal–hydrologic–mechanical–chemical processes in the evolution of engineered geothermal reservoirs," *Int. J. Rock Mech. Min. Sci.*, vol. 46, no. 5, pp. 855–864, Jul. 2009.

[10] D. Brown, "a Hot Dry Rock Geothermal Energy Concept Utilizing Supercritical Co2 Instead of Water," *Twenty-Fifth Work. Geotherm. Resevoir Eng.*, vol. 1995, no. April 1992, pp. 1–6, 2000.

[11] K. Pruess, "Enhanced geothermal systems (EGS) using CO2 as working fluid-A novel approach for generating renewable energy with simultaneous sequestration of carbon," *Geothermics*, vol. 35, no. 4, pp. 351–367, 2006.

[12] D. Duchane, "The History of HDR Research and Development," in *Proceedings, 4th International HDR Forum, Strasbourg, France, September 28-30, 1998*, 1999, no. 836.

[13] P. Schlagermann, "Overview of the Soultz geothermal project Overview of the Soultz geothermal project," no. 3. pp. 66–75, 2011.

[14] J. Vidal, A. Genter, and J. Schmittbuhl, "How do permeable fractures in the Triassic sediments of Northern Alsace characterize the top of hydrothermal convective cells? Evidence from Soultz geothermal boreholes (France)," *Geotherm. Energy*, vol. 3, no. 1, 2015.

[15] A. Genter, X. Goerke, J. Graff, N. Cuenot, G. Krall, M. Schindler, and G. Ravier, "Current Status of the EGS Soultz Geothermal Project (France)," *World Geotherm. Congr.*, vol. C, no. April, pp. 25–29, 2010.

[16] M. G. Blöcher, M. Cacace, B. Lewerenz, and G. Zimmermann, "Three dimensional

modelling of fractured and faulted reservoirs: Framework and implementation," *Chemie der Erde - Geochemistry*, vol. 70, no. SUPPL. 3, pp. 145–153, 2010.

[17] P. Baillieux, E. Schill, Y. Abdelfettah, and C. Dezayes, "Possible natural fluid pathways from gravity pseudo-tomography in the geothermal fields of Northern Alsace (Upper Rhine Graben)," *Geotherm. Energy*, vol. 2, no. 1, p. 16, 2014.

[18] L. Guillou-Frottier, C. Clément, B. Bourgine, V. Bouchot, and A. Genter, "Hydrothermal convection beneath an inclined basement-sediment interface : application to the Rhine graben and its Soultz-sous-Forêts temperature anomaly Case of impervious faults Case of permeable faults," in *AGU Fall Meeting, San Franscisco, 3-7 December, 2012*, 2012, p. 1.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A LOW DIFFUSIVE APPROACH FOR TWO PHASE FLOWS

A. Beccantini

CEA Saclay, Université Paris-Saclay,
Laboratoire de Dynamique des Structures, DEN/DM2S/SEMT/DYN,
F-91191 Gif-sur-Yvette Cedex
e-mail: alberto.beccantini@cea.fr

Keywords: Compressible multifluid flows, Discrete Equation Method, Upwind Downwind-Controlled Splitting, Anti-diffusive scheme, Tangent of hyperbola interface capturing

Abstract. *We present a work realized in the framework of the R&D program on the Hydrogen combustion of the I3P Project "accidents graves parc actuel". It is a sequel to recent publications (Tang et al., 2014) where high-order and low diffusive approaches (e.g. a multi-dimensional anti-diffusive approach) and an original Upwind Downwind-Controlled Splitting method (UDCS) were combined with the 1D and multi-dimensional formulations of DEM (Discrete Equation Method, Abgrall and Saurel 2003) and RDEM (Reactive Discrete Equation Method, Le Métayer et al., 2005).*

On one hand, the UDCS limited second-order method has already been combined with the explicit two-step Runge-Kutta scheme for time discretization and represents a significant improvement of the previous limited second-order RDEM approach in terms of robustness (it is now implemented in the fast dynamic fluid-structure interaction code EUROPLEXUS to compute reacting flows in 3D large geometries like the containment of a nuclear power plant).

On the other hand, as observed on the final 2D combustion problem presented in (Tang et al., 2014), it was necessary to improve the approach used to compute the normal vector at the flame interface so as to avoid that the computation of this normal spoils the accuracy improvement brought by the anti-diffusive UDCS/RDEM.

In this work a modification of the anti-diffusive approach has been realized using a hyperbolic tangent approximation of the phase function (or color function) to correctly compute its gradient and then normals to the flame interface. On one hand, numerical experiments have shown that this approach is more accurate than the limited second-order method. On the other hand, the problem of flame wrinkling (which also affects the limited second-order method) is not completely solved (as expected since numerical diffusion of the new approach is less important than in the limited second-order method), but it can be reduced if gradients are computed not directly to the phase function but on a field obtained by solving a diffusion problem having the phase function as initial condition.

1 Introduction

This work is a sequel to recent publications [1, 2, 3, 4] where high-order and low diffusive approaches (e.g. a multi-dimensional anti-diffusive approach) and an original Upwind Downwind-Controlled Splitting method (UDCS) were combined with the 1D and multi-dimensional formulations of DEM [5] and RDEM [6, 7]. First, the method was successfully developed in 1D for computing inert interfaces (e.g. impermeable water-gas shock tube problem) and flame interfaces (e.g. Chapman-Jouguet deflagration and strong detonation wave) with excellent robustness and accuracy properties. Secondly, it has been successfully extended to the multi-dimensional case for the modeling of multi-fluid flows on unstructured grids.

The limited second-order UDCS method has already been combined with the explicit two-step Runge-Kutta scheme for time discretization. This allow us to work with large enough CFL numbers when using a limited second-order reconstruction on all primitive variables (achieving quasi second-order accuracy in space and time). As shown in [4], this approach represents a significant improvement of the classical second-order RDEM approach in terms of robustness and is now implemented in the fast dynamic fluid-structure interaction code EUROPLEXUS [8] to compute reacting flows in 3D large geometries (e.g. the containment of a nuclear power plant). Details can be found in [9, 10, 11].

As shown in [3, 4], the anti-diffusive UDCS is much more accurate than the limited second-order UDCS. For instance, when computing a 1D flame acceleration with the latter approach, the flame region can become too large, thus weakening the precursor shock wave. On the opposite the anti-diffusive UDCS method keeps the flame region over one or two cells, thus avoiding the weakening of the precursor shock wave; however, in multi-dimensional computation, since the flame region is crisped, it is impossible to define a normal to the flame surface. For this reason we propose here a new approach, in which the the flame region is diffused over a region of fixed width; this is obtained by using the properties of the hyperbolic tangent function (tanh) to decide whether to add numerical diffusion to the antidiffusive approach. We point out that the idea of using an hyperbolic tangent function to represent a color function (or a phase function) has already been used in the THINC (tangent of hyperbola interface capturing) scheme of Xiao and coworkers, presented for the first time in [12]; some improvements on Cartesian grids are proposed in [13, 14, 15]; and extensions to unstructured meshes can be found in [16, 17]. However our approach, which is also devoted for computation in unstructured meshes, is quite different and does not require the reconstruction of tahn: the smoothing of the color function is obtained by adding numerical diffusion to the antidiffusive approach when necessary.

2 Numerical approach

This section is organized as follows. First of all we briefly introduce the hyperbolic tangent (Section 2.1). Since the smoothing of the color function is obtained by the modification of the anti-diffusive approach, for the sake of simplicity we first present the anti-diffusive approach of Lagoutière [18] and its tanh modification in 1D, in the particular case of monotone color function (respectively in Section 2.2 and Section 2.3). Then we present the 1D general case in Section 2.4. Finally Section 2.5 is devoted to the multi-dimensional case.

2.1 The hyperbolic tangent and the smoothed color function

The hyperbolic tangent $\tanh(x/l)$, with

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{\frac{1}{2}(\exp(x) - \exp(-x))}{\frac{1}{2}(\exp(x) + \exp(-x))},$$

can be considered as a smoothed Heaviside function varying from -1 to 1. As shown in Figure 1, it coincides with the Heaviside function as $l \to 0$, l being a length scale which determines the smoothness of the function.

A smoothed color function α, $\alpha_{\min} < \alpha < \alpha_{\max}$ can be described by the function

$$\begin{aligned} \alpha(x, l, \alpha_{\min}, \alpha_{\max}) &= \frac{1}{2}\alpha_{\min}\left(1 - \tanh\left(\frac{x}{l}\right)\right) + \frac{1}{2}\alpha_{\max}\left(1 + \tanh\left(\frac{x}{l}\right)\right) \\ &= \frac{\alpha_{\min} + \alpha_{\max}}{2} + \frac{\alpha_{\max} - \alpha_{\min}}{2}\tanh\left(\frac{x}{l}\right) \end{aligned} \tag{1}$$

The reciprocal function of the color function $x(\phi)$ is

$$x(\phi) = l\ \tanh^{-1}\left(\frac{\phi - \frac{\alpha_{\min}+\alpha_{\max}}{2}}{\frac{\alpha_{\max}-\alpha_{\min}}{2}}\right), \quad \alpha_{\min} < \phi < \alpha_{\max}.$$

Figure 2 represents the color function α (equation (1)) in the particular case $\alpha_{\min} = 0.1$, $\alpha_{\max} = 0.9$ ($l = 1$ and $l = 0.1$).

2.2 The 1D anti-diffusive approach for a monotone color function

Let us illustrate form a qualitative point of view how the anti-diffusive approach works on the equation

$$\frac{\partial \alpha}{\partial t} + D\frac{\partial \alpha}{\partial x} = 0$$

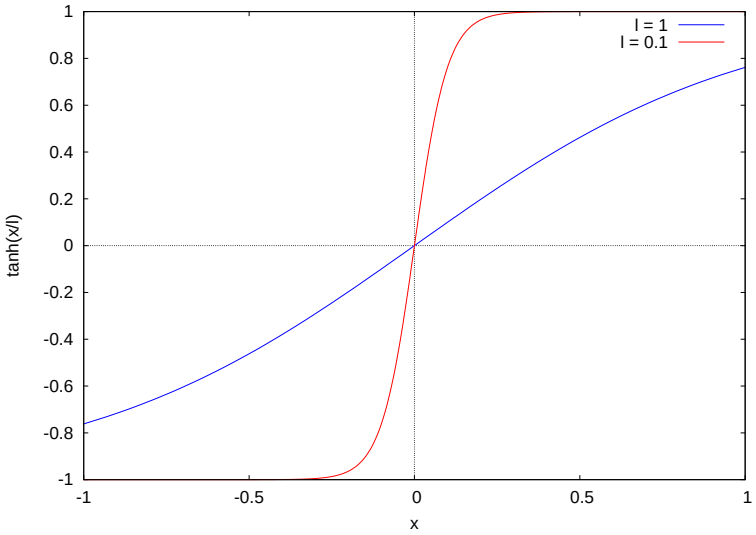


Figure 1: The hyperbolic tangent $\tanh(x/l)$.

with $D > 0$ and α monotone decreasing function

$$\frac{\partial \alpha}{\partial x} \leq 0,$$

(see for instance Figure3). With a conservative finite volume discretisation explicit in time we can write in the i-th cell

$$\frac{\alpha_i^{n+1} - \alpha_i^n}{\Delta t} = \frac{D}{\Delta x} \left(\alpha_{i-1/2}^n - \alpha_{i+1/2}^n \right), \tag{2}$$

α_i being value inside the i-th cell and $\alpha_{i\pm 1/2}$ intercell values (see Figure 4).
According to the monotone condition for α, we suppose $\alpha_{i-1}^n \geq \alpha_i^n \geq \alpha_{i+1}^n$.

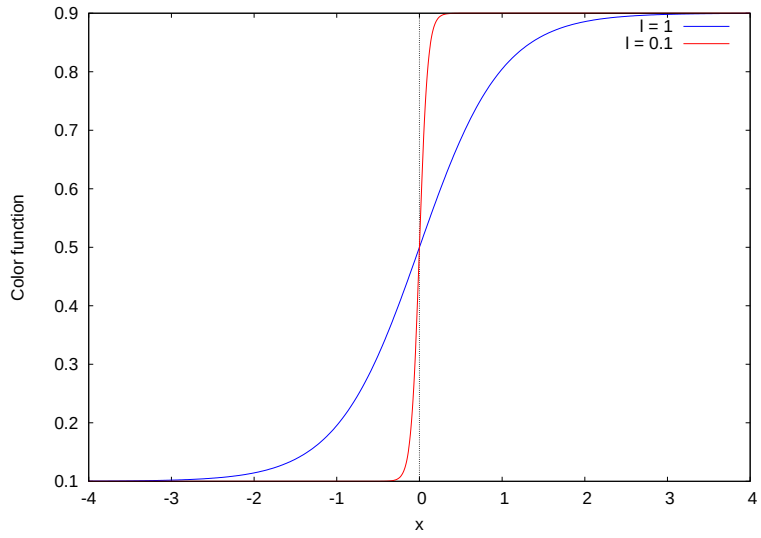


Figure 2: The color function (1) when $\alpha_{\min} = 0.1$, $\alpha_{\max} = 0.9$ ($l = 1$ and $l = 0.1$).

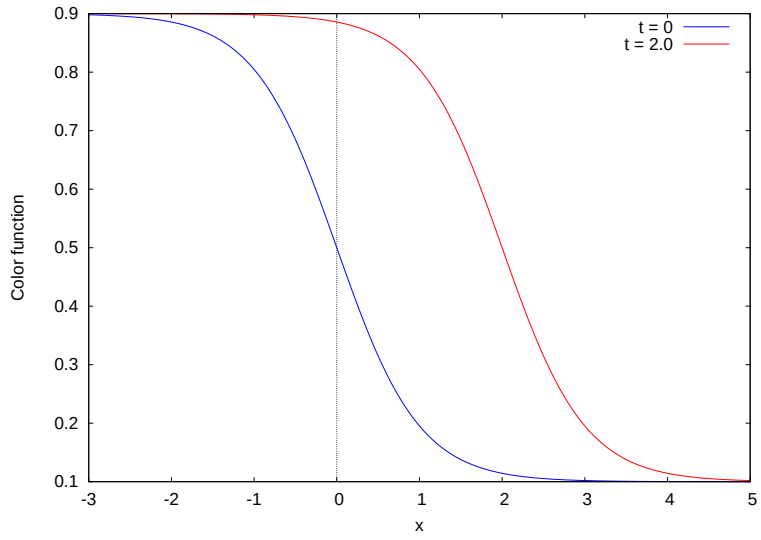


Figure 3: Example of translation of a monotone decreasing color function $\alpha(x)$ with speed 1, namely plot of $\alpha(x - t)$ with $t = 0$ and $t = 2$.

The choice of the intercell value $\alpha_{i+1/2}$ depends on the approach we use.

- $\alpha^n_{i+1/2} = \alpha^n_i$ corresponds to the **upwind scheme** (the speed D being positive), first order accurate and conditionally stable, which diffuses the Heaviside function.
- $\alpha^n_{i+1/2} = \alpha^n_{i+1}$ corresponds to the **downwind scheme**, first order accurate and unconditionally unstable, which "anti-diffuses" the Heaviside function.
- The purpose of the **controlled anti-diffusive scheme** is the combination of both approaches to prevent the diffusion of the Heaviside function. We look for $\alpha^n_{i+1/2}$ such that
$$\alpha^n_{i+1/2} \in (\alpha^n_{i+1}, \alpha^n_i)$$
More in particular we look for the closest value to the downwind one which does not create a new extremum, namely **we look for the minimum value of $\alpha^n_{i+1/2}$ which provides the maximum value of α^{n+1}_i with**
$$\alpha^{n+1}_i \leq M^n_i = \max\left(\alpha^n_i, \alpha^n_{i-1}\right) = \alpha^n_{i-1}. \tag{3}$$

From equation (2) we can write
$$\alpha^n_{i+1/2} = \alpha^n_{i-1/2} - \frac{\Delta x}{D\Delta t}\left(\alpha^{n+1}_i - \alpha^n_i\right).$$
In order to avoid a condition which couples all the intercell faces, we take for $\alpha^n_{i-1/2}$ its most conservative value, the one which maximizes α^{n+1}_i, namely we suppose that $\alpha^n_{i-1/2} = M^n_i$. It follows that condition (3) is satisfied if $\alpha^n_{i+1/2} \geq b^n_i$ with
$$\begin{aligned} b^n_i &= M^n_i - \frac{\Delta x}{D\Delta t}\left(M^n_i - \alpha^n_i\right) \\ &= \alpha^n_i - \left(\frac{\Delta x}{D\Delta t} - 1\right)\left(M^n_i - \alpha^n_i\right) \end{aligned} \tag{4}$$
independently from the choice of $\alpha^n_{i-1/2}$ (indeed we have supposed that it assumes its worst value for condition (3)).
b^n_i is lower than the upwind value (α^n_i) providing that a CFL condition is respected. At the end we choose
$$\alpha^n_{i+1/2} = \max(\alpha^n_{i+1}, b^n_i).$$
In this way $\alpha^n_{i+1/2} \in (\alpha^n_{i+1}, \alpha^n_i)$; namely its value is between the upwind and is the closest to the downwind one which does not create new extrema.

From a mathematical point of view one can show that using this scheme the 1D solution of Heaviside function presents zero, one or two intermediate regions at each time step, i.e. as the time increases the solution is not numerically diffused.

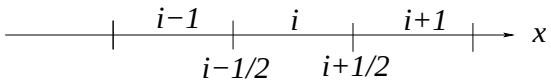


Figure 4: Conservative finite volume discretisation. x axis. Indexes $i-1$, i, $i+1$ refer to cell value. Indexes $i-1/2$, and $i+1/2$ refer to intercell values.

2.3 The tanh modification for a monotone color function

As in Section 2.2, we present the approach in the case

$$\frac{\partial \alpha}{\partial x} \leq 0,$$

and $\alpha_{i-1}^n \geq \alpha_i^n \geq \alpha_{i+1}^n$.
We suppose that in the i-th cell the color function has the shape

$$\alpha_{\tanh}\left(x - x_i\right) = \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{\beta(x - x_i)}{\Delta x_i} + \tilde{x_i}\right)$$

where Δx_i is the cell dimension, β is a parameter which controls the steepness of $\alpha_{\tanh}$ (the characteristic length l is $\Delta x_i/\beta$), $\tilde{x_i}$ is a parameter we have to determine. Instead of using a condition over the integral, as in the THINC approach, we prefer considering a condition over the value. Namely we compute $\tilde{x_i}$ such that

$$\alpha_{\tanh}\left(0\right) = \alpha_i^n$$

i.e.

$$\tilde{x_i} = \tanh^{-1}\left(1 - 2\alpha_i^n\right), \quad 0 < \alpha_i^n < 1.$$

Once determined $\tilde{x_i}$, since β is given, we increase the numerical diffusion of the anti-diffusive approach by imposing a constraint on the downwind-closest value for $\alpha_{i+1/2}^n$, namely we require that

$$\alpha_{i+1/2}^n > \alpha_{\tanh}(\underbrace{x_{i+1/2} - x_i}_{0.5\Delta x}). \tag{5}$$

We define

$$\tilde{\tilde{b_i^n}} = \max(\alpha_{i+1}^n, \alpha_{\tanh}(0.5\Delta x))$$

and we want now that

$$\alpha_{i+1/2}^n \in (\tilde{\tilde{b_i^n}}, \alpha_i^n)$$

which is more restrictive than $\alpha_{i+1/2}^n \in (\alpha_i^{n+1}, \alpha_i^n)$.
Finally, conditions (4) and (5) are satisfied if

$$\alpha_{i+1/2}^n = \max(\tilde{\tilde{b_i^n}}, b_i^n).$$

where b_i^n is defined via equation (4).

The following properties hold.

- Since $\tilde{\tilde{b_i^n}} \geq \alpha_i^{n+1}$, the scheme is more diffusive than the anti-diffusive approach presented in Section 2.2.
- If the tanh restriction is not so important (i.e. $\alpha_{\tanh}(0.5\Delta x) \leq \alpha_i^{n+1}$, which implies $\tilde{\tilde{b_i^n}} = \alpha_i^{n+1}$), the scheme degenerates to the anti-diffusive approach.

2.4 The tanh modification in the 1D general case

First of all, let us present the general formula for the anti-diffusive scheme. We define

$$
\begin{aligned}
m_i^n &= \min\left(\alpha_{i-1}^n, \alpha_i^n\right) \\
M_i^n &= \max\left(\alpha_{i-1}^n, \alpha_i^n\right).
\end{aligned}
$$

We suppose that speeds $D_{i-1/2} \geq 0$, $D_{i+1/2} > 0$.
We want to find intercell boundaries values for α such that

$$
m_i^n \leq \alpha_i^{n+1} \leq M_i^n \tag{6}
$$

$$
m_{i+1}^n \leq \alpha_{i+1/2}^n \leq M_{i+1}^n \tag{7}
$$

Condition (6) is satisfied if

$$
\left\{
\begin{array}{rcl}
M_i^n & \geq & \alpha_i^n \\
& + & \dfrac{D_{i-1/2}\Delta t}{\Delta x_i}(M_i^n - \alpha_i^n) \\
& - & \dfrac{D_{i+1/2}\Delta t}{\Delta x_i}(\alpha_{i+1/2}^n - \alpha_i^n) \\
m_i^n & \leq & \alpha_i^n \\
& + & \dfrac{D_{i-1/2}\Delta t}{\Delta x_i}(m_i^n - \alpha_i^n) \\
& + & \dfrac{D_{i+1/2}\Delta t}{\Delta x_i}(\alpha_i^n - \alpha_{i+1/2}^n)
\end{array}
\right.
$$

or

$$
\left\{
\begin{array}{rcccl}
\alpha_{i+1/2}^n & \geq & b_i^n & = & \alpha_i^n \\
& & & - & \dfrac{\Delta x_i}{D_{i+1/2}\Delta t}(M_i^n - \alpha_i^n) \\
& & & + & \dfrac{D_{i-1/2}}{D_{i+1/2}}(M_i^n - \alpha_i^n) \\
\alpha_{i+1/2}^n & \leq & B_i^n & = & \alpha_i^n \\
& & & + & \dfrac{\Delta x_i}{D_{i+1/2}\Delta t}(\alpha_i^n - m_i^n) \\
& & & - & \dfrac{D_{i-1/2}}{D_{i+1/2}}(\alpha_i^n - m_i^n)
\end{array}
\right. \tag{8}
$$

It is difficult to show that $b_i^n \leq \alpha_i \leq B_i^n$. Then we define

$$
\begin{aligned}
l_i &= \max(m_{i+1}^n, \min(b_i^n, \alpha_i^n))) \\
L_i &= \min(M_{i+1}^n, \max(B_i^n, \alpha_i^n))
\end{aligned}
$$

with b_i^n and B_i^n defined in (8). Since $l_i \leq \alpha_i \leq L_i$, the value of $\alpha_{i+1/2}^n$ defined as

$$\alpha_{i+1/2}^n = \begin{cases} L_i & \text{if } \alpha_{i+1}^n > L_i, \\ \alpha_{i+1}^n & \text{if } L_i \geq \alpha_{i+1}^n \geq l_i, \\ l_i & \text{if } l_i > \alpha_{i+1}^n \end{cases} \tag{9}$$

satisfies conditions (6) and (7).

We introduce numerical diffusion when the downwind value is too strong with respect to the tanh representation of α.

- If $|\alpha_{i-1}^n - \alpha_{i+1}^n| \leq 10^{-3}$ we do not modify formulae.
- If $\alpha_{i-1}^n > \alpha_{i+1}^n$ (negative centered gradient for α), then

$$\alpha_{\tanh}(x - x_i) = \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{\beta(x - x_i)}{\Delta x_i} + \tilde{x}_i\right)$$

$$\tilde{x}_i = \tanh^{-1}(1 - 2\alpha_i^n), \quad 0 < \alpha_i^n < 1.$$

$$\tilde{\tilde{b}}_i^n = \max(\alpha_{i+1}^n, \alpha_{\tanh}(0.5\Delta x_i))$$

We redefine

$$m_{i+1}^n = \min(\alpha_i^n, \tilde{\tilde{b}}_i^n)$$

and then we proceed like in classical downwind approach.

$$\begin{aligned} l_i &= \max(m_{i+1}^n, \min(b_i^n, \alpha_i^n))) \\ L_i &= \min(M_{i+1}^n, \max(B_i^n, \alpha_i^n)) \end{aligned}$$

where b_i^n and B_i^n are defined via formulae (8).
Finally

$$\alpha_{i+1/2}^n = \begin{cases} L_i & \text{if } \alpha_{i+1}^n > L_i, \\ \alpha_{i+1}^n & \text{if } L_i \geq \alpha_{i+1}^n \geq l_i, \\ l_i & \text{if } l_i > \alpha_{i+1}^n \end{cases}$$

- If $\alpha_{i-1}^n < \alpha_{i+1}^n$ (positive centered gradient for α), then

$$\alpha_{\tanh}(x - x_i) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{\beta(x - x_i)}{\Delta x_i} - \tilde{x}_i\right)$$

$$\tilde{x}_i = \tanh^{-1}(1 - 2\alpha_i^n), \quad 0 < \alpha_i^n < 1.$$

$$\tilde{\tilde{B}}_i^n = \min(\alpha_{i+1}^n, \alpha_{\tanh}(0.5\Delta x_i))$$

We redefine

$$M_{i+1}^n = \max(\alpha_i^n, \tilde{\tilde{B}}_i^n)$$

and then we proceed like in classical downwind approach.

$$\begin{aligned} l_i &= \max(m_{i+1}^n, \min(b_i^n, \alpha_i^n))) \\ L_i &= \min(M_{i+1}^n, \max(B_i^n, \alpha_i^n)) \end{aligned}$$

where b_i^n and B_i^n are defined via formulae (8).
Finally

$$\alpha_{i+1/2}^n = \begin{cases} L_i & \text{if } \alpha_{i+1}^n > L_i, \\ \alpha_{i+1}^n & \text{if } L_i \geq \alpha_{i+1}^n \geq l_i, \\ l_i & \text{if } l_i > \alpha_{i+1}^n \end{cases}$$

At the beginning of this section we have supposed that $D_{i-1/2} > 0$, $D_{i+1/2} > 0$. If we suppose that speed $D_{i-1/2} < 0$, $D_{i+1/2} < 0$, we proceed in the same manner. Finally, if $D_{i-1/2}.D_{i+1/2} < 0$ or if $D_{i+1/2} = 0$, we take the upwind value.

2.5 Extension to unstructured grids

The anti-diffusive approach for unstructured meshes we consider here is the one presented in [4]. Let us describe it from a qualitative point of view (see [4] for details).
As shown in Figure 5 in the i-th cell we have to distinguish between inlet and outlet interfaces, according to the scalar product of the velocity $\mathbf{D}$ and the normal vector pointing outside. At the outlet interfaces we define

$$\begin{aligned} m_{i,\text{out},j}^n &= \min\{\alpha_i^n, \alpha_{i,\text{out},j}^n\}, \\ M_{i,\text{out},j}^n &= \max\{\alpha_i^n, \alpha_{i,\text{out},j}^n\}, \\ m_{i,\text{out}}^n &= \min_j\{m_{i,\text{out},j}^n\}, \\ M_{i,\text{out}}^n &= \max_j\{M_{i,\text{out},j}^n\}, \end{aligned}$$

where index $''i, \text{out}, j''$ refers to the neighbor of the i-th cell sharing the j-th outlet interface.
At the outlet interfaces we compute the quantity $\Delta\text{up}_{i,\text{out}}^n$, which is the amount of α transmitted in the upwind case from the i-th cell to the neighbors sharing outlet interfaces. We also define the so-called downwind factor, which is the fraction of $\Delta\text{up}_{i,\text{out}}^n$ retained inside the i-th cell. A downwind factor equal to 0 means that $\Delta\text{up}_{i,\text{out}}^n$ is totally transmitted to its neighbors while a downwind factor equal to 1 means that all $\Delta\text{up}_{i,\text{out}}^n$ is totally kept inside the i-th cell.

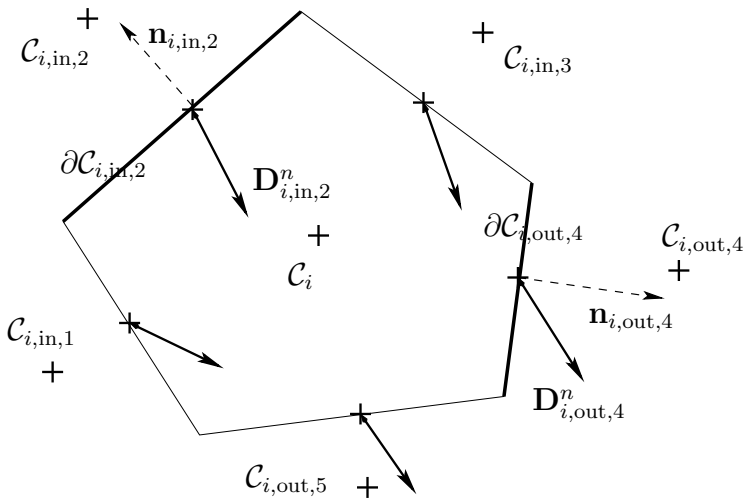


Figure 5: Two dimensional finite volume cell. Illustration for numerical resolution of the transport equation for α.

As in the 1D case, the tanh modification of the anti-diffusive scheme introduces a constraint to the downwind factor. First of all, we compute

$$\vec{n} = -\frac{\vec{\nabla}\alpha}{|\vec{\nabla}\alpha|}$$

$\vec{n}$ is directed versus the region of decreasing α. Then we suppose that, in the same direction as $\vec{n}$, α has the shape

$$\alpha_{\tanh}(s) = \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{\beta s}{\Delta x_i} + \tilde{x_i}\right) \tag{10}$$

where s is the curvilinear abscissa starting from the center of the cell and

$$\tilde{x_i} = \tanh^{-1}(1 - 2\alpha_i^n).$$

We also define

$$s_{\max} = \max_{j \in \text{face}}(|\mathbf{r}_{i,j} \cdot \mathbf{n}|),$$

- If $\Delta \text{up}_{i,\text{out}}^n > 0$, the function $\alpha_{\tanh}(s)$ (which is 1 at $s = -\infty$ and 0 at $s = +\infty$) is moving in the same direction as $\vec{n}$ (namely versus positive value of s). As in 1D, we enforce that the tanh modification should limit the value of $m_{i,\text{out}}^n$. We know that

  $$\alpha_{\tanh}(s_{\max}) < \alpha_i^n.$$

  We compute

  $$\tilde{\tilde{b}}_i^n = \max\left(\alpha_{\tanh}(s_{\max}), m_{i,\text{out}}^n\right)$$

  We define

  $$\tilde{\tilde{\Lambda}}_i^n = \frac{\alpha_i^n - \tilde{\tilde{b}}_i^n}{\alpha_i^n - m_{i,\text{out}}^n}$$

  and we enforce that the downwind factor cannot be larger than $\tilde{\tilde{\Lambda}}_i^n$.

- If $\Delta \text{up}_{i,\text{out}}^n < 0$, the function $\alpha_{\tanh}(s)$ (which is 1 at $s = -\infty$ and 0 at $s = +\infty$) is moving versus the negative value of s. As in 1D, we enforce that the tanh modification should limit the value of $M_{i,\text{out}}^n$. We know that

  $$\alpha_{\tanh}(-s_{\max}) > \alpha_i^n.$$

  We compute

  $$\tilde{\tilde{B}}_i^n = \min\left(\alpha_{\tanh}(-s_{\max}), M_{i,\text{out}}^n\right)$$

  We define

  $$\tilde{\tilde{\Lambda}}_i^n = \frac{\tilde{\tilde{B}}_i^n - \alpha_i^n}{M_{i,\text{out}}^n - \alpha_i^n}$$

  and we enforce that the downwind factor cannot be larger than $\tilde{\tilde{\Lambda}}_i^n$.

Concluding, the principle of the approach is the same as in the 1D case: we introduce a constraint which involves a space scale l to increase the numerical diffusion when the space scale in the i-cell at time t^n is lower than l.

3 Numerical results

We have computed several test cases with the new approaches, which are not presented here. For instance we have verified that both approaches degenerate to the upwind one as β tends to zero ($l = \Delta x/\beta$ tends to infinity). Similarly, both approaches degenerate to the anti-diffusive one as β tends to infinity ($l = \Delta x/\beta$ tends to zero).

The first test case we present here is the propagation of a reactive shock wave. It is a good example to see the behavior of the different approaches.

The second test case is the Chapman-Jouguet deflagration shock tube previously proposed in [7]. This test case is interesting because it allows observing how the poor accuracy on the flame region can affect the accuracy on the flame-generated shock wave. Here we show that the the anti-diffusive approach with tanh modification gives results more accurate that the classical limited second order approach. In particular the number of intermediate cells does not increase with time for the former.

The third test case is a non-reactive test case and consists in the expansion of a cylindrical high-pressure region. It confirms that even in multi-dimensional computations the different approaches behave in the same manner as in 1D computations. In particular the anti-diffusive approach with tanh modification shows to be more accurate than the one involving the Barth-Jespersen reconstruction.

3.1 Moving reactive shock wave

We compute a moving reactive shock wave for a stoichiometric mixture of hydrogen-air. For the unburnt mixture

$$(\rho, v, P) = (1.069100549993681, 545.2, 139000.0)$$

SI units. For the burnt mixture

$$(\rho, v, P) = (0.1468254285125072, 829.1210885558426, 125279.9833247705)$$

SI units. The fundamental speed of the shock is $K_0 = 45.2$ m/s, which gives a visible speed of $D = (545.2 - 45.2) = 500$ m/s. Specific heats are computed as fourth degree polynomial regressions of JANAF data.

Let us now investigate the behavior of different approaches (the anti-diffusive ones with tanh present $\beta = 0.5$). We take a mesh $0 < x < 20$ m with 100 cells, a CFL equal to 0.5 and we suppose that at the beginning the reactive shock is in $x = 10$ m.

In Figure 6 we represent the volume fraction of the burnt gases α obtained using the classical first order upwind (top) and the minmod limited second-order reconstruction (bottom). As expected the result obtained with the former approach is less accurate than the one obtained with the latter one. However in both approaches the number of intermediate regions increases with time.

In Figure 7 we represent the volume fraction of the burnt gases α obtained using the anti-diffusive approach (top) the anti-diffusive approach with tanh modification. As one can see, the anti-diffusive approach presents only one intermediate region while the other one presents a number of intermediate regions (10) which does not vary with time (as expected).

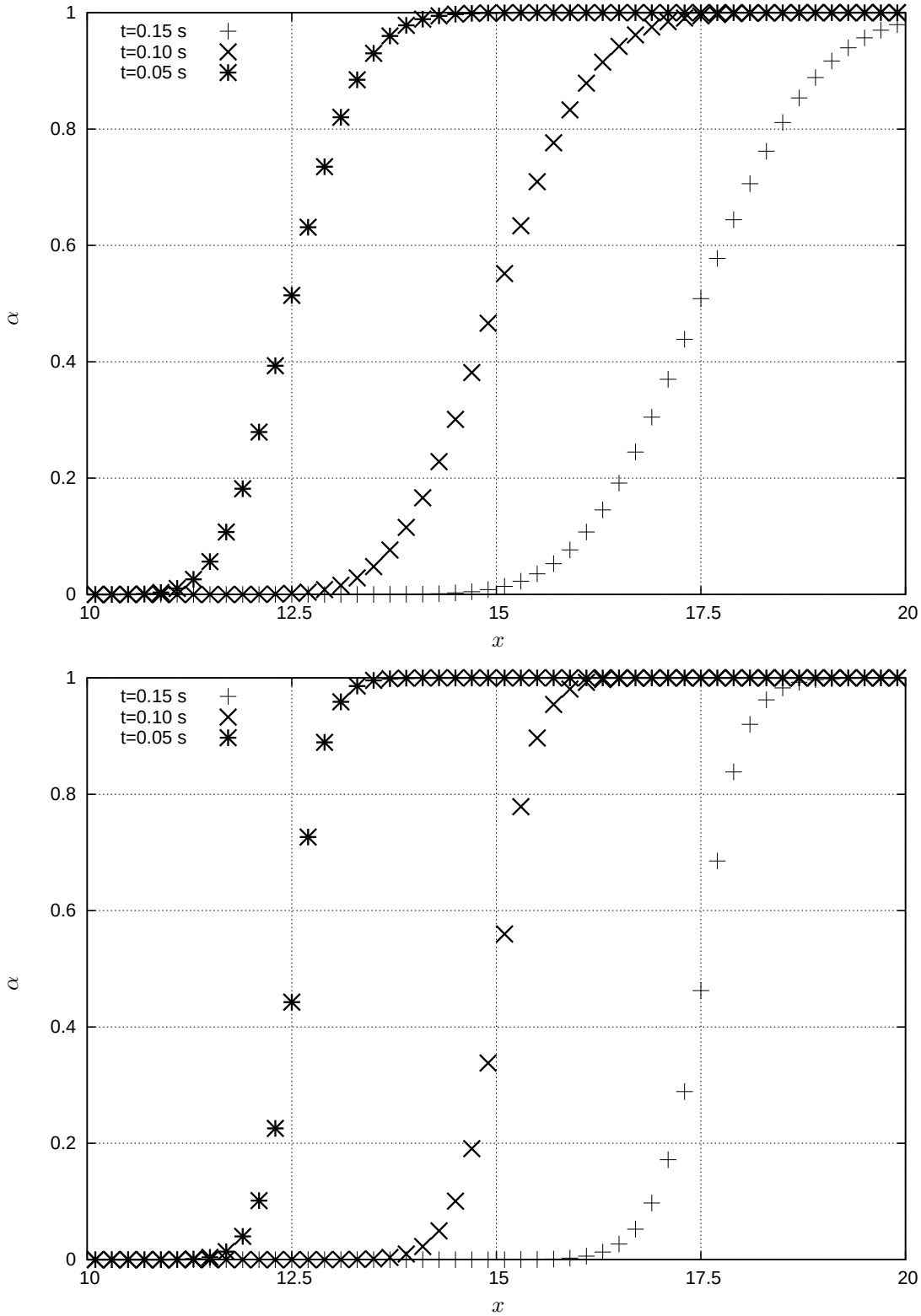


Figure 6: Propagation of a moving reactive shock wave. α, volume fraction of the burnt gases. On the top, first order upwind scheme. On the bottom, minmod limited second-order reconstruction

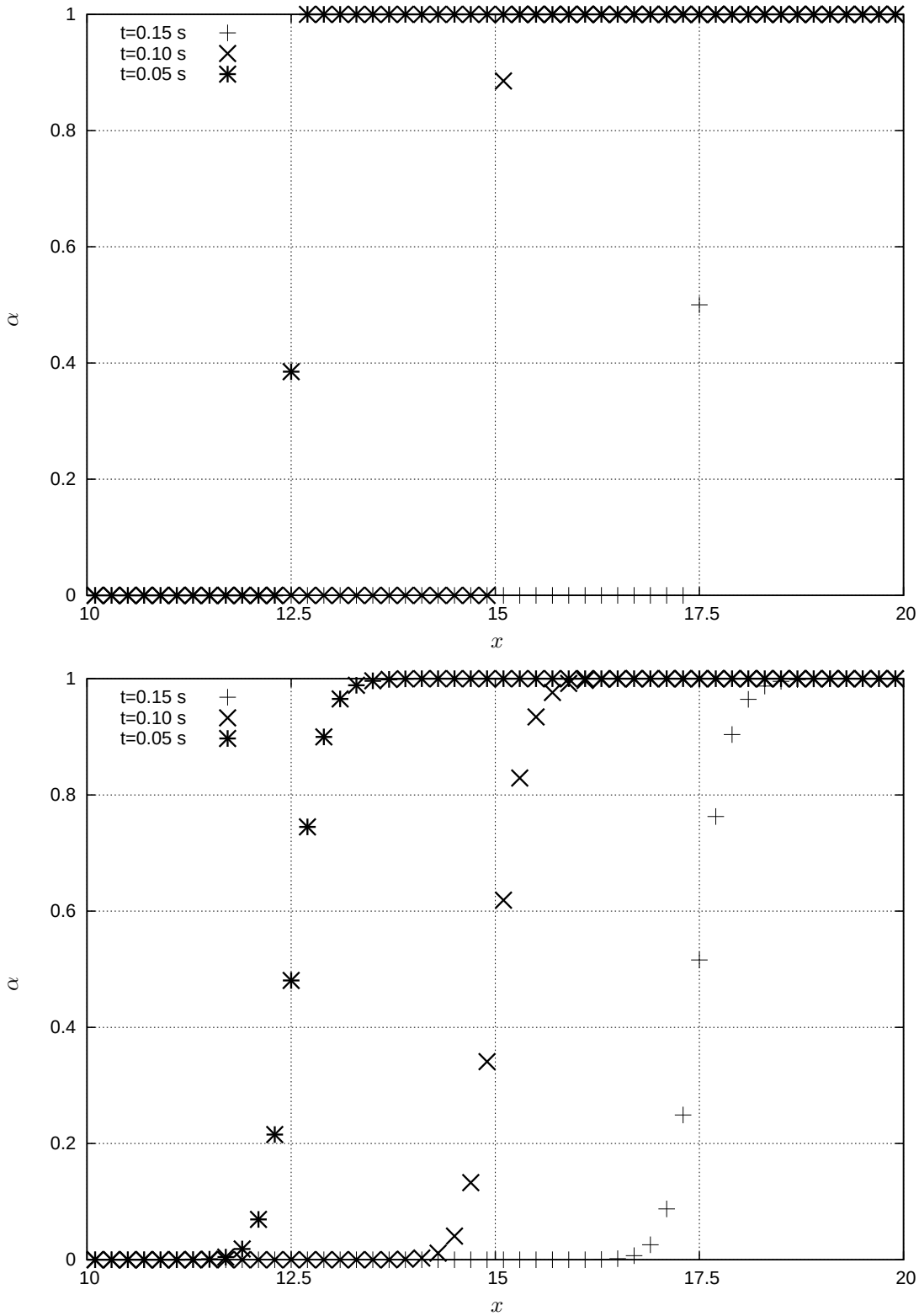


Figure 7: Propagation of a moving reactive shock wave. α, volume fraction of the burnt gases. On the bottom, anti-diffusive approach with tanh modification (case 2, limitation on the downwind-closest value).

3.2 Chapman-Jouguet deflagration shock tube

Let us consider the Chapman-Jouguet deflagration shock tube previously proposed in [7]. The 20 meter long shock tube involves on the right ($0\ \mathrm{m} \leq x \leq 10\ \mathrm{m}$) a stoichiometric mixture of hydrogen-air, with pressure and temperature equal to $1.013\ \mathrm{bar}$ and $290\ \mathrm{K}$ respectively. On the left ($-10\ \mathrm{m} \leq x \leq 0\ \mathrm{m}$), the burnt gas (due to the complete combustion of the stoichiometric mixture of hydrogen-air) is found with pressure and temperature equal to $2.013\ \mathrm{bar}$ and $2800\ \mathrm{K}$. The specific heats are computed as fourth degree polynomials of the temperature, obtained by interpolating data in JANAF tables. For the modeling and numerical reasons already explained in [5], the right part contains a very small volume of burnt gas (volume fraction is $\varepsilon = 10^{-8}$) and the left part contains a very small volume of unburnt gas (volume fraction is $\varepsilon = 10^{-8}$). The gases are initially at rest and the fundamental flame speed is set equal to $200\ \mathrm{m/s}$.

Let us now investigate the behavior of different approaches (the anti-diffusive one with tanh modification present $\beta = 0.5$). We take 1000 cells and a CFL equal to 0.75; second-order in time is achieved using an explicit Runge-Kutta scheme and we perform a minmod limited second-order reconstruction on primitive variables. Different approaches are tested on the volume fraction: minmod limited second-order reconstruction, anti-diffusive approach, anti-diffusive approach with tanh modification.

In Figure 8 we represent the burnt gas volume fraction α as function of x at different times. As usual, the minmod limited second-order reconstruction of α generates a numerical diffusion region dimension which increases with time. The anti-diffusive approach on α generates one intermediate region only. Finally the anti-diffusive approach with tanh modification generates a numerical diffusion region with about 10 cells which does not increase with time.
In Figure 9 we present the computed pressure close to the reactive front. As usual the anti-diffusive approach gives the most accurate results but the anti-diffusive approach with tanh modification also gives a correct representation of the flame region thanks to the fact that the numerical diffusion region does not increase with time.

3.3 2D Expansion of a cylindrical high-pressure high-density region

We have a quadrangular domain with 121x121 element with $\Delta x = 1$ (non-dimensional units), namely $-60.5 \leq x \leq 60.5$, $-60.5 \leq y \leq 60.5$. At the center we have a cylinder with $r = 6$ and we suppose that in the elements having their center inside this cylinder we have $P = 10^2$ and $\rho = 1.4 \cdot 10^2$ (and color function equal to 1). Outside we have $P = 1$ and $\rho = 1.4$ (and color function equal to 0). We suppose that we are dealing with a calorically perfect gas, with $\gamma = 1.4$.

Computations are realized with EUROPLEXUS using the DEMS material [8], the Euler explicit scheme for time integration, CFL = 0.25 and four different approaches:

- the upwind approach;
- UDCS combined with the Barth-Jespersen (limited second-order) reconstruction;
- UDCS combined with anti-diffusive approach;
- UDCS combined with anti-diffusive approach with tanh modification.

Because of the high-density high-pressure region, the cylindrical region expands in the low-density low-pressure region. In Figure 10 we represent the color function α at $t = 1$ and at

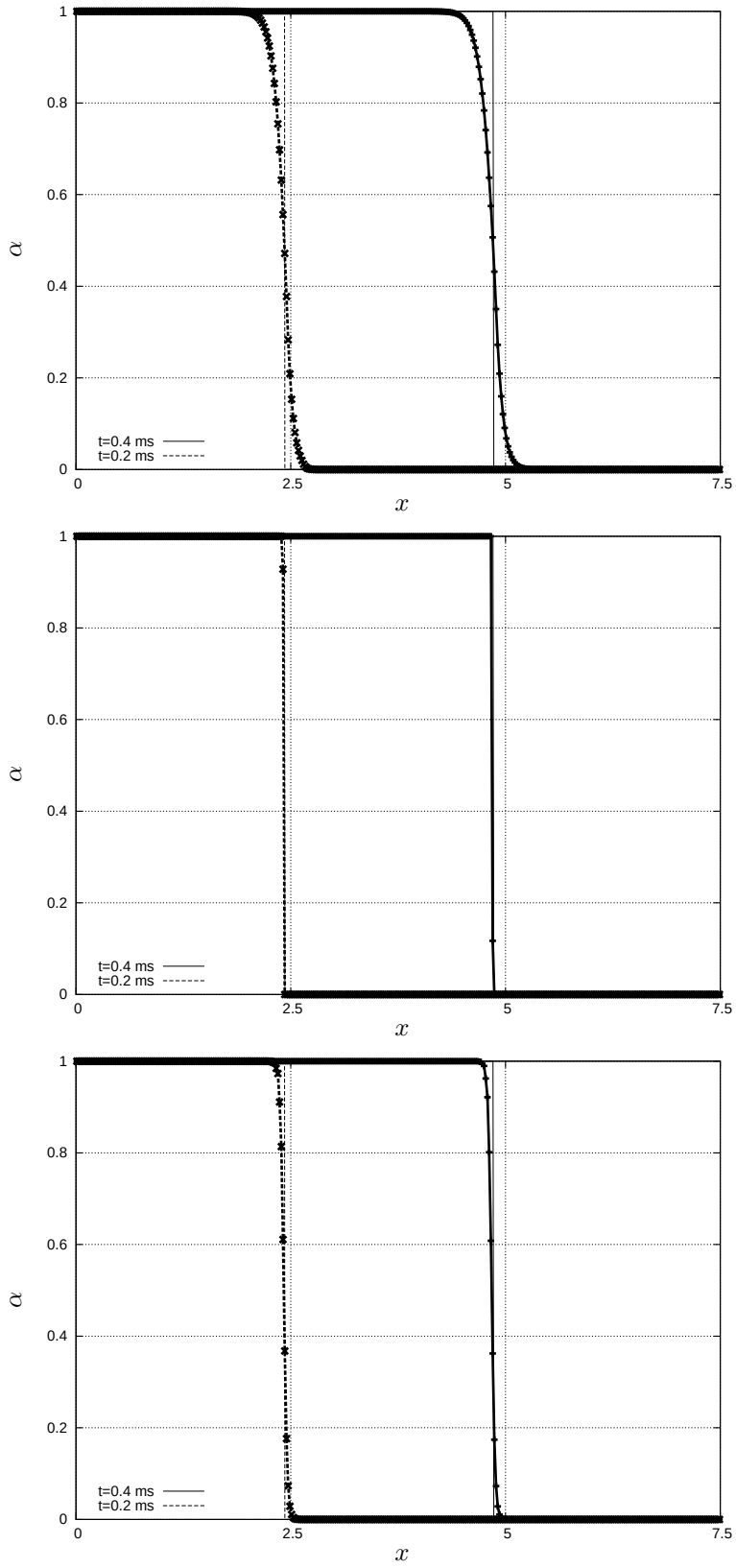


Figure 8: Chapman-Jouguet deflagration shock tube. Volume fraction of the burnt gases α as function of x (SI units). On the top, minmod limited second-order reconstruction on α. On the center anti-diffusive approach on α. On the bottom, anti-diffusive approach with tanh modification on α (the whole domain is -10 m $\leq x \leq$ 10 m).

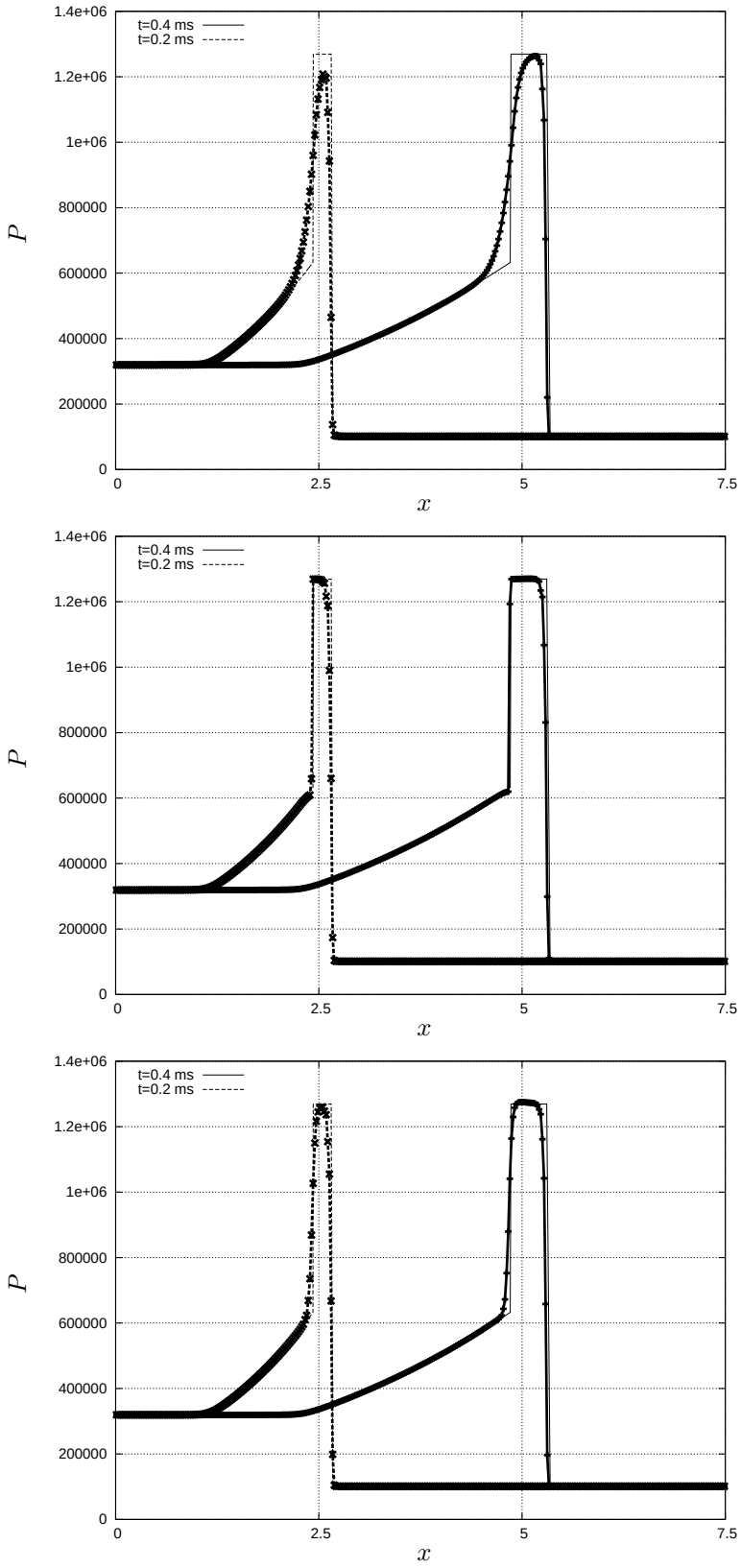


Figure 9: Chapman-Jouguet deflagration shock tube. Pressure P as function of x (SI units). On the top, minmod limited second-order reconstruction on α. On the center anti-diffusive approach on α. On the bottom, anti-diffusive approach with tanh modification on α (the whole domain is -10 m $\leq x \leq$ 10 m).

$t = 10$ obtained with the anti-diffusive approach.

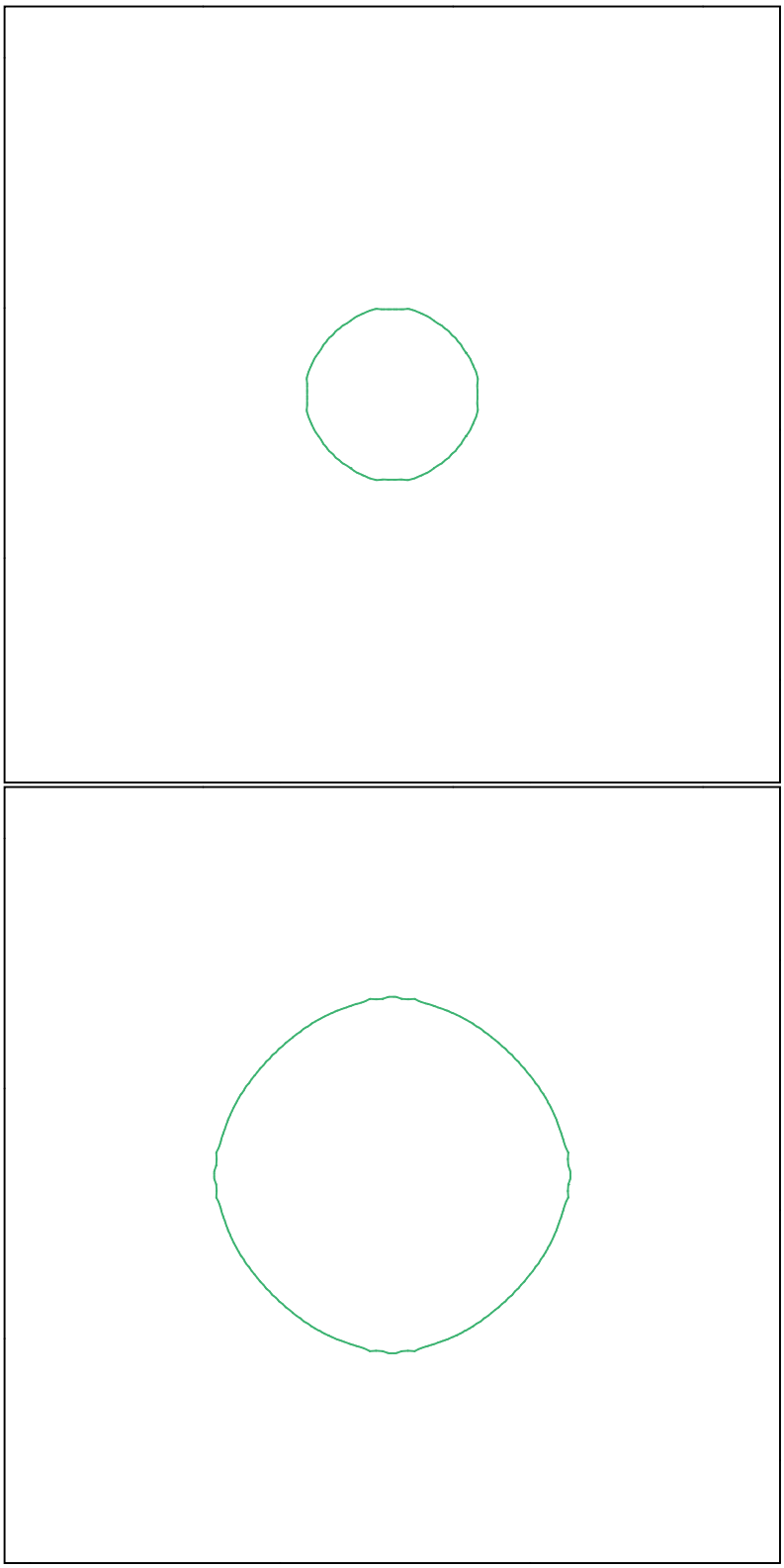


Figure 10: Expansion of a cylindrical high-pressure high-density region. Anti-diffusive approach. α at $t = 1$ (top) and at $t = 10$ (bottom).

In Figure 11 we represent the solution on the cell centers with $y = 0$ (the curvilinear abscissa involving centers begins at $x = 0$ and ends at $x = 120$). As one can see, for the anti-diffusive approach we have only one or two intermediate zones between α equal to 1 and 0. Solution for the anti-diffusive approach with tanh modification is obtained with $\beta = 0.75$. As one can see, at $t = 1$ this solution is as accurate as the one obtained with the Barth-Jespersen approach. At $t = 10$ the number of intermediate regions keeps the same for the former while increase for the latter one, as expected, namely the anti-diffusive approach with tanh modification becomes more accurate than the one involving the Barth-Jespersen reconstruction.

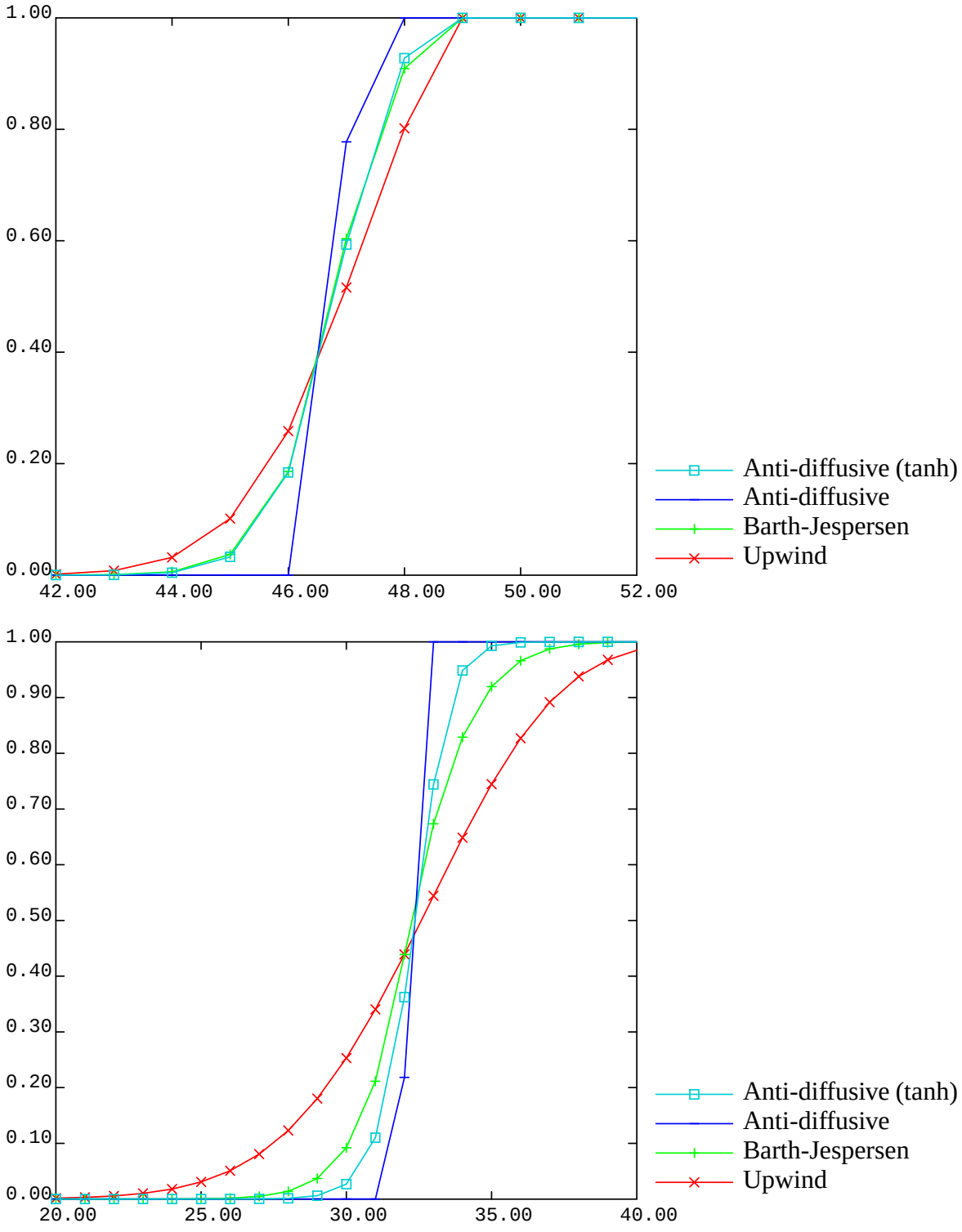


Figure 11: Expansion of a cylindrical high-pressure high-density region. α at $t = 1$ (top) and $t = 10$ (bottom) using different approaches. $y = 0$ (the curvilinear abscissa involving centers begins at $x = 0$ and ends at $x = 120$).

4 Conclusion and perspectives

A modification of the anti-diffusive approach is here proposed in 1D and multi-dimensional cases. This approach uses the hyperbolic tangent function to introduce a space scale for the diffusion region of the color function α. The existence of this diffusion region with a finite width allows a correct computation of gradient of α and then a correct computation of normals to the flame surface. Moreover, as shown by several numerical experiments here considered in reactive and non-reactive cases, the number of intermediate regions (i.e. its width) remains constant in time, making this approach more accurate than the ones which use a classical limited second order reconstruction.

Of course this way of computing normals does not resolve the problem of flame wrinkling, which in the case of the new approach is even more important than in case of approaches involving a limited second order reconstruction (as expected, since numerical diffusion is less important in the former case).

A way of reducing the flame wrinkling is to compute normals to the flame interface using a field obtained by the diffusion of the color function α. If the elimination of the flame wrinkling is a must, further studies should be done in this direction.

REFERENCES

[1] K. Tang, A. Beccantini, and C. Corre, "Combining Discrete Equations Method and upwind downwind-controlled splitting for non-reacting and reacting two-fluid computations," in *Seventh International Conference on Computational Fluid Dynamics (ICCFD7), Big Island, Hawaii, July 9-13*, 2012.

[2] K. Tang, *Combining Discrete Equations Method and Upwind Downwind-Controlled Splitting for Non-Reacting and Reacting Two-Fluid Computations*. PhD thesis, Université de Grenoble, dec 2012.

[3] K. Tang, A. Beccantini, and C. Corre, "Combining Discrete Equations Method and upwind downwind-controlled splitting for non-reacting and reacting two-fluid computations: One dimensional case," *Computers & Fluids*, vol. 93, pp. 74–90, 2014.

[4] K. Tang, A. Beccantini, and C. Corre, "Combining Discrete Equations Method and upwind downwind-controlled splitting for non-reacting and reacting two-fluid computations: Two dimensional case," *Computers & Fluids*, vol. 103, pp. 132–155, 2014.

[5] R. Abgrall and R. Saurel, "Discrete equations for physical and numerical compressible multiphase mixtures," *Journal of Computational Physics*, vol. 186, no. 2, pp. 361–396, 2003.

[6] O. L. Métayer, J. Massoni, R. Saurel, O. Le Métayer, J. Massoni, and R. Saurel, "Modelling evaporation fronts with reactive Riemann solvers," *Journal of Computational Physics*, vol. 205, no. 2, pp. 567–610, 2005.

[7] A. Beccantini and E. Studer, "The reactive Riemann problem for thermally perfect gases at all combustion regimes," *International Journal for Numerical Methods in Fluids*, vol. 64, pp. 269–313, 2010.

[8] "Europlexus, a computer program for the finite element simulation of fluid-structure systems under transient dynamic loading." `http://europlexus.jrc.ec.europa.eu`, visited in October 2015.

[9] E. Studer, A. Beccantini, S. Kudriakov, and A. Velikorodny, "Hydrogen Combustion Modelling in Large-Scale Geometries," in *Volume 6: Beyond Design Basis Events; Student Paper Competition*, p. V006T15A030, ASME, jul 2013.

[10] A. Velikorodny, E. Studer, S. Kudriakov, and A. Beccantini, "Combustion modeling in large scale volumes," in *International Conference on Hydrogen Safety*, (Brussels, Belgium), 2013.

[11] A. Velikorodny, E. Studer, S. Kudriakov, and A. Beccantini, "Combustion modeling in large scale volumes using EUROPLEXUS code," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 104–116, may 2015.

[12] F. Xiao, Y. Honma, and T. Kono, "A simple algebraic interface capturing scheme using hyperbolic tangent function," *International Journal for Numerical Methods in Fluids*, vol. 48, pp. 1023–1040, jul 2005.

[13] K. Yokoi, "Efficient implementation of THINC scheme: A simple and practical smoothed VOF algorithm," *Journal of Computational Physics*, vol. 226, no. 2, pp. 1985–2002, 2007.

[14] F. Xiao, S. Ii, and C. Chen, "Revisit to the THINC scheme: A simple algebraic VOF algorithm," *Journal of Computational Physics*, vol. 230, no. 19, pp. 7086–7092, 2011.

[15] S. Ii, K. Sugiyama, S. Takeuchi, S. Takagi, Y. Matsumoto, and F. Xiao, "An interface capturing method with a continuous function: The THINC method with multi-dimensional reconstruction," *Journal of Computational Physics*, vol. 231, no. 5, pp. 2328–2358, 2012.

[16] S. Ii, B. Xie, and F. Xiao, "An interface capturing method with a continuous function: The THINC method on unstructured triangular and tetrahedral meshes," *Journal of Computational Physics*, vol. 259, pp. 260–269, feb 2014.

[17] B. Xie, S. Ii, and F. Xiao, "An efficient and accurate algebraic interface capturing method for unstructured grids in 2 and 3 dimensions: The THINC method with quadratic surface representation," *International Journal for Numerical Methods in Fluids*, vol. 76, pp. 1025–1042, dec 2014.

[18] F. Lagoutière, *Modelisation mathematique et resolution numerique de problemes de fluides compressibles a plusieurs constituants*. PhD thesis, Paris 6, jan 2000.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A LARGE TIME STEP ROE SCHEME APPLIED TO TWO-PHASE FLOW

Sofia Lindqvist[1,2] **and Halvor Lund**[2,3]

[1]Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG, UK.
[2]SINTEF Energy Research, P. O. Box 4761 Sluppen, NO-7465 Trondheim, Norway.
[3]Corresponding author, e-mail: halvor.lund@sintef.no

Keywords: Large time step, Roe scheme, two-phase flow

Abstract. *We consider a large time step (LTS) Roe scheme, originally suggested by LeVeque [*Comm. Pure Appl. Math.*, 37 (1984), pp. 463–477] which has attained increased popularity in recent years. The scheme is based on a wave formulation of the ordinary Roe scheme, and the assumption that waves interact linearly. This assumption lets us propagate waves over more than one cell, thereby violating the ordinary CFL condition requiring the Courant number to be less than one. The LTS Roe scheme is applied to a two-phase flow model with phase transfer, modelled using the stiffened-gas equation of state. Results from various test cases show that the LTS Roe scheme with moderate Courant numbers is significantly more efficient than the ordinary Roe scheme, measured by an error-to-runtime ratio. The optimal Courant number is mainly a trade-off between reduced runtime due to fewer time steps, and increased error due to the assumption of linear wave interactions.*

1 INTRODUCTION

The motivation for our paper is solving hyperbolic conservation laws of the form

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0, \tag{1a}$$

$$\boldsymbol{u}(x,0) = \boldsymbol{u}_0(x). \tag{1b}$$

Here $\boldsymbol{u} \in \mathbb{R}^N$ is a vector of N unknowns, $\boldsymbol{f} : \mathbb{R}^N \to \mathbb{R}^N$ is a hyperbolic flux function, and $\boldsymbol{u}_0$ some initial condition. Numerical methods for solving such problems often fit into the finite volume framework, which solves eq. (1) in the integral form,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{U}_i = \frac{\boldsymbol{F}_{i-1/2} - \boldsymbol{F}_{i+1/2}}{\Delta x}, \tag{2}$$

where $\boldsymbol{U}_i$ approximates the average of $\boldsymbol{u}$ over volume i. The numerical flux function $\boldsymbol{F}_{i-1/2}$ approximates the flux between cells $i-1$ and i, using only the cell averages $\{\boldsymbol{U}_i\}$.

Standard explicit finite volume methods typically use a three-point stencil which limits the time steps to prevent waves from travelling more than one cell. Large time step (LTS) schemes seek to circumvent this restriction by using a larger stencil, to allow longer time steps and therefore more efficient numerical simulations. Explicit LTS schemes also have the advantage of being easily parallelizable, in contrast to implicit schemes, as pointed out by e.g. Norman et al. [20]. One of the first large time step schemes for scalar conservation laws was presented by LeVeque [9] in 1982. This scheme represents all waves as shocks, and handles the merging of shocks explicitly. LeVeque later introduced an LTS scheme for systems of conservation laws [11] using an exact Riemann solver, and proposed one using the Roe approximate Riemann solver [10]. The wave interactions are assumed to be linear, which removes the need to handle wave interactions explicitly. More recently, Qian and Lee [21] extended this scheme by taking wave interactions into account. Norman et al. [20] developed a LTS scheme similar to LeVeque's approach, based on a WENO discretization.

Harten [4] developed a class of LTS schemes of second order with $2K+3$ stencils, where K is the Courant number. He proves that the scheme is total-variation diminishing (TVD) for linearized systems, and introduces an entropy fix. This scheme was later picked up by Qian and Lee [22] and slightly modified to improve its stability. They also constructed an LTS scheme for Yee's upwind TVD scheme [27]. Other approaches to LTS schemes outside the framework of Godunov methods have also been presented, such as the front-tracking scheme of Holden et al. [6].

Murillo et al. [19] and Morales-Hernández et al. [17] picked up on LeVeque's ideas and applied an LTS Roe scheme to the shallow water equations. Recently, Thompson and Moeller [26] introduced their wave-in-cell numerical scheme for hyperbolic conservation laws. The scheme has many similarities with front-tracking schemes, and tracks the waves from each Riemann problem explicitly, assuming they interact linearly. For nonlinear systems, the solution is projected onto a grid after each time step. Both these schemes are equivalent to the LTS Roe scheme we present in this paper, although applied to a different system of equations.

The LTS Roe scheme presented in the current paper is built on the general framework of TVD LTS schemes described by Lindqvist et al. [13]. This work derived TVD requirements for a class of LTS schemes, and discusses how existing schemes like LTS Godunov and LTS Roe fit into this TVD LTS framework. They prove that an LTS extension of the Lax-Friedrichs scheme is the most diffusive scheme, and that the LTS Roe scheme is the least diffusive scheme.

In this paper, we apply the LTS Roe scheme to a two-phase flow model for CO_2 pipelines, with thermodynamics described by the stiffened-gas equation of state. CO_2 capture and storage (CCS) is expected to play a crucial role in reducing the world's emissions of greenhouse gases. In the International Energy Agency's two-degree scenario (2DS), CCS will contribute to reducing the global CO_2 emissions by about seven gigatonnes per year in 2050 [7]. CO_2 pipelines will be one of the key elements in a CCS infrastructure, and accurate models and efficient numerical methods are necessary to predict their operation.

1.1 Paper outline

Our paper is organized as follows: In section 2, we give a short introduction to the ordinary Roe scheme and its large time step (LTS) extension. Section 3 describes the homogeneous equilibrium model for two-phase flow, which we will use in our numerical test cases. The model is applied to two-phase flow of CO_2, the thermodynamics of which is described by the stiffened-gas equation of state. Results from various test cases relevant to compressible two-phase flow are presented in section 4, where we investigate how the LTS Roe scheme performs for Courant numbers greater than 1. The different Courant numbers are compared in terms of error versus runtime. Finally, section 5 summarizes our findings and outlines how the scheme can be improved and extended to higher order.

2 NUMERICAL SCHEME

To solve the hyperbolic equation system (1) numerically, we divide the x and t axes into finite volumes of size $(\Delta x, \Delta t)$, and integrate over one such cell,

$$\boldsymbol{U}_i^{n+1} = \boldsymbol{U}_i^n + \frac{\Delta t}{\Delta x}(\boldsymbol{F}_{i-1/2}^n - \boldsymbol{F}_{i+1/2}^n), \tag{3}$$

where $\boldsymbol{U}_i^n$ is the average of $\boldsymbol{u}$ over cell i at time t^n, and $\boldsymbol{F}_{i-1/2}^n$ approximates the flux at the cell interface between cells $i-1$ and i.

In the classical Roe scheme [24], the numerical flux $\boldsymbol{F}_{i-1/2}^n$ is given by

$$\boldsymbol{F}_{i-1/2}^n = \frac{1}{2}(\boldsymbol{F}(\boldsymbol{U}_i^n) + \boldsymbol{F}(\boldsymbol{U}_{i-1}^n)) - \frac{1}{2}|\hat{\boldsymbol{A}}(\boldsymbol{U}_{i-1}^n, \boldsymbol{U}_i^n)|(\boldsymbol{U}_i^n - \boldsymbol{U}_{i-1}^n), \tag{4}$$

where $\hat{\boldsymbol{A}}$ is the Roe matrix, which must satisfy three properties:

1. Conserving: $\hat{\boldsymbol{A}}(\boldsymbol{U}_\mathrm{L}, \boldsymbol{U}_\mathrm{R})(\boldsymbol{U}_\mathrm{R} - \boldsymbol{U}_\mathrm{L}) = \boldsymbol{F}(\boldsymbol{U}_\mathrm{R}) - \boldsymbol{F}(\boldsymbol{U}_\mathrm{L})$
2. Hyperbolic: $\hat{\boldsymbol{A}}(\boldsymbol{U}_\mathrm{L}, \boldsymbol{U}_\mathrm{R})$ is diagonalizable with real eigenvalues
3. Consistent: $\hat{\boldsymbol{A}}(\boldsymbol{U}_\mathrm{L}, \boldsymbol{U}_\mathrm{R}) \to \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}}$ smoothly as $\boldsymbol{U}_\mathrm{L}, \boldsymbol{U}_\mathrm{R} \to \boldsymbol{U}$

Here, $\boldsymbol{U}_\mathrm{L}$/$\boldsymbol{U}_\mathrm{R}$ are the states to the immediate left/right of the cell interface.

If $\hat{\boldsymbol{A}} = \hat{\boldsymbol{R}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{R}}^{-1}$ is a diagonalization of the Roe matrix, the absolute Roe matrix is $|\hat{\boldsymbol{A}}| = \hat{\boldsymbol{R}}|\hat{\boldsymbol{\Lambda}}|\hat{\boldsymbol{R}}^{-1}$. Here $|\hat{\boldsymbol{\Lambda}}| = \mathrm{diag}(|\hat{\lambda}_1|, \ldots, |\hat{\lambda}_N|)$ is the diagonal matrix with the absolute eigenvalues of $\hat{\boldsymbol{A}}$. It is instructive for our further derivations to rewrite the scheme in terms of waves, as

$$\boldsymbol{U}_i^{n+1} = \boldsymbol{U}_i^n - \frac{\Delta t}{\Delta x}(\hat{\boldsymbol{A}}_{i-1/2}^+ \Delta_{i-1/2} + \hat{\boldsymbol{A}}_{i+1/2}^- \Delta_{i+1/2}), \tag{5}$$

where $\Delta_{i-1/2} = \boldsymbol{U}_i - \boldsymbol{U}_{i-1}$, $\hat{\boldsymbol{A}}^\pm = \hat{\boldsymbol{R}}\hat{\boldsymbol{\Lambda}}^\pm\hat{\boldsymbol{R}}^{-1}$, and $\hat{\boldsymbol{\Lambda}}^\pm = \pm\mathrm{diag}(\max(\pm\hat{\lambda}_1, 0), \ldots, \max(\pm\hat{\lambda}_N, 0))$. This formulation illustrates how the fluctuations $\hat{\boldsymbol{A}}_{i-1/2}^+ \Delta_{i-1/2}$ (right-going) and $\hat{\boldsymbol{A}}_{i+1/2}^- \Delta_{i+1/2}$ (left-going) propagate as waves in each direction.

2.1 Large time step (LTS) Roe scheme

In this section, we present the most important aspects of the LTS Roe scheme, which was first proposed by Leveque [10]. For more details, the reader is referred to Lindqvist et al. [13]. The time step Δt in the ordinary Roe scheme in eq. (5) is limited by the CFL condition [2], which requires the Courant number $\mathcal{C}$ to be less than one,

$$\mathcal{C} = \frac{\Delta t}{\Delta x} \max_{k,i}(|\lambda_{k,i}|) < 1. \tag{6}$$

Here $\lambda_{k,i}$ is the kth eigenvalue of $\partial \boldsymbol{F}/\partial \boldsymbol{U}$ in the ith cell. This condition ensures that no waves travel more than one cell in a single time step, or equivalently that the numerical domain of dependence includes the mathematical domain of dependence. The motivation behind the LTS Roe scheme (or any LTS scheme, for that matter) is to enable larger time steps by extending the numerical domain of dependence.

If we assume that all wave interactions are linear, a natural extension of the Roe scheme is

$$\boldsymbol{U}_j^{n+1} = \boldsymbol{U}_j^n - \frac{\Delta t}{\Delta x} \sum_{i=0}^{\infty} \left(\hat{\boldsymbol{A}}^{i+}_{j-1/2-i} \Delta_{j-1/2-i} + \hat{\boldsymbol{A}}^{i-}_{j+1/2+i} \Delta_{j+1/2+i} \right), \tag{7}$$

where

$$\hat{\boldsymbol{A}}^{i\pm} = \hat{\boldsymbol{R}} \hat{\boldsymbol{\Lambda}}^{i\pm} \hat{\boldsymbol{R}}^{-1}, \tag{8a}$$

$$\hat{\boldsymbol{\Lambda}}^{i\pm} = \text{diag}(\hat{\lambda}^{i\pm}), \tag{8b}$$

$$\hat{\lambda}^{i+} = \max\left(0, \min\left(\hat{\lambda} - i\frac{\Delta x}{\Delta t}, \frac{\Delta x}{\Delta t}\right)\right), \tag{8c}$$

$$\hat{\lambda}^{i-} = \min\left(0, \max\left(\hat{\lambda} + i\frac{\Delta x}{\Delta t}, -\frac{\Delta x}{\Delta t}\right)\right). \tag{8d}$$

Although the sum in eq. (7) appears to be infinite, $\hat{\boldsymbol{A}}^{i\pm}$ will be zero for sufficiently large i. This reduces to the ordinary Roe scheme for Courant numbers less than 1. More importantly, it allows arbitrarily large Courant numbers while still respecting the condition that the numerical domain of dependence include the mathematical one. For scalar equations, we also know that the scheme is TVD [13]. For a Courant number of $\mathcal{C}$, the scheme's stencil will include a maximum of $2\lceil \mathcal{C} \rceil + 1$ cells.

2.2 Entropy fix

Lindqvist et al. [13] show that the scheme is unconditionally total-variation diminishing for scalar equations for all Courant numbers. From the Lax-Wendroff theorem [8] we know that TVD schemes converge to a weak solution for scalar equations, but not necessarily the correct entropy solution [5]. In [13], it is suggested to vary the time step slightly, which ensures that entropy-violating shocks are eliminated. In this work, we employ this entropy fix, by using a modified Courant number $\hat{C}$ given by

$$\hat{C} = \mathcal{C} - r, \tag{9}$$

where r is sampled randomly and uniformly from the interval $(0, 1)$.

Lindqvist et al. [13] also prove an extension of Harten's lemma [3]: All unconditionally TVD schemes have numerical viscosities that are bounded by the LTS Lax-Friedrichs and the LTS Roe schemes. The related work of Qui and Shu [23] also concerns the entropy properties of TVD LTS schemes.

3 TWO-PHASE FLOW MODEL

For our numerical test cases, we model the two-phase (gas-liquid) flow of CO_2 using the one-dimensional homogeneous equilibrium model (HEM) [25], in which the two phases have the same pressure, temperature, chemical potential and velocity. The advantage of this model is that it does not require explicit modelling of the interaction between the phases, such as interfacial friction, heat and mass. It can nevertheless describe two-phase flow reasonably well in cases where the two phases are tightly coupled, such as in bubbly flow or in small diameter pipes. This model has been used for simulation of two-phase flow by e.g. Clerc [1], Lund et al. [14].

The homogeneous equilibrium model can be expressed using the Euler equations,

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho v)}{\partial x} = 0, \tag{10a}$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial (\rho v^2 + p)}{\partial x} = 0, \tag{10b}$$

$$\frac{\partial E}{\partial t} + \frac{\partial [v(E+p)]}{\partial x} = 0, \tag{10c}$$

where ρ is the mixture density, v is the velocity, p is the pressure and E is the mixture energy density. The latter can be expressed as

$$E = \rho \left(e + \frac{1}{2} v^2 \right), \tag{11}$$

where e is the mixture internal energy density.

To close the equation system (10), we need an equation of state (EOS) to relate the pressure p to the mixture density ρ and the mixture internal energy density e. In this work, we construct a mixture EOS by describing each phase using the stiffened-gas EOS, as described by e.g. Menikoff [15], Menikoff and Plohr [16]. For each phase, this can be viewed as a linearization of a reference EOS, and has the advantage of being simple and highly efficient, while still giving relatively accurate results close to a chosen reference point. The two stiffened-gas equations of state are combined into a mixture EOS by requiring that both phases have the same pressure, temperature and chemical potential.

The pressure, internal energy and chemical potential (for each phase) are given by

$$p_k(\rho_k, T) = \rho_k(\gamma_k - 1) c_{v,k} T - p_{\infty,k}, \tag{12}$$

$$e_k(\rho_k, T) = c_{v,k} T + \frac{p_{\infty,k}}{\rho_k} + e_{*,k}, \tag{13}$$

$$\mu_k(\rho_k, T) = \gamma_k c_{v,k} T + e_{*,k} - c_{v,k} T \ln \left(\frac{T}{T_{0,k}} \left(\frac{\rho_{0,k}}{\rho} \right)^{\gamma_k - 1} \right) - s_{0,k} T, \tag{14}$$

where γ_k, $c_{v,k}$, $p_{\infty,k}$, $e_{*,k}$, $T_{0,k}$, $\rho_{0,k}$ and $s_{0,k}$ are parameters specific for phase $k = \{\mathrm{g}, \ell\}$ where g denotes gas and ℓ denotes liquid. These parameters can be fitted using experimental data or a reference EOS for a given fluid.

The pressure p must be expressed as a function of mixture density $\rho \equiv \alpha_\mathrm{g}\rho_\mathrm{g} + \alpha_\ell\rho_\ell$ and mixture energy $e \equiv (\alpha_\mathrm{g}\rho_\mathrm{g}e_\mathrm{g} + \alpha_\ell\rho_\ell e_\ell)/\rho$, with the condition that the phase pressures, temperatures and chemical potentials be equal, $p_\mathrm{g} = p_\ell = p$, $T_\mathrm{g} = T_\ell = T$ and $\mu_\mathrm{g} = \mu_\ell$. This leads to a transcendental equation system that is solved numerically using a Newton method, as in Lund et al. [14].

In the LTS Roe scheme for the homogeneous equilibrium model (10), we use the Roe matrix described by Morin et al. [18]. The Roe matrix is diagonalized numerically.

4 RESULTS

In this section, we will present results from various test cases relevant to compressible flow. We will compare the LTS Roe scheme to an ordinary Roe scheme with a Courant number less than one. The performance of the two methods will be measured by the error achieved for a certain runtime, which in this case is more relevant than the error on a certain grid, as long as we are not limited by memory. The error is measured in the L^2 norm by comparing the numerical solutions to reference solutions calculated with a Roe method with minmod flux limiter on a 6000 cell grid. More precisely, the error $\mathcal{E}$ can be expressed as

$$\mathcal{E} = \frac{\sqrt{\sum_{i=1}^{n}(U_i - U_{\mathrm{ref},i})^2}}{n}, \tag{15}$$

where U_i and $U_{\mathrm{ref},i}$ are the numerical and the reference solution at point i, respectively.

The properties used for the stiffened gas equation of state are listed in table 1. These properties are fitted to match the real properties of carbon dioxide at $p = 45\,\mathrm{bar}$ and saturation conditions ($T = 283.13\,\mathrm{K}$).

Table 1: Stiffened gas parameters, fitted to CO_2 at $45\,\mathrm{bar}$ and saturation conditions.

Symbol	Unit	Gas value	Liquid value
γ	–	1.06	1.23
p_∞	Pa	8.86×10^5	1.32×10^8
c_v	$\mathrm{J/(kg\,K)}$	2411	2436
e_*	$\mathrm{J/kg}$	-3.00×10^5	-6.23×10^5
T_0	K	283.13	283.13
ρ_0	$\mathrm{kg/m^3}$	135	861.27
s_0	$\mathrm{J/(kg\,K)}$	1784	1088

4.1 Advected volume fraction discontinuity

In this case, there is an initial discontinuity in the volume fraction which is advected in the positive x direction at a constant speed. The initial conditions are

$$p = 20\,\mathrm{bar}, \tag{16a}$$

$$u = 10\,\mathrm{m/s}, \tag{16b}$$

$$\alpha_\ell = \begin{cases} 0.1 \text{ if } x \leq 2500\,\mathrm{m}, \\ 0.9 \text{ if } x > 2500\,\mathrm{m}, \end{cases} \tag{16c}$$

$$T = T_\mathrm{eq}(p), \tag{16d}$$

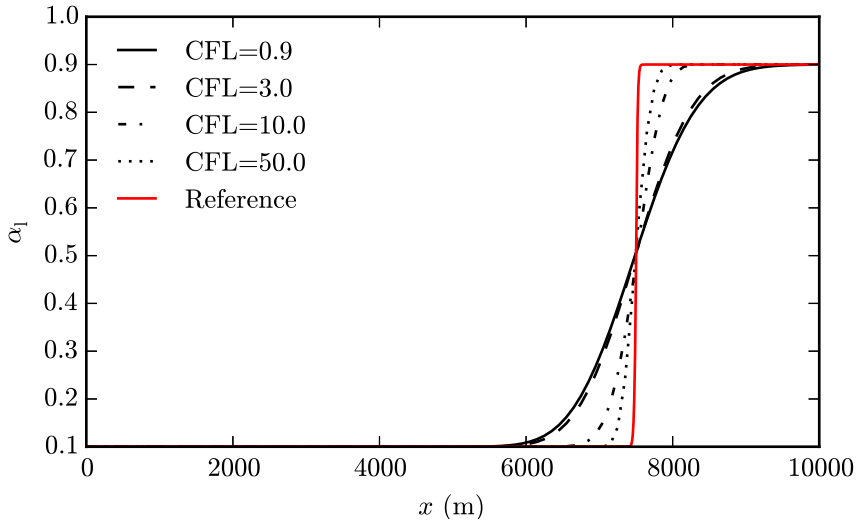


Figure 1: Liquid volume fraction discontinuity at time $t = 500\,\mathrm{s}$ on a grid with 100 cells for initial conditions given in eq. (16).

where $T_{\mathrm{eq}}(p)$ denotes the equilibrium (boiling) temperature at the pressure p.

Figure 1 shows the liquid volume fraction at time $t = 500\,\mathrm{s}$ for four different Courant numbers, compared to the reference solution. Since each time step introduces some diffusion, the LTS Roe scheme has a significant advantage over the ordinary Roe scheme for this case. As an example, Roe with $\mathcal{C} = 0.9$ will need roughly 56 times more time steps than LTS Roe with $\mathcal{C} = 50$, leading to more diffusion and a less accurate solution. In addition, fewer time steps means a shorter computational time, which adds to the advantages of using LTS Roe for this case. This seems to indicate that it is beneficial to increase the Courant number indefinitely, however this is not the case for all initial conditions, as we will see in the following sections.

4.2 Rarefaction and shock

This case starts with a discontinuity in both volume fraction and pressure, which leads to both a rarefaction wave and a shock wave. The initial condition is given by

$$u = 0, \tag{17a}$$

$$\alpha_\ell = \begin{cases} 0.1 & \text{if } x < 5000\,\mathrm{m}, \\ 0.9 & \text{if } x > 5000\,\mathrm{m}, \end{cases} \tag{17b}$$

$$p = \begin{cases} 20\,\mathrm{bar} & \text{if } x < 5000\,\mathrm{m}, \\ 60\,\mathrm{bar} & \text{if } x > 5000\,\mathrm{m}, \end{cases} \tag{17c}$$

$$T = T_{\mathrm{eq}}(p). \tag{17d}$$

Figures 2 and 3 show the pressure and liquid volume fraction, respectively, at $t = 25\,\mathrm{s}$. In this case, it is clear that a Courant number of 50 is far too high, and also $\mathcal{C} = 10.0$ leads to some oscillations. $\mathcal{C} = 3.0$ seems more appropriate and gives a solution similar to that achieved with $\mathcal{C} = 0.9$, although some oscillations are present. However, since $\mathcal{C} = 0.9$ requires about 3 times as many time steps, $\mathcal{C} = 3.0$ might well be the most appropriate choice when it comes to runtime per error.

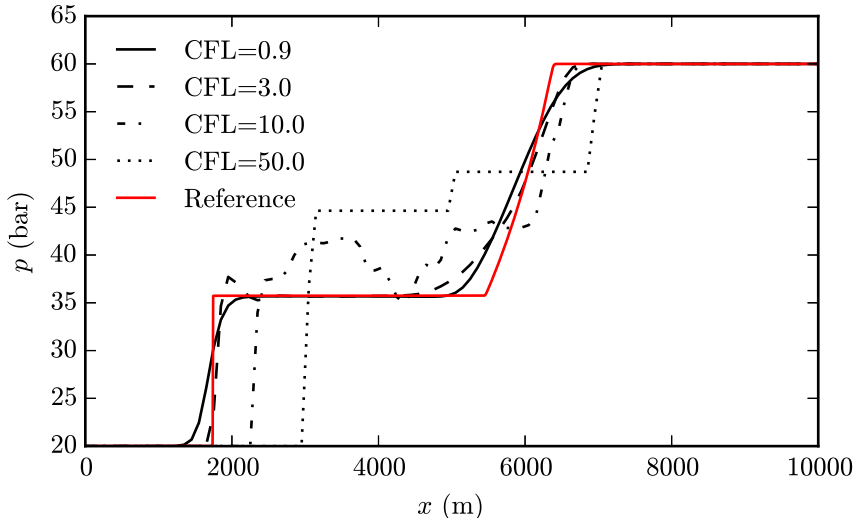


Figure 2: Pressure for at time $t = 25\,\mathrm{s}$ on a grid with 100 cells for initial conditions given in eq. (17).

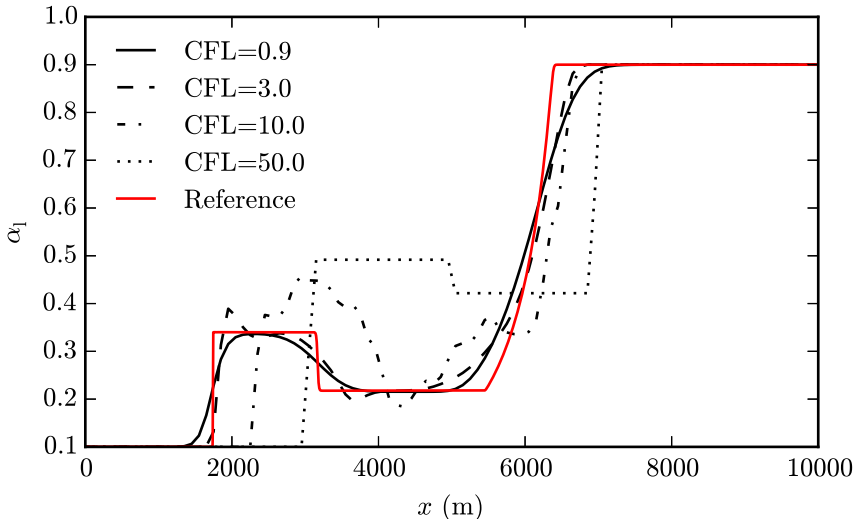
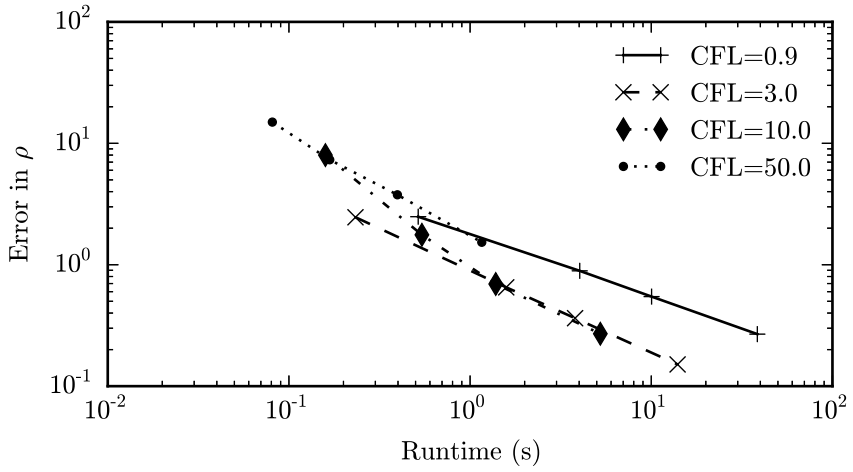
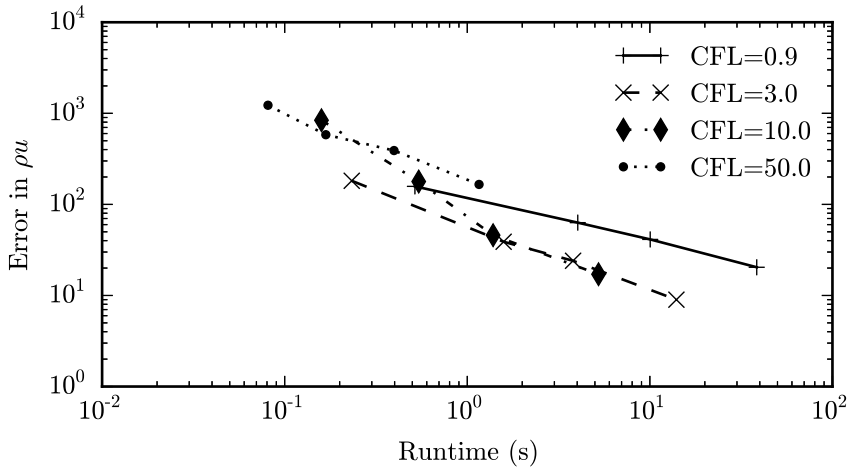


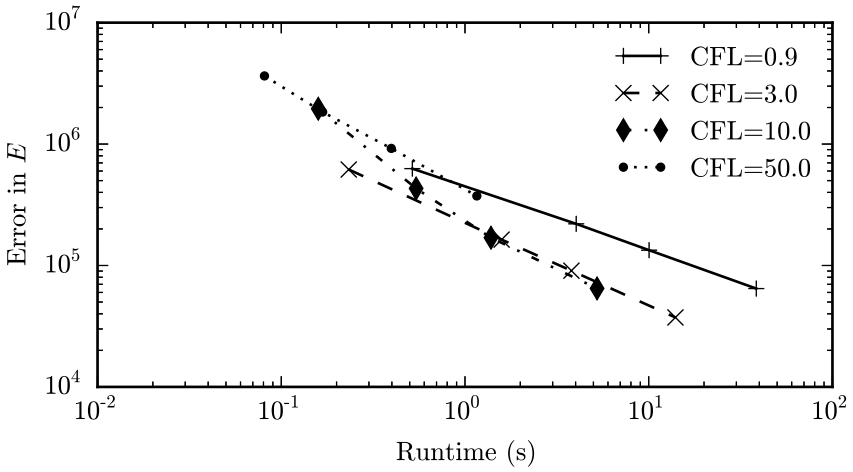
Figure 3: Pressure for at time $t = 25\,\mathrm{s}$ on a grid with 100 cells for initial conditions given in eq. (17).



(a) Error in ρ.



(b) Error in ρu.



(c) Error in E.

Figure 4: Errors plotted against runtime for initial conditions given in eq. (17).

Figure 4 shows the error in the three conservative variables as function of runtime, for the four different Courant numbers considered. It is apparent that a Courant number of 3 gives a significantly lower error for a given runtime compared to $\mathcal{C} = 0.9$. This picture seems to be consistent for all the three conserved variables. Even $\mathcal{C} = 50.0$ achieves runtimes that are comparable to that of $\mathcal{C} = 0.9$ for a given error, although this requires a grid with about 7 times more cells.

4.3 Liquid-gas shock tube

In this case, the left part of the domain contains liquid at $60\,\mathrm{bar}$, whereas the right contains gas at $20\,\mathrm{bar}$. This is expected to be more challenging than the previous cases, since it requires handling both single-phase and two-phase regions. More precisely, the initial state is given by

$$u = 0, \tag{18a}$$

$$\alpha_\ell = \begin{cases} 1 & \text{if } x < 5000\,\mathrm{m}, \\ 0 & \text{if } x > 5000\,\mathrm{m}, \end{cases} \tag{18b}$$

$$p = \begin{cases} 60\,\mathrm{bar} & \text{if } x < 5000\,\mathrm{m}, \\ 20\,\mathrm{bar} & \text{if } x > 5000\,\mathrm{m}, \end{cases} \tag{18c}$$

$$T = T_{\mathrm{eq}}(p). \tag{18d}$$

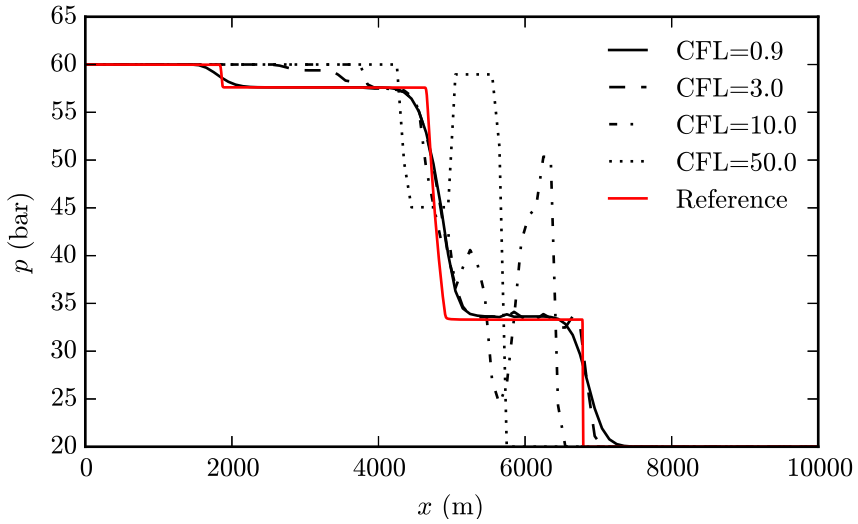
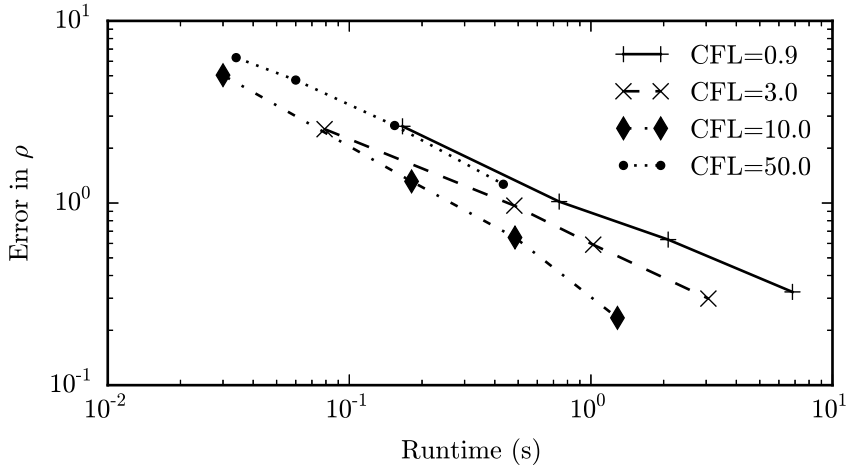


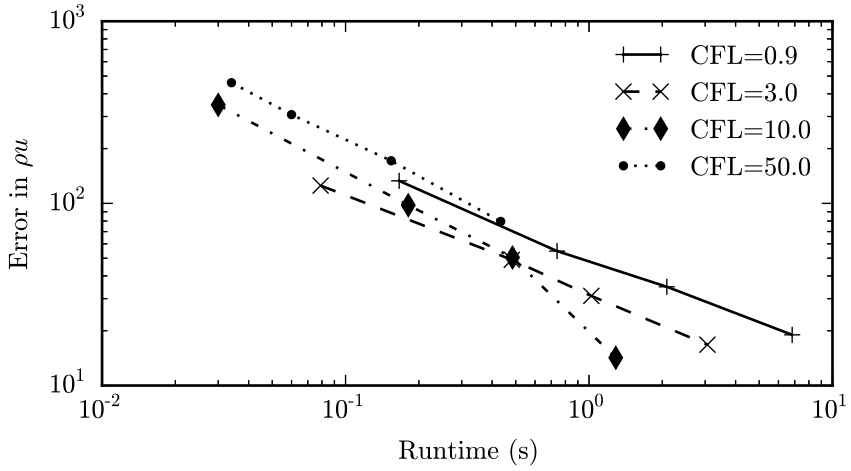
Figure 5: Pressure at time $t = 7\,\mathrm{s}$ on a grid with 100 cells for initial conditions given by eq. (18).

Figure 5 shows the pressure at $t = 7\,\mathrm{s}$ on a grid with 100 cells, compared to the reference solution. Both $\mathcal{C} = 10.0$ and $\mathcal{C} = 50.0$ lead to some significant oscillations, whereas $\mathcal{C} = 3.0$ performs quite well. However, it is apparent that the speed of the left-moving rarefaction wave (around $x = 2000\,\mathrm{m}$) is not resolved correctly and is notably different from the exact one.

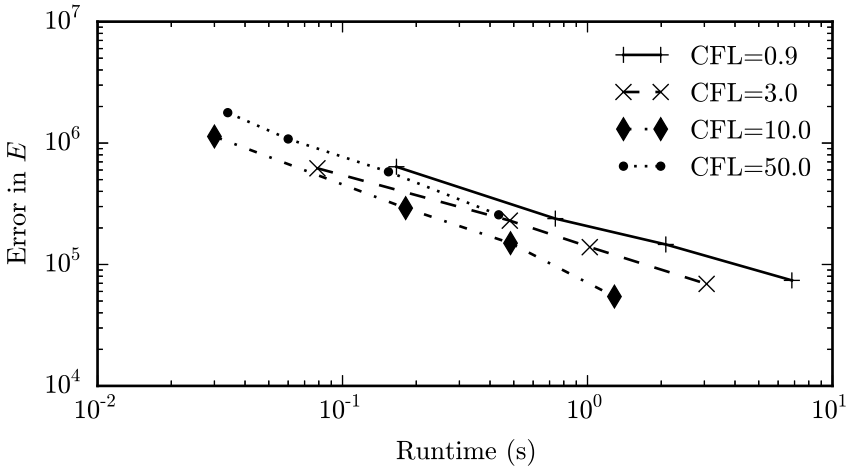
Figure 6 shows the errors in the three conserved variables as function of runtimes. Also for this rather challenging case, the LTS Roe scheme outperforms the ordinary Roe scheme for moderate Courant numbers. Even a Courant number as high as 10 seems to be the best choice



(a) Error in ρ.



(b) Error in ρu.



(c) Error in E.

Figure 6: Errors plotted against runtime for initial conditions given in eq. (18).

for this case when it comes to runtime, somewhat contrary to what one would expect when only considering fig. 5.

4.4 Smooth initial state

As a final test of how LTS Roe handles smoother initial states, we consider the following initial state:

$$\alpha_\ell = 0.6 \sin\left(\pi \frac{x}{10\,000\,\mathrm{m}}\right) + 0.2 \tag{19a}$$

$$u = \left(2.0 \cos\left(4\pi \frac{x}{10\,000\,\mathrm{m}}\right) + 2\right) \mathrm{m/s} \tag{19b}$$

$$p = \left(20 + \sin\left(1 + 4\frac{x}{10\,000\,\mathrm{m}}\right)\right) \mathrm{bar} \tag{19c}$$

$$T = T_{\mathrm{eq}}(p) \tag{19d}$$

Figure 7 shows the errors and runtimes for the three conserved variables. For this case, with rather smooth solutions, the LTS Roe scheme performs very well, and seems able to handle very large Courant numbers.

4.5 Discussion

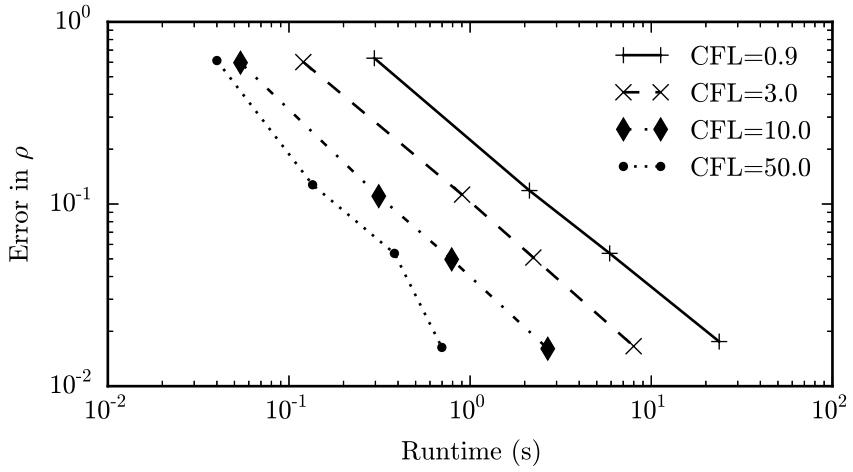
From the results presented above, the main impression is that the LTS Roe scheme with moderate Courant numbers works well for a variety of different cases. For smooth solutions, very high Courant numbers can be tolerated and still produce results with a lower error-to-runtime ratio. The presence of shocks and strong nonlinearities seems to reduce the maximal suitable Courant number, which is to be expected since we assume all waves interact linearly. Nevertheless a Courant number of around 3 proved to be a good choice in all cases. Interestingly, this is very similar to what was reported by LeVeque [11] in his related scheme.

We have presented the dependence of the error on runtime and Courant number for the various cases. It is important to note that the time needed for different steps of the solution procedure might vary significantly depending on the application. In our case, a large portion of the simulation time is spent solving the two-phase equilibrium problem, mentioned in section 3. This makes it especially beneficial to increase the Courant number, since this leads to fewer executions of the two-phase equilibrium solver. The same considerations are likely to be relevant for other applications that involve computationally expensive thermodynamics.

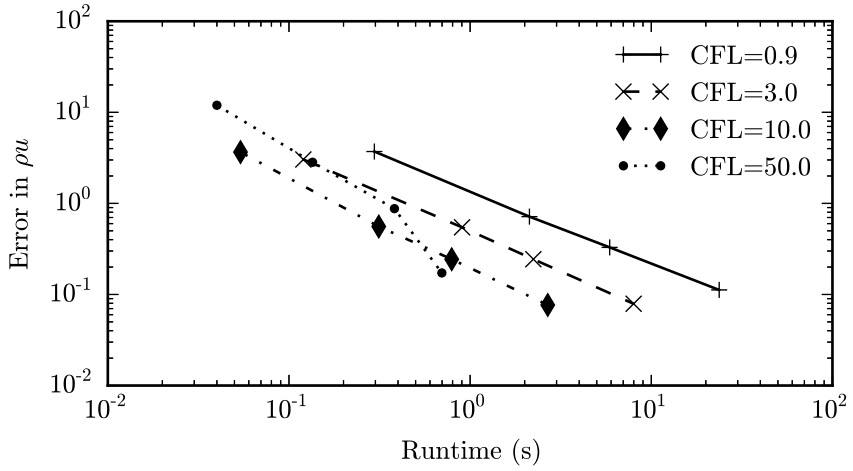
Simply put, the solution procedure for a single time step consists of three parts:

1. Calculating the waves emanating from each cell interface by solving the Riemann problems,
2. propagating the waves over one or more cells (depending on the Courant number),
3. and projecting the resulting solution back onto the grid.

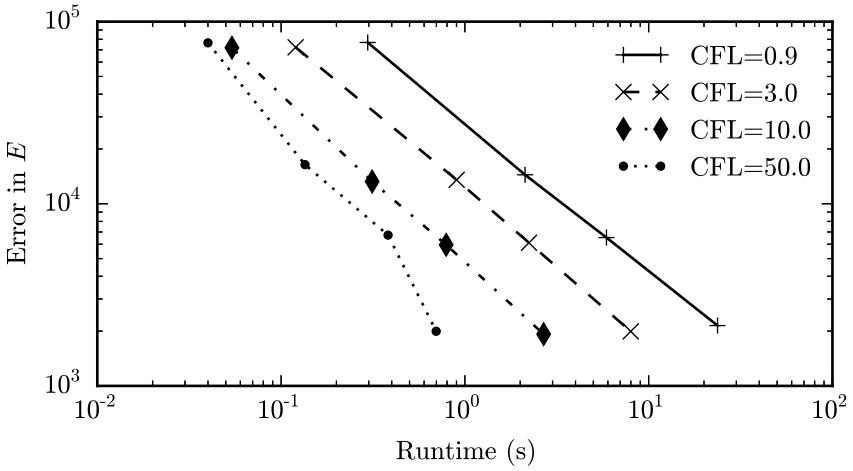
Each step involves a certain cost and introduces a certain error. In a typical application, step 1 incurs the highest costs, whereas steps 2 and 3 introduce the most errors. In step 2, the LTS Roe scheme is based on the assumption that the waves interact linearly, which leads to some errors. In addition, each projection in step 3 leads to projection errors. When increasing the



(a) Error in ρ.



(b) Error in ρu.



(c) Error in E.

Figure 7: Errors plotted against runtime for initial conditions given in eq. (19).

Courant number, we make a trade-off between increased error in step 2, reduced error (due to fewer executions) in step 3, and reduced cost (fewer executions) of step 1. The Courant number that gives the optimal trade-off (error versus runtime) will depend on the application.

5 SUMMARY

We have presented a two-phase flow model using the stiffened gas equation of state to describe the gas and liquid phases. The model was solved numerically with a novel large time step (LTS) Roe scheme. The LTS Roe scheme is based on the ordinary Roe scheme, but has a larger stencil which allows longer time steps while still being total-variation diminishing (TVD). Waves emanating from the Riemann problems are assumed to interact linearly, and propagated over one or more cells depending on the wave speed.

The LTS Roe scheme was applied to a number of test cases relevant to two-phase flow. It was shown that increasing the Courant number above 1 can give more accurate results for a certain runtime. For most cases, there is an optimal, finite Courant number that gives the optimal trade-off between runtime and error. Since the scheme is fully explicit, it is easily parallelizable, which is an advantage over implicit schemes.

As presented, the LTS Roe scheme is of first order, but we expect the scheme to be easily extendable to second order. Harten [4] expanded his similar scheme to second-order by applying to a modified flux using a minmod flux limiter. Furthermore, it is not straightforward to include source terms in a LTS scheme. A possibility is to use a fractional-step method such as Godunov or Strang splitting (see e.g. LeVeque [12, Ch. 17]), or to integrate the calculation of source terms directly with the wave propagation. This work verifies that the LTS Roe scheme, already successfully applied to shallow water flow [17, 19], aerodynamics [22] and Maxwell's equations [26], can be extended to more general equation systems.

Acknowledgements

This publication has been produced with support from the NORDICCS Centre, performed under the Top-level Research Initiative CO_2 Capture and Storage program, and Nordic Innovation. The authors acknowledge the following partners for their contributions: Statoil, Gassco, Norcem, Reykjavik Energy, and the Top-level Research Initiative (Project number 11029). The research has also received funding from the European Community's Seventh Framework Programme (FP7-ENERGY-20121-1-2STAGE) under grant agreement no. 308809 (The IMPACTS project). The authors acknowledge the project partners and the following funding partners for their contributions: Statoil Petroleum AS, Lundin Norway AS, Gas Natural Fenosa, MAN Diesel & Turbo SE and Vattenfall AB.

We would also like to thank our colleagues Peder Aursand, Tore Flåtten and Alexandre Morin for invaluable advice and discussions.

References

[1] S. Clerc. Numerical simulation of the homogeneous equilibrium model for two-phase flows. *Journal of Computational Physics*, 161(1):354–375, June 2000. doi: 10.1006/jcph.2000.6515.

[2] R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2):215–234, 1967. doi: 10.1147/rd.112.0215.

[3] A. Harten. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 49(3):357–393, Mar. 1983. doi: 10.1016/0021-9991(83)90136-5.

[4] A. Harten. On a large time-step high resolution scheme. *Mathematics of Computation*, 46 (174):379–399, 1986. doi: 10.2307/2007983.

[5] A. Harten, J. M. Hyman, P. D. Lax, and B. Keyfitz. On finite-difference approximations and entropy conditions for shocks. *Communications on Pure and Applied Mathematics*, XXIX:297–322, 1976. doi: 10.1002/cpa.3160290305.

[6] H. Holden, K.-A. Lie, and N. H. Risebro. An unconditionally stable method for the Euler equations. *Journal of Computational Physics*, 150:76–96, 1999. doi: 10.1006/jcph.1998.6164.

[7] IEA. *Energy Technology Perspectives*. 2014. ISBN 978-92-64-20801-8. doi: 10.1787/energy_tech-2014-en.

[8] P. Lax and B. Wendroff. Systems of conservation laws. *Communications on Pure and Applied Mathematics*, 2:217–237, 1960. doi: 10.1002/cpa.3160130205.

[9] R. J. LeVeque. Large time step shock-capturing techniques for scalar conservation laws. *SIAM Journal on Numerical Analysis*, 19(6):1091–1109, 1982. doi: 10.1137/0719080.

[10] R. J. Leveque. Convergence of a large time step generalization of Godunov's method for conservation laws. *Communications on Pure and Applied Mathematics*, 37(4):463–477, 1984. ISSN 1097-0312. doi: 10.1002/cpa.3160370405.

[11] R. J. LeVeque. A large time step generalization of Godunov's method for systems of conservation laws. *SIAM Journal on Numerical Analysis*, 22(6):1051–1073, 1985. doi: 10.1137/0722063.

[12] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, New York, 2002.

[13] S. Lindqvist, P. Aursand, T. Flåtten, and A. A. Solberg. Large time step TVD schemes for hyperbolic conservation laws, 2015. URL `http://arxiv.org/abs/1511.07027`. Submitted.

[14] H. Lund, T. Flåtten, and S. T. Munkejord. Depressurization of carbon dioxide in pipelines – models and methods. In J. Gale, C. Hendriks, and W. Turkenberg, editors, *GHGT-10 – 10th International Conference on Greenhouse Gas Control Technologies*, pages 2984–2991, Amsterdam, The Netherlands, Sept. 2011. IEAGHGT, Energy Procedia vol. 4. doi: 10.1016/j.egypro.2011.02.208.

[15] R. Menikoff. Empirical equations of state for solids. *Shock Wave Science and Technology Reference Library*, 2:143–188, 2007.

[16] R. Menikoff and B. J. Plohr. The Riemann problem for fluid flow of real materials. *Review of Modern Physics*, 61(1):75–130, Jan.–Mar. 1989. doi: 10.1103/RevModPhys.61.75.

[17] M. Morales-Hernández, P. García-Navarro, and J. Murillo. A large time step 1D upwind explicit scheme (CFL $>$ 1): Application to shallow water equations. *J. Comput. Phys.*, 231(19):6532–6557, Aug. 2012. ISSN 0021-9991. doi: 10.1016/j.jcp.2012.06.017.

[18] A. Morin, P. K. Aursand, T. Flåtten, and S. T. Munkejord. Numerical resolution of CO_2 transport dynamics. In *SIAM Conference on Mathematics for Industry: Challenges and Frontiers (MI09)*, San Francisco, CA, USA, Oct. 2009.

[19] J. Murillo, P. García-Navarro, P. Brufau, and J. Burguete. Extension of an explicit finite volume method to large time steps (CFL $>$ 1): Application to shallow water flows. *International Journal for Numerical Methods in Fluids*, 50(1):63–102, 2006. ISSN 1097-0363. doi: 10.1002/fld.1036.

[20] M. R. Norman, R. D. Nair, and F. H. M. Semazzi. A low communication and large time step explicit finite-volume solver for non-hydrostatic atmospheric dynamics. *Journal of Computational Physics*, 230(4):1567–1584, 2011. ISSN 0021-9991. doi: 10.1016/j.jcp.2010.11.022.

[21] Z. Qian and C.-H. Lee. A class of large time step Godunov schemes for hyperbolic conservation laws and applications. *Journal of Computational Physics*, 230(19):7418 – 7440, 2011. doi: 10.1016/j.jcp.2011.06.008.

[22] Z. Qian and C.-H. Lee. On large time step TVD scheme for hyperbolic conservation laws and its efficiency evaluation. *Journal of Computational Physics*, 231:7415–7430, 2012. doi: 10.1016/j.jcp.2012.07.015.

[23] J.-M. Qui and C.-W. Shu. Convergence of Godunov-type schemes for scalar conservation laws under large time steps. *SIAM Journal on Numerical Analysis*, 46(5):2211–2237, 2007. doi: 10.2307/25663061.

[24] P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43(2):357–372, Oct. 1981. doi: 10.1016/0021-9991(81)90128-5.

[25] H. Städtke. *Gasdynamic aspects of two-phase flow*. Wiley-VCH, Weinheim, Germany, 2006. ISBN 3-527-40578-X.

[26] R. J. Thompson and T. Moeller. A discontinuous wave-in-cell numerical scheme for hyperbolic conservation laws. *Journal of Computational Physics*, 299:404–428, 2015. ISSN 0021-9991. doi: 10.1016/j.jcp.2015.06.047.

[27] H. C. Yee. Linearized form of implicit TVD schemes for the multidimensional Euler and Navier-Stokes equations. *Computers & Mathematics with Applications*, 12(4–5, Part A):413–432, 1986. ISSN 0898-1221. doi: 10.1016/0898-1221(86)90172-0.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ENERGY CONSERVING TIME INTEGRATION BASED ON GALERKIN-VARIATIONAL INTEGRATORS WITH CONSTRAINTS

Matthias Bartelt[1] **and Michael Groß**[2]

Technische Universität Chemnitz, Professorship of applied mechanics and dynamics
Reichenhainer Straße 70, D-09126 Chemnitz
[1] e-mail: matthias.bartelt@mb.tu-chemnitz.de [2] e-mail: michael.gross@mb.tu-chemnitz.de

Keywords: Variational integrator, energy conservation, finite element method, higher-order time integration, Lagrange multiplier method

Abstract. *Often, Lagrange's formalism based on the Lagrangian is the preferred way in order to derive equations of motion for a mechanical system. Therefore, the first step is the formulation of the total kinetic and total potential energy. In this formalism, holonomic constraints can be taken into account by using the Lagrange multiplier method, and external and dissipative forces can be included variationally consistently by means of the Lagrange-D'Alembert principle. In this way, we are directly arrive at a weak formulation of the equations of motion without using a strong form and integration by parts. The variational time integration method is now based on the discretization of the action functional and the discrete variational calculus. This method leads to time stepping schemes called variational integrators.*
In this paper, we show that variational integrators based on higher-order shape functions and quadrature rules leads to a higher-order time approximation, which preserve the total linear and total angular momentum of the mechanical system. A further topic of the paper is a new implementation of the Lagrange multiplier method for holonomic constraints in the higher-order variational integrator, in order to compute bearing forces by means of Lagrange multipliers.
Usually, variational integrators show an excellent long time behavior and bind the total energy error per time step, but are not able to preserve the total energy exactly. Therefore, we introduce a discrete gradient of the total potential energy in the variational integrator to preserve the total energy. We show different discrete gradients with different results for numerical stability. We can show that these time stepping schemes preserve the total linear momentum, total angular momentum as well as the total energy for every time step size.
As numerical examples, we show motions of three-dimensional continua, which are discretized in space by the finite element method. We start with a free flying hyper-elastic rotor to show the preservation of total linear and total angular momentum in combination with higher-order accuracy in time. In the second example, we consider a hyper-elastic beam and apply the Lagrange multiplier method in order to calculate the bearing forces at fixed nodes. In the last example, the energy conservation is shown for different discrete gradients.

1 INTRODUCTION

This paper shows the application of a variational formulation in time to create an *energy-conserving* variational integrator of higher order accuracy for finite element models. A three-dimensional body will be discretized with linear finite elements in space. We also apply DIRICHLET boundaries and show how these can be included in the variational principle. For this, we use LAGRANGE multiplier with a special NEWTON-COTÊS approximation. The calculation of the bearing forces will be described, which is needed for the calculation of the balance of total angular momentum. A *standard* variational integrator is total linear and total angular momentum conserving, however does not preserve the total energy of a discrete conservative system. But, the introduction of the enhanced derivative or gradient, respectively, in Reference [1] in the variational integrator leads to a variational based time integration algorithm, which preserves the discrete balance of total energy.

First of all, the introduction of the enhanced gradient is consistent with the LAGRANGE-D'ALEMBERT principle and therefore generates an extended variational integrator. This knowledge can be derived from the generalization of the discrete derivative or gradient, respectively, in Reference [8] to higher order schemes [1]. For higher order schemes, the discrete gradient of the total energy is always split into two terms of a sum. The first term is a conservative part, which arises in a variational integrator from the discretized Lagrangian of the dynamical system. The second term as extension of the conservative part can be considered as a special non-conservative extension of the dynamical system. The special feature of this term is, that it does not lead to a physically based energy dissipation or a physically based external work, but rather to an algorithmic energy conservation in the discrete variational system and to a zero term in the continuous variational system. In Reference [1] is also shown, that both parts can be derived from a separate constraint variational problem, which can be concatenated with the variational formulation of the dynamical system.

Another interesting point of view is to recognize in the second term of the enhanced gradient an energy error controller. In Reference [11], there is described a so-called *affine Hamiltonian control system* with an extended Hamiltonian

$$H(\mathbf{q},\mathbf{p},\mathbf{u}) = H_0(\mathbf{q},\mathbf{p}) - \sum_{j=1}^{n_{\mathrm{con}}} H_j(\mathbf{q},\mathbf{p})\, u_j \tag{1}$$

which depends only linearly or affine on the control variables u_j, $j = 1,\ldots,n_{\mathrm{con}}$. The function H_0 is called the internal Hamiltonian of the dynamical system, and the functions H_j are called coupling Hamiltonians. Accordingly, the natural outputs y_j, $j = 1,\ldots,n_{\mathrm{con}}$ of the Hamiltonian system are given by

$$y_j = \frac{\partial H}{\partial u_j} \tag{2}$$

Taking into account Hamilton's principle, we arrive in the continuous setting at the affine Hamiltonian control system

$$\dot{\mathbf{q}} = \frac{\partial H_0}{\partial \mathbf{p}} - \sum_{j=1}^{n_{\mathrm{con}}} \frac{\partial H_j}{\partial \mathbf{p}}\, u_j \tag{3}$$

$$\dot{\mathbf{p}} = -\frac{\partial H_0}{\partial \mathbf{q}} + \sum_{j=1}^{n_{\mathrm{con}}} \frac{\partial H_j}{\partial \mathbf{q}}\, u_j \tag{4}$$

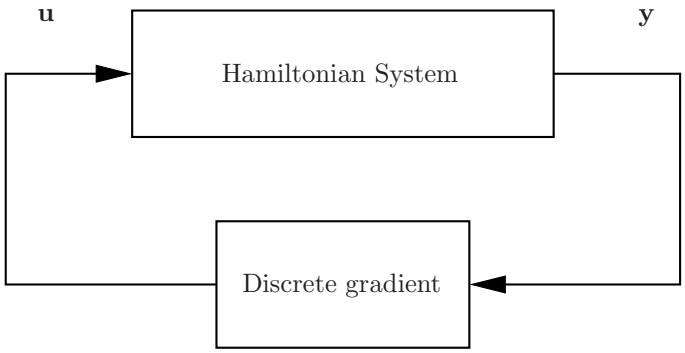


Figure 1: Discrete gradient as energy error controller of a time integrator

with the corresponding total energy balance

$$\frac{\mathrm{d}H_0}{\mathrm{d}t} = \sum_{j=1}^{n_{\mathrm{con}}} \dot{y}_j \, u_j \tag{5}$$

Hence, in the continuous setting the input variables have to vanish for total energy conservation, but in the discrete setting the sum balance the energy error of the time integrator. Comparing this Hamiltonian formulation with that in Reference [1], it is obvious that the control inputs u_j are the components of the second term of the enhanced gradient, following from a costoptimized constraint optimization problem and leading to an energy consistent algorithmic *closed loop control problem*.

Consequently, the enhanced gradient as higher order accurate discrete gradient can be seen at least in two ways as a consistent extension of a variational based time integrator. In this work, the first method of using the LAGRANGE-D'ALEMBERT principle is shown in a continuous as well as discrete setting.

2 FINITE ELEMENT APPROXIMATION

We choose a standard spatial finite element discretization for the solid continuum. First we define a model in the reference configuration $\mathcal{B}_0$ at time $t = 0$. The deformed model is denoted by the configurations $\mathcal{B}_t$ for $t > 0$. The postion vector in configuration $\mathcal{B}_0$ is described by the vector $\mathbf{X} \in \mathbb{R}^n$. For the configuration $\mathcal{B}_t$, the new position vector depends on time t with $\mathbf{x}(t) \in \mathbb{R}^n$. The index n means the number of degrees of freedom for the three-dimensional discretized continuum.

$$\mathbf{X} \in \mathbb{R}^n \text{ in } \mathcal{B}_0 \text{ and } \mathbf{x}(t) \in \mathbb{R}^n \text{ in } \mathcal{B}_t \tag{6}$$

With an application of the isoparametric concept [3], we define a function $\boldsymbol{\varphi}$. This makes it possible to calculate the actual postion from the reference configuration and the actual time t. The actual position depends on a linear mapping

$$\mathbf{x}(t) = \boldsymbol{\varphi}(\mathbf{X}, t). \tag{7}$$

The position vectors are approximated in space with linear LAGRANGE-polynomials N^A on the reference element with coordination vector $\boldsymbol{\xi}$. This finite element is also known as C3D8 cube.

$$\mathbf{X}^{e,\mathrm{h}} = \sum_{A=1}^{n_{\mathrm{s}}} N^A(\boldsymbol{\xi}) \mathbf{X}^{e,A} \tag{8}$$

$$\mathbf{x}^{e,\mathrm{h}} = \sum_{A=1}^{n_\mathrm{s}} N^A(\boldsymbol{\xi})\mathbf{q}^{e,A} \tag{9}$$

The deformation gradient is defined by the derivative of the function $\boldsymbol{\varphi}$ with respect to $\mathbf{X}$,

$$\mathbf{F}(\mathbf{X},t) = \nabla_{\mathbf{X}}\boldsymbol{\varphi}(\mathbf{X},t) \tag{10}$$

with

$$\mathbf{F}^e = \sum_{A=1}^{n_\mathrm{s}} \mathbf{q}^{e,A} \otimes \left((\mathbf{J}^e)^{-\mathrm{T}} \cdot \frac{\partial N^A(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right). \tag{11}$$

and

$$\mathbf{J}^e = \sum_{A=1}^{n_\mathrm{s}} \mathbf{X}^{e,A} \otimes \frac{\partial N^A(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \tag{12}$$

The deformation gradient $\mathbf{F}$ provides a relationship between tangent vectors in the reference and actual configuration of a continuum. For the description of the deformation measurement we choose the right-CAUCHY-GREEN tensor

$$\mathbf{C} = \mathbf{F}^{\mathrm{T}}\mathbf{F} \tag{13}$$

The third invariant of $\mathbf{F}$ can also be defined in $\mathbf{C}$ as

$$I_3(\mathbf{F}) = \sqrt{\det(\mathbf{C})}. \tag{14}$$

We also define the derivative of the tensor $\mathbf{C}$ with respect to $\mathbf{q}$ as the linearized strain operator

$$\mathbb{B}^{e,A} = 2\frac{\partial \mathbf{C}^e}{\partial \mathbf{q}^e} = 2\nabla_{\mathbf{X}} N^A \otimes \mathbf{F}^{e^{\mathrm{T}}}. \tag{15}$$

2.1 MATERIAL FORMULATION

The used material model is described by the following free energy function see also [5] and [10].

$$\psi = \frac{\mu}{2}\left[\mathrm{tr}(\mathbf{C}) - n_{\mathrm{dim}} - 2\ln(I_3(\mathbf{C}))\right] + \psi^{\mathrm{vol}} \tag{16}$$

$$\psi^{\mathrm{vol}} = \frac{\lambda}{2}\left[\ln(I_3(\mathbf{C}))^2 + (I_3(\mathbf{C}) - 1)^2\right] \tag{17}$$

The constants λ and μ are called LAMÉ-parameters. In the three-dimensional case, n_{dim} is set to three. This matrial model, in combination with the formulation of the right-CAUCHY-GREEN tensor, allows us to decribe large deformations for the physical model.

3 THE LAGRANGIAN

The LAGRANGE function, defined by the kinetic energy T^{kin} and potential energy V^{int}, is given by

$$\mathcal{L} = T^{\mathrm{kin}} - V^{\mathrm{int}} \tag{18}$$

As potential energy $V^{\text{int},e}$, we define the integral over free energy function ψ^e for every element and the gravity with respect to the reference configuration $\mathcal{B}_0$. The kinetic energy $T^{\text{kin},e}$ for one element depends on the mass matrix $\mathbb{M}^e$ and the velocity $\dot{\mathbf{q}}^e$.

$$V^{\text{int},e} = \int_{\mathcal{B}_0^e} \rho_0 \psi^e(\mathbf{C}^e(\mathbf{q}^e)) + \rho_0 \mathbf{g}^e \cdot \mathbf{q}^e \mathrm{d}V \tag{19}$$

$$T^{\text{kin}} = \frac{1}{2} \int_{\mathcal{B}_0^e} \rho_0 \dot{\mathbf{q}} \cdot \dot{\mathbf{q}} \tag{20}$$

$$T^{\text{kin},e} = \frac{1}{2} \dot{\mathbf{q}}^e \cdot \mathbb{M}^e \cdot \dot{\mathbf{q}}^e \tag{21}$$

The gravitational vector $\mathbf{g}^e$ is defined by $\mathbf{g}^e = g[0, -e_y, 0]^{\mathrm{T}}$ with g as the gravitational constant.

3.1 APPROXIMATION IN TIME

The postion vector $\mathbf{q}$ and velocity vector $\dot{\mathbf{q}}$ are discretized in the following way:

$$\mathbf{q}^{\mathrm{h}} = \sum_{i=0}^{n_{\mathrm{p}}} M_i(\alpha) \mathbf{q}_{\frac{i}{n_{\mathrm{p}}}} \qquad \mathbf{q}^{\mathrm{d}} = \sum_{i=0}^{n_{\mathrm{p}}} M_i^{'}(\alpha) \mathbf{q}_{\frac{i}{n_{\mathrm{p}}}} \qquad i = 0 \ldots n_{\mathrm{p}} \tag{22}$$

n_{p} is the number of polynomial degree and M represents the LAGRANGE polynomials in the interval from zero to one. For this approximation, we need specific quadrature points $\alpha_j, j = 1, \ldots, n_{\mathrm{q}}$. In this paper, we only use GAUSSIAN quadrature points and LAGRANGE polynomials with equidistant points in the interval from zero to one. Note that, it is also possible to choose other quadrature points or other shape functions in time. The approximated velocity vector $\dot{\mathbf{q}}$ is defined by the derivative of the position vector with respect to the time t. This derivative will be replaced by a derivative of $M(\alpha)$ with respect to α.

$$\dot{\mathbf{q}}(t) = \sum_{i=0}^{n_{\mathrm{p}}} \frac{\mathrm{d}M(\alpha(t))}{\mathrm{d}t} \mathbf{q}_{\frac{i}{n_{\mathrm{p}}}} = \sum_{i=0}^{n_{\mathrm{p}}} \frac{1}{h_n} \frac{\partial M(\alpha)}{\partial \alpha} \mathbf{q}_{\frac{i}{n_{\mathrm{p}}}} = \sum_{i=0}^{n_{\mathrm{p}}} \frac{1}{h_n} M_i^{'}(\alpha) \mathbf{q}_{\frac{i}{n_{\mathrm{p}}}} = \frac{1}{h_n} \mathbf{q}^{\mathrm{d}} \tag{23}$$

3.2 THE DISCRETE STATE SPACE DEFINITION

We divide one time interval with the size h_n into different micro time steps. The number of micro time steps depends on the polynomial degree of approximation in time. This will be applied to the position vector $\bar{\mathbf{q}}_{k+1}$ and the LAGRANGE-multiplier vector $\bar{\boldsymbol{\lambda}}_{k+1}$.

$$\bar{\mathbf{q}}_{k+1} = \left[\mathbf{q}_k, \mathbf{q}_{k+\frac{1}{n_{\mathrm{p}}}}, \mathbf{q}_{k+\frac{2}{n_{\mathrm{p}}}}, \ldots, \mathbf{q}_{k+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \mathbf{q}_{k+1} \right]^{\mathrm{T}} \tag{24}$$

$$\bar{\boldsymbol{\lambda}}_{k+1} = \left[\boldsymbol{\lambda}_k, \boldsymbol{\lambda}_{k+\frac{1}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{k+\frac{2}{n_{\mathrm{p}}}}, \ldots, \boldsymbol{\lambda}_{k+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{k+1} \right]^{\mathrm{T}}$$

The total number of all state variables is summarized in $\bar{\mathbf{q}}^{\text{tot}}$ and $\bar{\boldsymbol{\lambda}}^{\text{tot}}$. The number of all macro time steps is equal to n_{t}.

$$\bar{\mathbf{q}}^{\text{tot}} = \left[\mathbf{q}_{0+\frac{0}{n_{\mathrm{p}}}}, \mathbf{q}_{0+\frac{1}{n_{\mathrm{p}}}}, \mathbf{q}_{0+\frac{2}{n_{\mathrm{p}}}}, \ldots, \mathbf{q}_{0+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \mathbf{q}_{1+\frac{0}{n_{\mathrm{p}}}}, \ldots, \mathbf{q}_{n_{\mathrm{t}}-1+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \mathbf{q}_{n_{\mathrm{t}}} \right]^{\mathrm{T}} \tag{25}$$

$$\bar{\boldsymbol{\lambda}}^{\text{tot}} = \left[\boldsymbol{\lambda}_{0+\frac{0}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{0+\frac{1}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{0+\frac{2}{n_{\mathrm{p}}}}, \ldots, \boldsymbol{\lambda}_{0+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{1+\frac{0}{n_{\mathrm{p}}}}, \ldots, \boldsymbol{\lambda}_{n_{\mathrm{t}}-1+\frac{n_{\mathrm{p}}-1}{n_{\mathrm{p}}}}, \boldsymbol{\lambda}_{n_{\mathrm{t}}} \right]^{\mathrm{T}}$$

The continuous LAGRANGIAN will be approximated with discretized variables, see below.

3.3 THE DISCRETE LAGRANGIAN

$$\tilde{\mathcal{L}}_{\mathrm{d}} = \int_{t_k}^{t_{k+1}} \frac{1}{2}\dot{\mathbf{q}}^{\mathrm{h}} \cdot \mathbb{M} \cdot \dot{\mathbf{q}}^{\mathrm{h}} \mathrm{d}t - \int_{t_k}^{t_{k+1}} \int_{\mathcal{B}_0} \rho_0 \psi(\mathbf{C}(\mathbf{q}^{\mathrm{h}})) + \rho_0 \mathbf{g} \cdot \mathbf{q}^{\mathrm{h}} \mathrm{d}V \mathrm{d}t - \int_{t_k}^{t_{k+1}} \boldsymbol{\lambda} \cdot \boldsymbol{\Phi}\left(\mathbf{q}\right) \mathrm{d}t \tag{26}$$

First we define a discrete functional $\mathcal{L}_{\mathrm{d},k+1}$ without the LAGRANGE multiplier vector $\boldsymbol{\lambda}$.

$$\mathcal{L}_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}) = h_n \sum_{m=1}^{n_{\mathrm{q}}} w_m \mathcal{L}(\mathbf{q}_{k+1}, \dot{\mathbf{q}}_{k+1}) \tag{27}$$

The reason is, we want to fulfill constraints at the micro time steps and not at the quadrature points. Therefore we apply a NEWTON-COTÊS quadrature for the constraint terms. The application of the NEWTON-COTÊS quadrature, with an equidistant choice of quadrature points, decouples the following equations and defines the LAGRANGE multiplier at the micro time steps.

$$\tilde{\mathcal{L}}_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}, \bar{\boldsymbol{\lambda}}_{k+1}) = \mathcal{L}_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}) - h_n \sum_{i=0}^{n_{\mathrm{q}}} w_{\mathrm{c}} M_{\mathrm{c},i}(\alpha_{\mathrm{c}}) \tilde{\boldsymbol{\lambda}}_{k+\frac{i}{n_{\mathrm{p}}}} \cdot \boldsymbol{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_{\mathrm{p}}}}\right) \tag{28}$$

The shape functions in time for the NEWTON-COTÊS quadrature are always one for the quadrature points and zero for all other shape functions.

$$\sum_{i=0}^{n_{\mathrm{q}}} w_{\mathrm{c}} M_{\mathrm{c},i}(\alpha_{\mathrm{c}}) \tilde{\boldsymbol{\lambda}}_{k+\frac{i}{n_{\mathrm{p}}}} = \sum_{i=0}^{n_{\mathrm{q}}} w_{\mathrm{c}} \tilde{\boldsymbol{\lambda}}_{k+\frac{i}{n_{\mathrm{p}}}} \tag{29}$$

We also define a new LAGRANGE multiplier vector $\boldsymbol{\lambda}_{k+\frac{i}{n_{\mathrm{p}}}}$ which includes the weight from the quadrature.

$$\tilde{\mathcal{L}}_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}, \bar{\boldsymbol{\lambda}}_{k+1}) = \mathcal{L}_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}) - h_n \sum_{i=0}^{n_{\mathrm{q}}} \boldsymbol{\lambda}_{k+\frac{i}{n_{\mathrm{p}}}} \cdot \boldsymbol{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_{\mathrm{p}}}}\right) \tag{30}$$

We build the discrete action sum $\mathcal{S}_{\mathrm{d}}$ as a sum over all time steps.

$$\begin{aligned} \delta \mathcal{S}_{\mathrm{d}} &= \delta \sum_{k=0}^{n_{\mathrm{t}}-1} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \\ &= \frac{\partial}{\partial \bar{\mathbf{q}}^{\mathrm{tot}}} \left(\sum_{k=0}^{n_{\mathrm{t}}-1} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \right) \cdot \delta \bar{\mathbf{q}}^{\mathrm{tot}} + \frac{\partial}{\partial \bar{\boldsymbol{\lambda}}^{\mathrm{tot}}} \left(\sum_{k=0}^{n_{\mathrm{t}}-1} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \right) \cdot \delta \bar{\boldsymbol{\lambda}}^{\mathrm{tot}} \end{aligned} \tag{31}$$

The first variation with respect to $\bar{\mathbf{q}}^{\mathrm{tot}}$ gives us:

$$\begin{aligned} &\frac{\partial}{\partial \bar{\mathbf{q}}^{\mathrm{tot}}} \left(\sum_{k=0}^{n_{\mathrm{t}}-1} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \right) \cdot \delta \bar{\mathbf{q}}^{\mathrm{tot}} \\ &= \sum_{k=0}^{n_{\mathrm{t}}-1} \left[\mathrm{D}_1 \tilde{\mathcal{L}}_{\mathrm{d},k+1} \cdot \delta \mathbf{q}_k + \sum_{i=1}^{n_{\mathrm{p}}-1} \left(\mathrm{D}_{1+i} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \cdot \delta \mathbf{q}_{k+\frac{i}{n_{\mathrm{p}}}} \right) + \mathrm{D}_{1+n_{\mathrm{p}}} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \cdot \delta \mathbf{q}_{k+1} \right] \\ &= \sum_{k=0}^{n_{\mathrm{t}}-1} \sum_{i=1}^{n_{\mathrm{p}}-1} \left(\mathrm{D}_{1+i} \tilde{\mathcal{L}}_{\mathrm{d},k+1} \cdot \delta \mathbf{q}_{k+\frac{i}{n_{\mathrm{p}}}} \right) + \sum_{k=1}^{n_{\mathrm{t}}-1} \left(\mathrm{D}_1 \tilde{\mathcal{L}}_{\mathrm{d},k+1} + \mathrm{D}_{1+n_{\mathrm{p}}} \tilde{\mathcal{L}}_{\mathrm{d},k} \right) \cdot \delta \mathbf{q}_k \\ &+ \mathrm{D}_1 \tilde{\mathcal{L}}_{\mathrm{d},1} \cdot \delta \mathbf{q}_0 + \mathrm{D}_{1+n_{\mathrm{p}}} \tilde{\mathcal{L}}_{\mathrm{d},n_{\mathrm{t}}} \cdot \delta \mathbf{q}_{n_{\mathrm{t}}} \end{aligned} \tag{32}$$

By definition, $\delta\mathbf{q}_0 = \delta\mathbf{q}_{n_\mathrm{t}} = \mathbf{0}$ are always zero. We also define the derivative $\mathrm{D}_{1+i}\tilde{\mathcal{L}}_{\mathrm{d},k+1}$ in the following way:

$$\mathrm{D}_{1+i}\tilde{\mathcal{L}}_{\mathrm{d},k+1} = \frac{\partial\tilde{\mathcal{L}}_{\mathrm{d},k+1}}{\partial\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}} \quad \text{with} \quad i \in \{0,\dots n_\mathrm{p}\} \tag{33}$$

We collect the terms for similar variations and get:

$$\mathbf{0} = \mathrm{D}_{1+n_\mathrm{p}}\tilde{\mathcal{L}}_{\mathrm{d},k} + \mathrm{D}_1\tilde{\mathcal{L}}_{\mathrm{d},k+1} \tag{34}$$

$$\mathbf{0} = \mathrm{D}_{1+i}\tilde{\mathcal{L}}_{\mathrm{d},k+1} \quad \forall\, i = 1,\dots,n_\mathrm{p}-1 \tag{35}$$

In the same way, we define the vector of linear momentum. The vector of linear momentum only exists at every macro time step.

$$\bar{\mathbf{p}}^{\mathrm{tot}} = \left[\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n_\mathrm{t}-1}, \mathbf{p}_{n_\mathrm{t}}\right]^{\mathrm{T}} \tag{36}$$

The discrete EULER-LAGRANGE equations are invariant up to rigid body motions. Therefore, we can reformulate the EULER-LAGRANGE equations in the so called position-momentum form [7].

$$\mathbf{p}_k = \mathrm{D}_{1+n_\mathrm{p}}\tilde{\mathcal{L}}_{\mathrm{d},k} = -\mathrm{D}_1\tilde{\mathcal{L}}_{\mathrm{d},k+1} \tag{37}$$

$$-\mathbf{p}_k = \mathrm{D}_1\tilde{\mathcal{L}}_{\mathrm{d},k+1} \tag{38}$$

$$\mathbf{p}_{k+1} = \mathrm{D}_{1+n_\mathrm{p}}\tilde{\mathcal{L}}_{\mathrm{d},k+1} \tag{39}$$

The variation with respect to $\bar{\boldsymbol{\lambda}}^{\mathrm{tot}}$ gives:

$$\frac{\partial}{\partial\bar{\boldsymbol{\lambda}}^{\mathrm{tot}}}\left(\sum_{k=0}^{n_\mathrm{t}-1}\tilde{\mathcal{L}}_{\mathrm{d},k+1}\right)\cdot\delta\bar{\boldsymbol{\lambda}}^{\mathrm{tot}} = -\sum_{k=0}^{n_\mathrm{t}-1}h_n\sum_{i=0}^{n_\mathrm{p}}\boldsymbol{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}\right)\cdot\delta\bar{\boldsymbol{\lambda}}_{k+\frac{i}{n_\mathrm{p}}} \tag{40}$$

As a result, the constraints have to be fulfilled at every micro time step.

$$\boldsymbol{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}\right)\Big|_{i=0}^{n_\mathrm{p}-1} = \mathbf{0} \tag{41}$$

In order to guarantee the symplecticity construction for our time step scheme, the following condition has to be implemented to fulfill the constraints on velocity level at the last micro time step in every macro time step [2], [7].

$$\frac{\mathrm{d}\boldsymbol{\Phi}\left(\mathbf{q}_{k+1}\right)}{\mathrm{d}t} = \boldsymbol{\Psi}\left(\mathbf{q}_{k+1},\mathbf{p}_{k+1}\right) = \mathbf{0} \tag{42}$$

With the equations below we get the discrete EULER-LAGRANGE equations in the position-momentum form. In these equations we first calculate implicitly the unknown values of the postion vector for every micro time step and the unknown LAGRANGE multiplier in Equation (43). In the "update" step we calculate also implicitly the unknown LAGRANGE multiplier

and the linear momentum vector at the end of the micro time step, Equation (44) [6], [9].

$$
\begin{aligned}
-\mathbf{p}_k &= \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_k} - h_n \left(\frac{\partial \mathbf{\Phi}(\mathbf{q}_k)}{\partial \mathbf{q}_k} \right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_k \qquad (43)\\
\mathbf{0} &= \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}} - h_n \left(\frac{\partial \mathbf{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}\right)}{\partial \mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}} \right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_{k+\frac{i}{n_\mathrm{p}}} \Bigg|_{i=1}^{n_\mathrm{p}-1}\\
\mathbf{0} &= \mathbf{\Phi}(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}) \Big|_{i=1}^{n_\mathrm{p}-1}\\
\mathbf{p}_{k+1} &= \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_{k+1}} - h_n \left(\frac{\partial \mathbf{\Phi}(\mathbf{q}_{k+1})}{\partial \mathbf{q}_{k+1}} \right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_{k+1} \qquad (44)\\
\mathbf{0} &= \mathbf{\Psi}\left(\mathbf{q}_{k+1}, \mathbf{p}_{k+1}\right)
\end{aligned}
$$

The derivative of the discrete LAGRANGE function with respect to $\mathbf{q}$ for every element can be written as

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \bar{\mathbf{q}}^e_{k+1}} &= \sum_{i=1}^{n_\mathrm{q}} w_i \frac{\partial \left(\frac{1}{2} \mathbf{q}^{\mathrm{d}^e}(\alpha_i) \cdot \mathbb{M}^e \cdot \mathbf{q}^{\mathrm{d}^e}(\alpha_i) \right)}{\partial \bar{\mathbf{q}}^e_{k+1}} \qquad (45)\\
&- \sum_{i=1}^{n_\mathrm{q}} w_i \frac{\partial \left(\int\limits_{\mathcal{B}_0^e} \left[\rho_0 \psi^e \left(\tilde{\mathbf{C}}^e(\mathbf{q}^{\mathrm{h}^e}(\alpha_i)) \right) + \rho_0 \mathbf{g}^e \cdot \mathbf{q}^{\mathrm{h}^e} \right] \mathrm{d}V \right)}{\partial \bar{\mathbf{q}}^e_{k+1}}.
\end{aligned}
$$

4 CONSERVATION PROPERTIES

Every variational integrator conserves the balance of total linear momentum and the balance of total angular momentum. Without any constrains and no gravitational force, it is easy to calculate these conservation properties, but we want to show how these properties can be calculated and show that they are preserved, and hence constraints and a gravity force are included in the simulation.

4.1 BALANCE OF TOTAL LINEAR MOMENTUM

In the case without a DIRICHLET boundary, the conservation of the balance of total linear momentum in the x, y, z-direction can be calculated by the sum over all n_{nodes} nodes:

$$
\underbrace{\sum_{i=1}^{n_{\text{nodes}}} \left[\mathbf{p}_{j,i_{k+1}} - \mathbf{p}_{j,i_k} \right]}_{P_{i+1} - P_i} = 0 \qquad \forall j = [x, y, z] \qquad (46)
$$

In the case with boundary nodes, the conservation of the balance of total linear momentum in the x, y, z-direction can be calculated by the sum over all n_{nodes} nodes and the sum over the

LAGRANGE multiplier vector:

$$\sum_{i=1}^{n_{\text{nodes}}} \left[\mathbf{p}_{j,i_{k+1}} - \mathbf{p}_{j,i_k}\right] + h_n \sum_{i=1}^{n_{\text{nodes}}} \sum_{m=0}^{n_{\text{p}}} \boldsymbol{\lambda}_{j,i_{k+\frac{m}{n_{\text{p}}}}} = 0 \qquad \forall j = [x,y,z] \quad (47)$$

$$\underbrace{\sum_{i=1}^{n_{\text{nodes}}} \left[\mathbf{p}_{j,i_{k+1}} - \mathbf{p}_{j,i_k}\right]}_{P_{i+1}-P_i} + \underbrace{h_n \sum_{i=1}^{n_{\text{nodes}}} \sum_{m=0}^{n_{\text{p}}} w_{\text{c}} M_{\text{c},m}(\alpha_{\text{c}}) \tilde{\boldsymbol{\lambda}}_{j,i_{k+\frac{m}{n_{\text{p}}}}}}_{\bar{F}^{\text{ext}}} = 0 \qquad \forall j = [x,y,z] \quad (48)$$

4.2 BALANCE OF TOTAL ANGULAR MOMENTUM

Without a DIRICHLET boundary and external forces, the conservation of the balance of total angular momentum can be easily calculated by

$$\underbrace{\sum_{i=1}^{n_{\text{nodes}}} \left[\mathbf{q}_{i_{k+1}} \times \mathbf{p}_{i_{k+1}} - \mathbf{q}_{i_k} \times \mathbf{p}_{i_k}\right]}_{L_{i+1}-L_i} = \mathbf{0} \quad (49)$$

In the case of DIRICHLET boundary conditions, the bearing forces at the boundary nodes have to be calculated. Also the gravity force has to be included in the calculation to show the conservation of the balance of the balance of total angular momentum.

$$\underbrace{\sum_{i=1}^{n_{\text{nodes}}} \left[\mathbf{q}_{i_{k+1}} \times \mathbf{p}_{i_{k+1}} - \mathbf{q}_{i_k} \times \mathbf{p}_{i_k}\right]}_{L_{i+1}-L_i} + \underbrace{\sum_{i=1}^{n_{\text{nodes}}} \left[\sum_{m=0}^{n_{\text{p}}} M_m(\alpha) \mathbf{q}_{i,m_{k+\frac{m}{n_{\text{p}}}}} \times \mathbf{F}^{\text{ext},h}_{i_{k+\frac{m}{n_{\text{p}}}}}\right]}_{\bar{M}^{\text{ext}}} = \mathbf{0} \quad (50)$$

Then the approximated external forces $\mathbf{F}^{\text{ext},h}_{i_{k+\frac{m}{n_{\text{p}}}}}$ only depends on the gravitational force $\mathbf{F}_{\text{g}}$, and the external forces are equal to the gravity force.

$$\mathbf{F}^{\text{ext},h}_{i_{k+\frac{m}{n_{\text{p}}}}} = \mathbf{F}_{\text{g}} \quad (51)$$

In the case with boundary nodes the bearing forces at the boundary nodes have to be calculated. Therefore, we introduce a force $\mathbf{F}_\lambda$ which is theoretically approximated at the GAUSSIAN points in time and the force $\mathbf{F}_{\lambda\text{c}}$. Now we project $\mathbf{F}_{\lambda\text{c}}$ onto $\mathbf{F}_\lambda$ with the following relation.

$$\bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}_{\text{g}} \mathbf{F}_\lambda = \bar{\mathbf{M}}_{\text{c}} \mathbf{w}_{\text{c}} \mathbf{M}_{\text{c}} \mathbf{F}_{\lambda\text{c}} \quad (52)$$

$$\bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \left(\mathbf{M}^{\text{n}}_{\text{g}} \mathbf{F}^{\text{n}}_\lambda + \mathbf{M}^{1}_{\text{g}} \mathbf{F}^{1}_\lambda\right) = \bar{\mathbf{M}}_{\text{c}} \mathbf{w}_{\text{c}} \mathbf{M}_{\text{c}} \mathbf{F}_{\lambda\text{c}} \quad (53)$$

$$\bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{\text{n}}_{\text{g}} \mathbf{F}^{\text{n}}_\lambda = \bar{\mathbf{M}}_{\text{c}} \mathbf{w}_{\text{c}} \mathbf{M}_{\text{c}} \mathbf{F}_{\lambda\text{c}} - \bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{1}_{\text{g}} \mathbf{F}^{1}_\lambda \quad (54)$$

$$\mathbf{F}^{\text{n}}_\lambda = \left(\bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{\text{n}}_{\text{g}}\right)^{-1} \left[\bar{\mathbf{M}}_{\text{c}} \underbrace{\mathbf{w}_{\text{c}} \mathbf{M}_{\text{c}} \mathbf{F}_{\lambda\text{c}}}_{\tilde{\boldsymbol{\lambda}}_{\text{F}}} - \bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{1}_{\text{g}} \mathbf{F}^{1}_\lambda\right] \quad (55)$$

$$\mathbf{F}^{\text{n}}_\lambda = \left(\bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{\text{n}}_{\text{g}}\right)^{-1} \left[\bar{\mathbf{M}}_{\text{c}} \tilde{\boldsymbol{\lambda}}_{\text{F}} - \bar{\mathbf{M}}_{\text{g}} \mathbf{w}_{\text{g}} \mathbf{M}^{1}_{\text{g}} \mathbf{F}^{1}_\lambda\right] \quad (56)$$

$\bar{\mathbf{M}}_{\text{g}}$ and $\bar{\mathbf{M}}_{\text{c}}$ are trial functions for a GAUSSIAN and NEWTON-COTÊS quadrature. $\mathbf{M}_{\text{g}}$ and $\mathbf{M}_{\text{c}}$ are also shape functions for a GAUSSIAN and NEWTON-COTÊS quadrature at all quadrature

points and the corresponding quadrature weights $\mathbf{w}_\mathrm{g}$ and $\mathbf{w}_\mathrm{c}$. $\mathbf{F}_\lambda^\mathrm{n}$ represents the unknown vector of the forces in one macro time step.
Furthermore, a non-diagonal mass matrix is used and presented in the time step scheme, because it is essential to calculate the terms from nondiagonal entries in the mass matrix multiplied with the approximated velocity. We subtract the linear momentum values $\mathbf{p}_\mathrm{kin}^\mathrm{bc}$ from the calculated LAGRANGE multiplier vector $\boldsymbol{\lambda}$.

$$h_n\tilde{\boldsymbol{\lambda}}_\mathrm{F} = h_n\boldsymbol{\lambda} - \mathbf{p}_\mathrm{kin}^\mathrm{bc} \tag{57}$$

$$\tilde{\boldsymbol{\lambda}}_\mathrm{F} = \boldsymbol{\lambda} - \frac{\mathbf{p}_\mathrm{kin}^\mathrm{bc}}{h_n} \tag{58}$$

$$\mathbf{p}_\mathrm{kin}^\mathrm{nd} = \mathbb{M}\cdot\dot{\mathbf{q}} = \sum_{i=1}^{n_\mathrm{q}} M'(\alpha_i)w_i\frac{[\mathbb{M}-\mathrm{diag}(\mathbb{M})]\cdot\mathbf{q}^\mathrm{d}(\alpha_i)}{h_n} \tag{59}$$

For the calculation of the force vector $\tilde{\boldsymbol{\lambda}}_\mathrm{F}$ only the values at boundary nodes $\mathbf{p}_\mathrm{kin}^\mathrm{bc}$ are needed.

$$\mathbf{p}_\mathrm{kin}^\mathrm{bc} = \left.\mathbf{p}_\mathrm{kin}^\mathrm{nd}\right|_\mathrm{boundary} \tag{60}$$

$$\tilde{\boldsymbol{\lambda}}_\mathrm{F} = \boldsymbol{\lambda} - \frac{\mathbf{p}_\mathrm{kin}^\mathrm{bc}}{h_n} = \boldsymbol{\lambda} - \mathbf{F}_\mathrm{kin}^\mathrm{bc} \tag{61}$$

Now we use Equation (56) and calculate for every macro time step the unknown forces at the nodes $\mathbf{F}_\lambda^\mathrm{n}$ for every micro time step.
The calculated forces can now be used for the calculation of balance of the total angular momentum in Equation (50). We now define $\mathbf{F}^{\mathrm{ext},h}_{i_{k+\frac{m}{n_\mathrm{p}}}}$ as

$$\mathbf{F}^{\mathrm{ext},h}_{i_{k+\frac{m}{n_\mathrm{p}}}} = \mathbf{F}_\mathrm{g} + \sum_{m=0}^{n_\mathrm{p}} wM_m(\alpha)\mathbf{F}_{\lambda,i_{k+\frac{m}{n_\mathrm{p}}}} \tag{62}$$

Hence, it is possible to show the conservation of the balance of total angular momentum in the x, y, z-direction for a system with homolonic constraints.

4.3 THE DISCRETE GRADIENT

We start analogous to [1] with two functionals $\mathcal{F}^e$ and $\mathcal{G}^e$.

$$\mathcal{F}^e(\mathbf{D}^\mathrm{G}W_e(\alpha)) = \frac{1}{2}\int_0^1 \left\|\mathbf{D}^\mathrm{G}W_e(\alpha) - \mathbf{D}W_e(\alpha)\right\|^2 \mathrm{d}\alpha \tag{63}$$

$$\mathcal{G}^e(\mathbf{D}^\mathrm{G}W_e(\alpha)) = W_e(1) - W_e(0) - \int_0^1 \mathbf{D}^\mathrm{G}W_e(\alpha) : \frac{\partial\mathbf{C}}{\partial\alpha}\mathrm{d}\alpha \tag{64}$$

We choose the functional $\mathcal{G}^e$ as a constraint for every element with a LAGRANGE multiplier γ_e. The new functional $\mathcal{H}^e$ should only depend on $\mathbf{D}^\mathrm{G}W_e(\alpha)$ and γ_e

$$\mathcal{H}^e(\mathbf{D}^\mathrm{G}W_e(\alpha), \gamma_e) = \mathcal{F}^e(\mathbf{D}^\mathrm{G}W_e(\alpha)) + \gamma_e\mathcal{G}^e(\mathbf{D}^\mathrm{G}W_e(\alpha)) \tag{65}$$

The variation of $\mathcal{H}^e$ with respect to $\mathbf{D}^{\mathrm{G}}W_e(\alpha)$ gives us the first equation

$$\frac{\partial \mathcal{H}^e}{\partial \mathbf{D}^{\mathrm{G}}W_e(\alpha)} = \int_0^1 \left[\left(\mathbf{D}^{\mathrm{G}}W_e(\alpha) - \mathbf{D}W_e(\alpha) \right) - \gamma_e \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \right] \mathrm{d}\alpha \delta \mathbf{D}^{\mathrm{G}}W_e(\alpha) = \mathbf{0} \quad (66)$$

$$\mathbf{0} = \left(\mathbf{D}^{\mathrm{G}}W_e(\alpha) - \mathbf{D}W_e(\alpha) \right) - \gamma_e \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \quad (67)$$

$$\mathbf{D}^{\mathrm{G}}W_e(\alpha) = \mathbf{D}W_e(\alpha) + \gamma_e \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \quad (68)$$

The variation of $\mathcal{H}^e$ with respect to γ_e gives us the second equation

$$\frac{\partial \mathcal{H}^e}{\partial \gamma_e} = \left[W_e(1) - W_e(0) - \int_0^1 \mathbf{D}^{\mathrm{G}}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \right] \delta \gamma_e = \mathbf{0} \quad (69)$$

$$0 = W_e(1) - W_e(0) - \int_0^1 \mathbf{D}^{\mathrm{G}}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \quad (70)$$

Now we replace the $\mathbf{D}^{\mathrm{G}}W_e(\alpha)$ term in the Equation (70) with $\mathbf{D}^{\mathrm{G}}W_e(\alpha)$ from Equation (68).

$$\mathbf{D}^{\mathrm{G}}W_e(\alpha) = \mathbf{D}W_e(\alpha) + \gamma_e \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \quad (71)$$

$$W_e(1) - W_e(0) = \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha + \gamma_e \int_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \quad (72)$$

Rearranging the equation with respect to γ_e and replacing it in Equation (68) we obtain

$$\gamma_e = \frac{W_e(1) - W_e(0) - \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha}{\int_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha} \quad (73)$$

$$\mathbf{D}^{\mathrm{G}}W_e(\alpha) = \mathbf{D}W_e(\alpha) + \frac{W_e(1) - W_e(0) - \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha}{\int_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha} \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \quad (74)$$

Until now, the approximation of $\mathbf{C}(\alpha)$ and $\mathbf{D}W_e(\alpha)$ are not fixed. So we take a look at which kind of possiblities we have to choose an approximation. To fulfill the balance of energy in every time step, we define integral over one time step as

$$\int_0^1 \frac{\mathrm{d}W_e(\alpha)}{\mathrm{d}\alpha} \mathrm{d}\alpha = \int_0^1 \tilde{\mathbf{S}}(\alpha) : \frac{1}{2} \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha = \int_0^1 \tilde{\mathbf{S}}(\alpha) : \frac{1}{2} \frac{\partial \mathbf{C}(\alpha)}{\partial \mathbf{q}^{\mathrm{h}}} \frac{\partial \mathbf{q}^{\mathrm{h}}}{\partial \alpha} \mathrm{d}\alpha = W_e(1) - W_e(0). \quad (75)$$

To fulfill Equation (75), a special $\tilde{\mathbf{S}}(\alpha)$ is needed. We define $\tilde{\mathbf{S}}(\alpha) \equiv 2\mathbf{D}^{\mathrm{G}}W_e(\alpha)$ and test the relation.

$$\begin{aligned} W_e(1) - W_e(0) &= \int_0^1 \tilde{\mathbf{S}}(\alpha) : \frac{1}{2}\frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \\ &= \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \\ &+ \frac{W_e(1) - W_e(0) - \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha}{\int_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha} \int_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha \\ W_e(1) - W_e(0) &= W_e(1) - W_e(0) \end{aligned} \tag{76}$$

It is possible to choose a different approximation of $\mathbf{C}$ and also fulfill Equation (75).

$$\int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}_a}{\partial \alpha} \mathrm{d}\alpha + \frac{W_e(1) - W_e(0) - \int_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}_a}{\partial \alpha} \mathrm{d}\alpha}{\int_0^1 \frac{\partial \mathbf{C}_b}{\partial \alpha} : \frac{\partial \mathbf{C}_a}{\partial \alpha} \mathrm{d}\alpha} \int_0^1 \frac{\partial \mathbf{C}_b}{\partial \alpha} : \frac{\partial \mathbf{C}_a}{\partial \alpha} \mathrm{d}\alpha \tag{77}$$

One possible case is:

$$\begin{aligned} \mathbf{C}_a &\equiv \mathbf{C}(\mathbf{q}^{\mathrm{h}}) \\ \mathbf{C}_b = \mathbf{C}_a \quad &\text{and} \quad \mathbf{D}W_e(\alpha) = \mathbf{D}W_e(\mathbf{C}_a) \end{aligned} \tag{78}$$

One other case is:

$$\begin{aligned} \mathbf{C}_a &\equiv \mathbf{C}(\mathbf{q}^{\mathrm{h}}) \\ \mathbf{C}_b = \mathbf{C}^{\mathrm{h}} \neq \mathbf{C}_a \quad &\text{and} \quad \mathbf{D}W_e(\alpha) = \mathbf{D}W_e(\mathbf{C}_b) \end{aligned} \tag{79}$$

This special choice of $\mathbf{C}_a$ and $\mathbf{C}_b$ is also known as assumed strain formulation [1].

4.4 ENERGY CONSERVING

A variational integrator only bind the energy error in every time step, but not under the NEWTON tolerance. We implement a discrete gradient to correct energy error [8], [1]. The variational integrator loses its symplecticity, but the absolute value of the energy error in every time step is bounded by the NEWTON tolerance. Splitting Equation (74) in two parts allows us to put one part in potential energy and the other part can be introduced by the LAGRANGE-

D'ALEMBERT principle.

$$\begin{aligned}\int\limits_{\mathcal{B}_0^e} \frac{\partial \rho_0 \psi^e\left(\mathbf{C}^{\mathrm{h}^e}\right)}{\partial \bar{\mathbf{q}}_{k+1}^e} \mathrm{d}V &= \int\limits_{\mathcal{B}_0^e} \mathbf{S}^e\left(\mathbf{C}^{\mathrm{h}^e}\right) : \mathbb{B}^e\left(\mathbf{q}^{\mathrm{h}^e}(\alpha)\right) \mathrm{d}V \\ &= \int\limits_{\mathcal{B}_0^e} 2\mathbf{D}W_e(\mathbf{C}^{\mathrm{h}^e}) : \mathbb{B}^e\left(\mathbf{q}^{\mathrm{h}^e}(\alpha)\right) \mathrm{d}V\end{aligned} \tag{80}$$

The second part will be placed in the LAGRANGE-D'ALEMBERT principle.

$$\mathcal{K}^{DG}\left(\mathbf{C}(\alpha)\right) = 2 \frac{W_e(1) - W_e(0) - \int\limits_0^1 \mathbf{D}W_e(\alpha) : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha}{\int\limits_0^1 \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} : \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \mathrm{d}\alpha} \frac{\partial \mathbf{C}(\alpha)}{\partial \alpha} \tag{81}$$

The virtual work can be written as

$$\begin{aligned}\delta W_{\mathrm{d},k+1}(\bar{\mathbf{q}}_{k+1}^e) &= -h_n \sum_{i=1}^{n_\mathrm{q}} w_i \mathcal{K}^{DG}(\mathbf{C}^{\mathrm{h}^e}) : \mathbb{B}^e\left(\mathbf{q}^{\mathrm{h}^e}(\alpha_i)\right) \delta \mathbf{q}^{\mathrm{h}^e}(\alpha_i) \\ &= -\sum_{i=0}^{n_\mathrm{q}} \mathbf{\Gamma}_{k+\frac{i}{n_\mathrm{p}}}^{DG}\left(\mathbf{C}^{\mathrm{h}^e}, \mathbf{q}^{\mathrm{h}^e}\right) \delta \mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}^e\end{aligned} \tag{82}$$

The new time step scheme leads to

$$-\mathbf{p}_k = \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_k} - h_n \left(\frac{\partial \mathbf{\Phi}(\mathbf{q}_k)}{\partial \mathbf{q}_k}\right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_k + \mathbf{\Gamma}_k^{DG}\left(\mathbf{C}^{\mathrm{h}^e}, \mathbf{q}^{\mathrm{h}^e}\right) \tag{83}$$

$$\begin{aligned}\mathbf{0} &= \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}} - h_n \left(\frac{\partial \mathbf{\Phi}\left(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}\right)}{\partial \mathbf{q}_{k+\frac{i}{n_\mathrm{p}}}}\right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_{k+\frac{i}{n_\mathrm{p}}} + \mathbf{\Gamma}_{k+\frac{i}{n_\mathrm{p}}}^{DG}\left(\mathbf{C}^{\mathrm{h}^e}, \mathbf{q}^{\mathrm{h}^e}\right)\Bigg|_{i=1}^{n_\mathrm{p}-1} \\ \mathbf{0} &= \mathbf{\Phi}(\mathbf{q}_{k+\frac{i}{n_\mathrm{p}}})\Big|_{i=1}^{n_\mathrm{p}}\end{aligned}$$

$$\begin{aligned}\mathbf{p}_{k+1} &= \frac{\partial \mathcal{L}_{\mathrm{d},k+1}}{\partial \mathbf{q}_{k+1}} - h_n \left(\frac{\partial \mathbf{\Phi}(\mathbf{q}_{k+1})}{\partial \mathbf{q}_{k+1}}\right)^{\mathrm{T}} \cdot \boldsymbol{\lambda}_{k+1} + \mathbf{\Gamma}_{k+1}^{DG}\left(\mathbf{C}^{\mathrm{h}^e}, \mathbf{q}^{\mathrm{h}^e}\right) \\ \mathbf{0} &= \mathbf{\Psi}\left(\mathbf{q}_{k+1}, \mathbf{p}_{k+1}\right)\end{aligned} \tag{84}$$

With respect to Equation (77), we define the time step scheme with the following approximation of the right CAUCHY-GREEN tensor as VDG1:

$$\mathbf{C}^{\mathrm{h}^e} \equiv \mathbf{C}^e\left(\mathbf{q}^{\mathrm{h}^e}\right) \quad \text{and} \quad \mathbf{D}W_e\left(\mathbf{C}^{\mathrm{h}^e}\right) = \mathbf{D}W_e\left(\mathbf{C}^e\left(\mathbf{q}^{\mathrm{h}^e}\right)\right) \tag{85}$$

This approximation of the right CAUCHY-GREEN tensor in the time step scheme will be named as VDG2:

$$\mathbf{C}^{\mathrm{h}^e} \equiv \sum_{i=0}^{n_\mathrm{q}} M_i(\alpha) \mathbf{C}_i^e(\mathbf{q}_i) \quad \text{and} \quad \mathbf{D}W_e\left(\mathbf{C}^{\mathrm{h}^e}\right) = \mathbf{D}W_e\left(\mathbf{C}_i^e\right) \tag{86}$$

The difference with respect to the energy error and the stability will be shown for the numerical examples.

5 NUMERICAL EXAMPLES

First we show the convergence for a higher order accuracy in position and linear momentum for a free flying rotor. In the next numerical example we also proof the balance of linear momentum, the balance of angular momentum and the balance of energy for the free flying rotor and the vibration of a cantilever beam.

5.1 CONVERGENCE STUDY FOR THE FREE FLYING HYPERELASTIC ROTOR

The convergence was calculated for a simulation time $t = 1\,\mathrm{s}$. After this time, the values of position and linear momentum vectors were used as $\mathbf{q}_{\mathrm{end}}$ and $\mathbf{p}_{\mathrm{end}}$. The reference solution for $\mathbf{q}_{\mathrm{ref}}$ and $\mathbf{p}_{\mathrm{ref}}$ for every polynomial degree n_{p} was calculated with the smallest time step size $h_n \approx 5 \cdot 10^{-4}$. A convergence study is preformed with the following simulation parameters:

n_{el}	560	n_{dof}	$3156 \cdot n_{\mathrm{p}}$
h_n	$2^{-1,\ldots,-9}\,\mathrm{s}$	T_{end}	$1\,\mathrm{s}$
n_{p}	$1, \ldots, 4$	ρ_0	8.93
$\rho_0 \lambda$	66.25	$\rho_0 \mu$	650
ϵ	$1 \cdot 10^{-8}$	$\mathbf{r}_0$	$[0, 0, 0.3]^{\mathrm{T}}$
$\boldsymbol{\omega}$	$[1, 0, 1]^{\mathrm{T}}$	$\mathbf{v}$	$[10, 0, 15]^{\mathrm{T}}$

Table 1: Simulation parameters for the convergence study

The free flying rotor in the reference configuartion is printed in the Figure 2. The motion is subject to a gravitational force field $\mathbf{g}$, and initiated by an initial translational velocity field $\mathbf{v}$ and an intial angular velocity field $\boldsymbol{\omega}$ with the corresponding position vector $\mathbf{r}_0$.

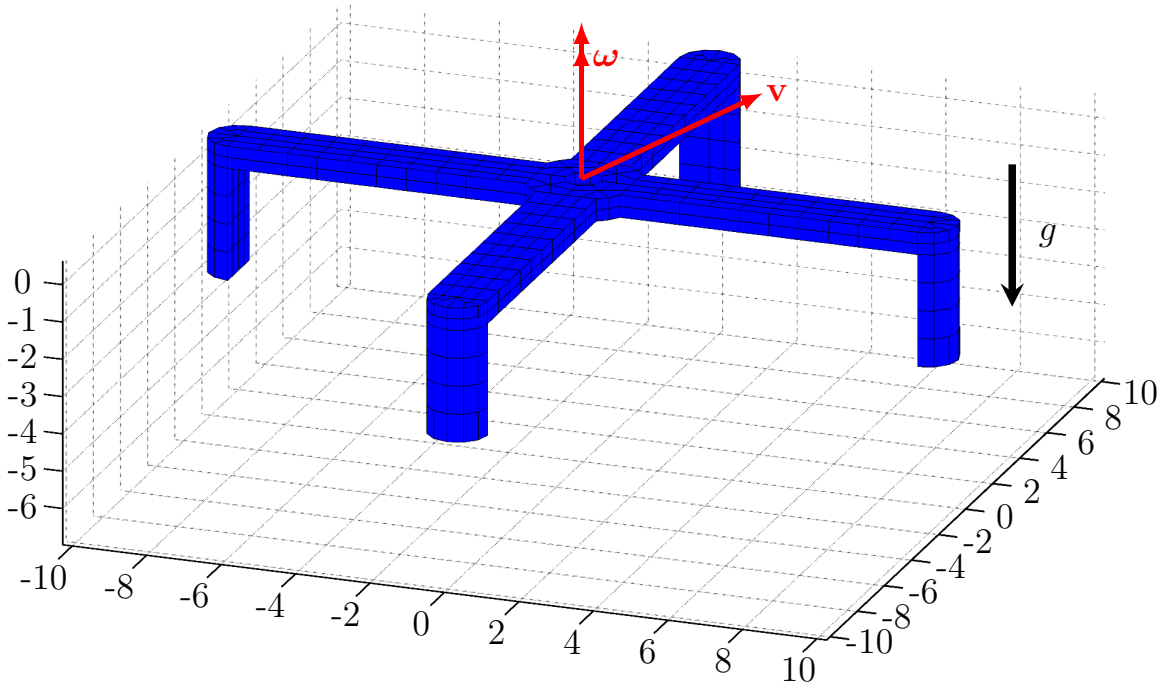


Figure 2: Reference configuration for the free flying hyperelastic rotor for the time $t = 0\,\mathrm{s}$

As error criterion, the relative error in position and linear momentum is defined using the EUCLIDIAN norm. Theoretically, convergence order for position vector and linear momentum vector should be identical; in fact, they are almost identical in dependence of the stiffness of the material. The theoretical values were reached until $n_\mathrm{p} = 3$. For $n_\mathrm{p} = 4$, the theoretical values are not obtained, but we assume that with more calculation points it would be possible to show the theoretical convergence order.

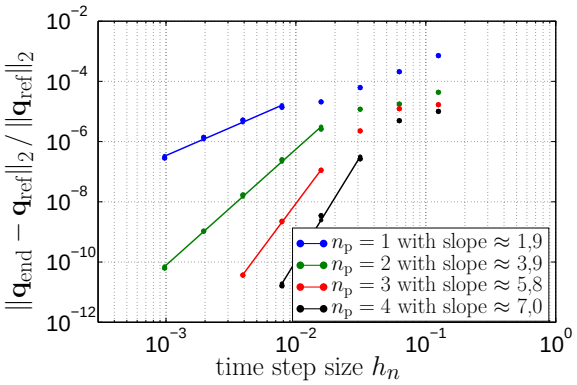


Figure 3: Convergence in position without discrete gradient

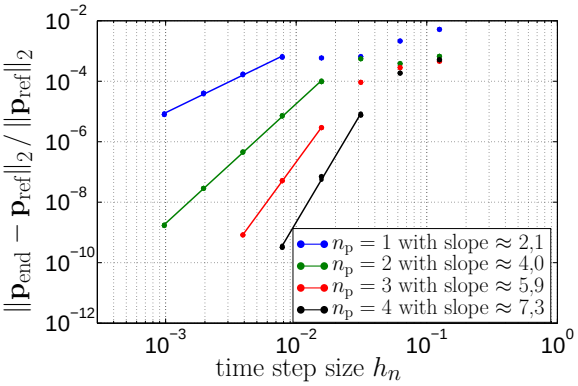


Figure 4: Convergence in linear momentum without discrete gradient

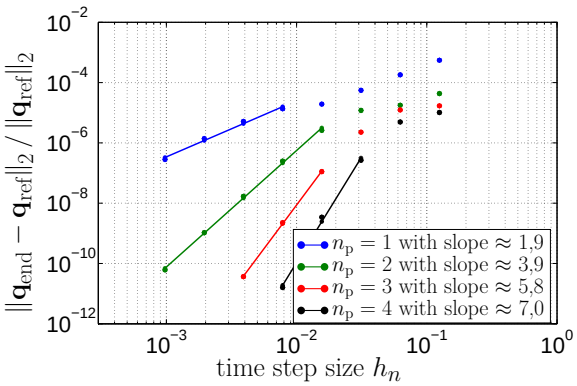


Figure 5: Convergence in position with VDG2

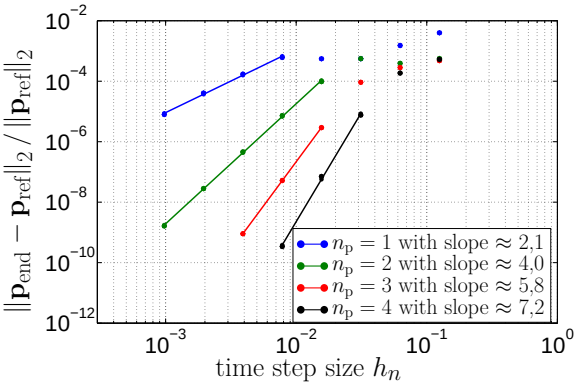


Figure 6: Convergence in linear momentum with VDG2

5.2 FREE FLYING HYPERELASTIC ROTOR

In the Figure 7 the reference configuration of the free flying rotor is shown with a finer mesh. In this example the rotor gets an initial linear velocity $\mathbf{v}$ and an initial angular velocity $\boldsymbol{\omega}$ around the point $\mathbf{r}_0$. The simulation parameters can be found in the Table 2. The printed total CPU-time

n_{el}	18639	n_{dof}	$74136 \cdot n_{\text{p}} = 222408$
h_n	0.02 s	T_{end}	4 s
n_{p}	3	ρ_0	8.93
$\rho_0\lambda$	$66.25 \cdot 10^3$	$\rho_0\mu$	$65 \cdot 10^5$
ϵ	$1 \cdot 10^{-7}$	$\mathbf{r}_0$	$[0, 0, 0.3]^{\text{T}}$
$\boldsymbol{\omega}$	$[1, 0, 1]^{\text{T}}$	$\mathbf{v}$	$[10, 0, 15]^{\text{T}}$
VI without discrete gradient			
t_{cpu}	14790 s	t_{solver}	11660 s
VDG2			
t_{cpu}	17660 s	t_{solver}	12380 s

Table 2: Simulation parameters for the free flying rotor

and SOLVER-time in the table both are measured for a parallelization of the element calculation in CUDA. For a large number of degrees of freedom n_{dof} resulting from the polynomial degree n_{p} for the selected approximation, the CPU-time t_{cpu} more and more depends on the SOLVER-time t_{solver}.
The overhead for a calculation with a higher order approximation and the effort of using a discrete gradient in the element routine can be compensated by a parallelization of the element routine with CUDA.

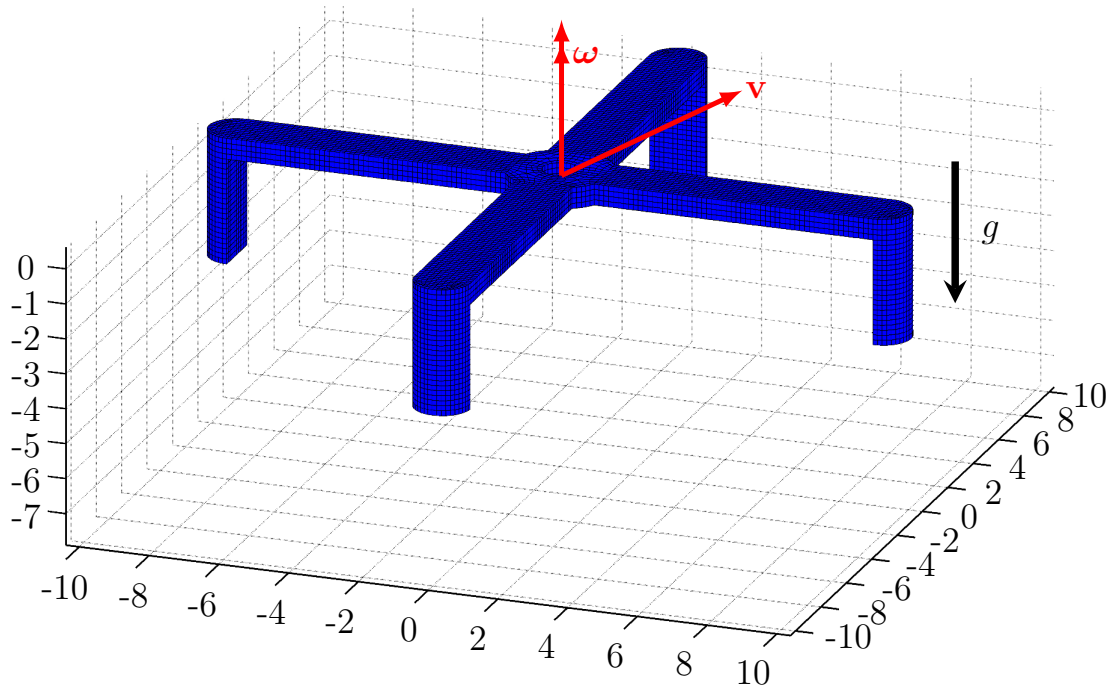


Figure 7: Reference configuration for the free flying hyperelastic rotor for the time $t = 0$ s

The Figures 8-11 show the parabolic free fly of the rotor in the gravitational field for different time points. The color bar at the right-hand side of each figure marks the VON MISES stress at the spatial nodes.

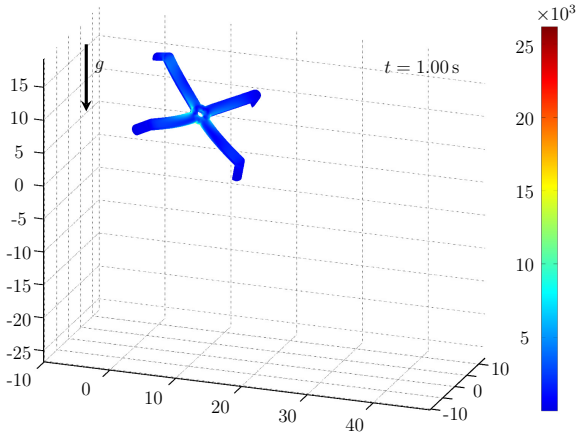


Figure 8: Actual configuration for the free flying hyperelastic rotor for the time $t = 1\,\mathrm{s}$

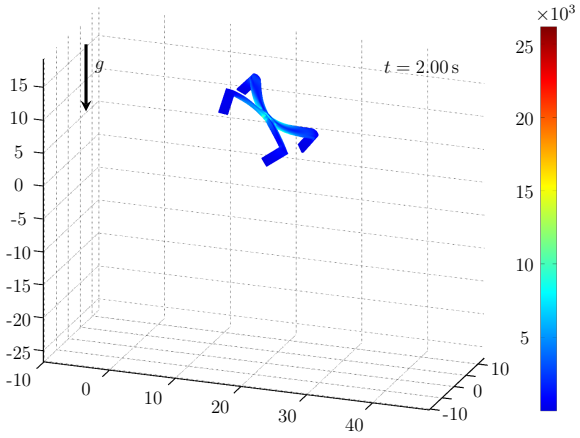


Figure 9: Actual configuration for the free flying hyperelastic rotor for the time $t = 2\,\mathrm{s}$

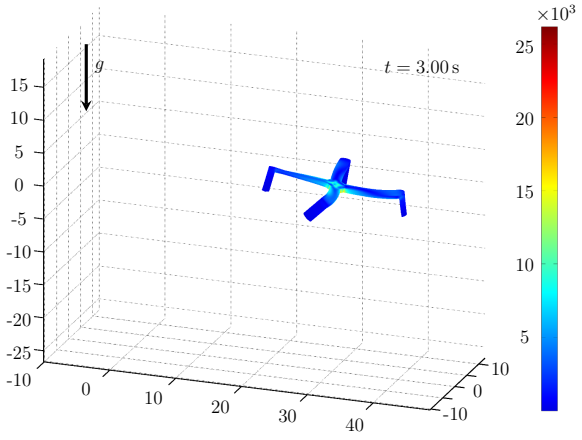


Figure 10: Actual configuration for the free flying hyperelastic rotor for the time $t = 3\,\mathrm{s}$

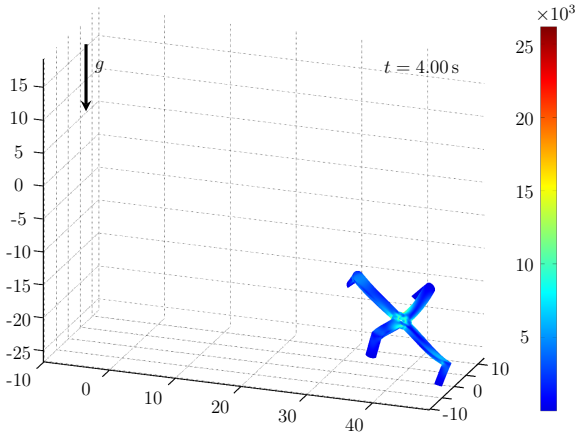


Figure 11: Actual configuration for the free flying hyperelastic rotor for the time $t = 4\,\mathrm{s}$

For this simulation, we calculated the balance of total linear momentum, the balance of total angular momentum, and the balance of total energy for a higher order variational integrator without using a discrete gradient. All balances were also divided by the NEWTON criterion ϵ, see Table 2. The balances of total linear momentum and balance of total angular momentum

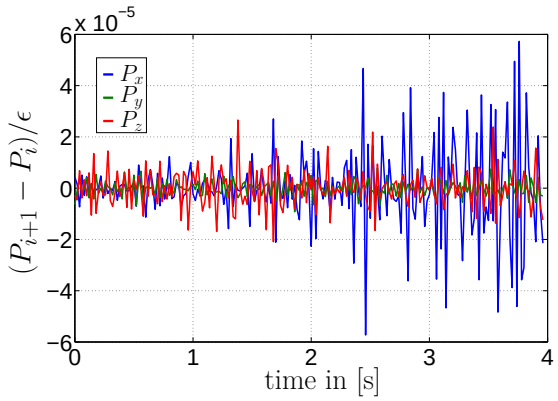


Figure 12: Balance of total linear momentum without discrete gradient for free flying rotor with $n_\mathrm{p} = 2$

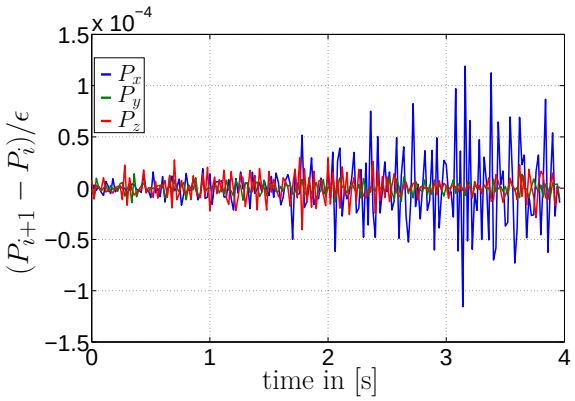


Figure 13: Balance of total linear momentum without discrete gradient for free flying rotor with $n_\mathrm{p} = 3$

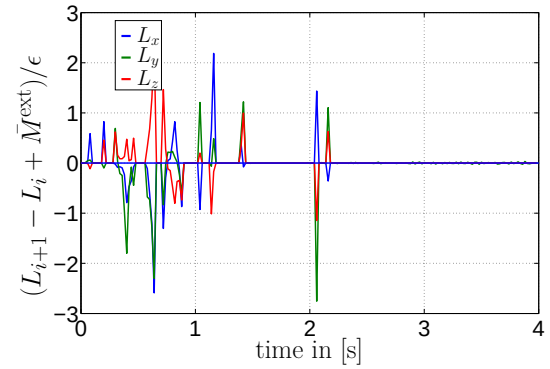


Figure 14: Balance of total angular momentum without discrete gradient for free flying rotor with $n_\mathrm{p} = 2$

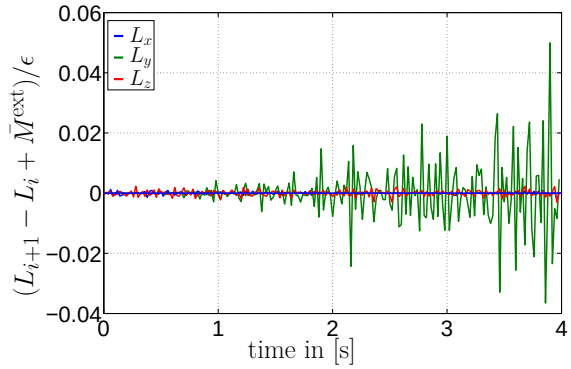


Figure 15: Balance of total angular momentum without discrete gradient for free flying rotor with $n_\mathrm{p} = 3$

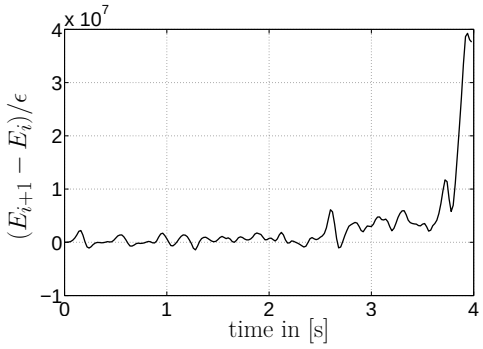


Figure 16: Balance of total energy without discrete gradient for free flying rotor with $n_\mathrm{p} = 2$

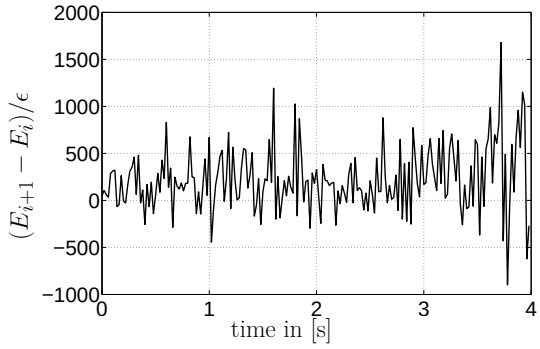


Figure 17: Balance of total energy without discrete gradient for free flying rotor with $n_\mathrm{p} = 3$

are met, but not the balance of total energy. The balance of the total angular momentum is calculated under influence of the gravitational force, see Equation (50).
The same kind of calculations are preformed with the discrete gradient for the time step scheme VDG2. The balance of total linear momentum and balance of total angular momentum are also met.

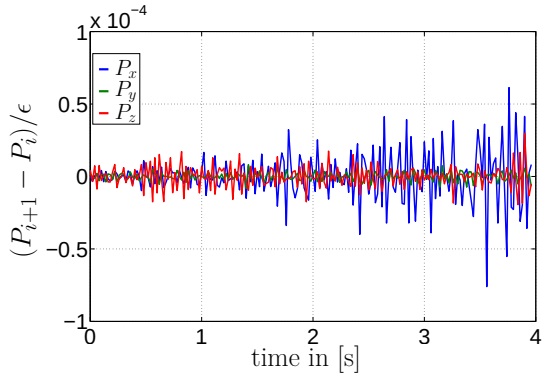


Figure 18: Balance of total linear momentum with VDG2 for free flying rotor with $n_\mathrm{p} = 2$

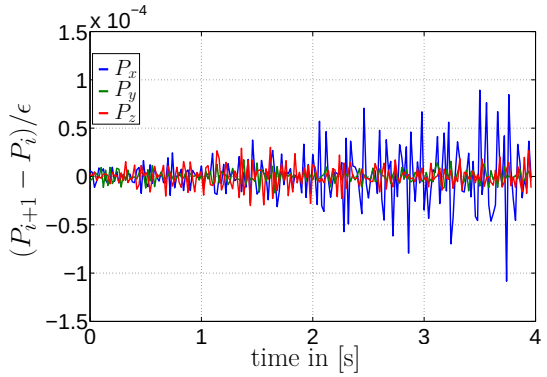


Figure 19: Balance of total linear momentum with VDG2 for free flying rotor with $n_\mathrm{p} = 3$

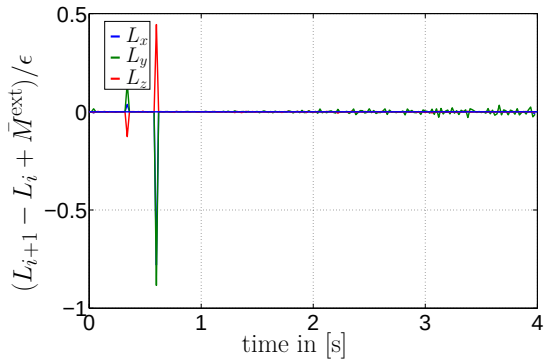


Figure 20: Balance of total angular momentum with VDG2 for free flying rotor with $n_\mathrm{p} = 2$

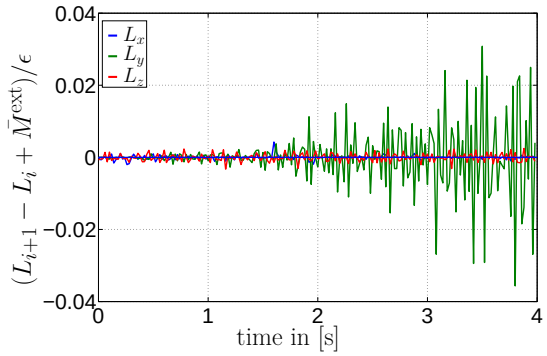


Figure 21: Balance of total angular momentum with VDG2 for free flying rotor with $n_\mathrm{p} = 3$

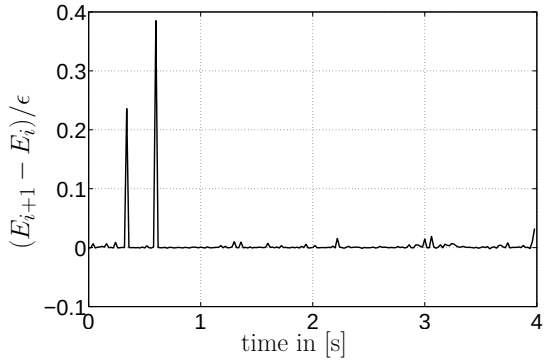


Figure 22: Balance of total energy with VDG2 for free flying rotor with $n_\mathrm{p} = 2$

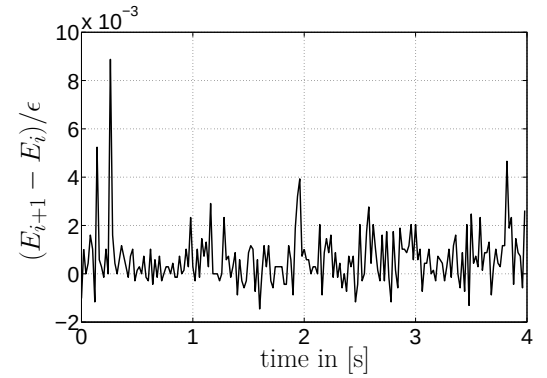


Figure 23: Balance of total energy with VDG2 for free flying rotor with $n_\mathrm{p} = 3$

Importanted are the Figures 22-23 showing the balances of the total energy for VDG2 time step scheme. The absolute value of the error in the balance of total energy for every time step is below the chosen NEWTON criterion ϵ. Hence, the VDG2 time step scheme preserves the total energy over the total simulation time. Another interesting point is the balance of total angular momentum for VDG2 with $n_\mathrm{p} = 2$ Figure 20, here a no peaks over the NEWTON criterion arise. This simulation was calculated with the same parameters like the simulation in Figure 14, but the discrete gradient shows a better fulfillment of the balance of total angular momentum.

5.3 VIBRATION OF HYPERELASTIC CANTILEVER BEAM

The second numerical example is a vibration of a hyperelastic cantilever beam. We used the rotor once more, but now we fixed the degrees of freedom $n_{\text{q,fix}}$ in all directions in the hub of the rotor, see Figure 24. Under influence of the gravitational force, every blade of the rotor starts to swing around the static equilibrium position.

n_{el}	18639	n_{dof}	$74136 \cdot n_{\text{p}} = 222408$
h_n	$0.025\,\text{s}$	T_{end}	$4\,\text{s}$
n_{p}	3	ρ_0	8.93
$\rho_0\lambda$	$16.25 \cdot 10^4$	$\rho_0\mu$	$15 \cdot 10^6$
ϵ	$1 \cdot 10^{-7}$	$n_{\text{q,fix}}$	$360 \cdot n_{\text{p}} = 1080$
VI without discrete gradient			
t_{cpu}	$14560\,\text{s}$	t_{solver}	$11620\,\text{s}$
VDG2			
t_{cpu}	$17850\,\text{s}$	t_{solver}	$12730\,\text{s}$

Table 3: Simulation parameters for the vibration of the hyperelastic cantilever beam

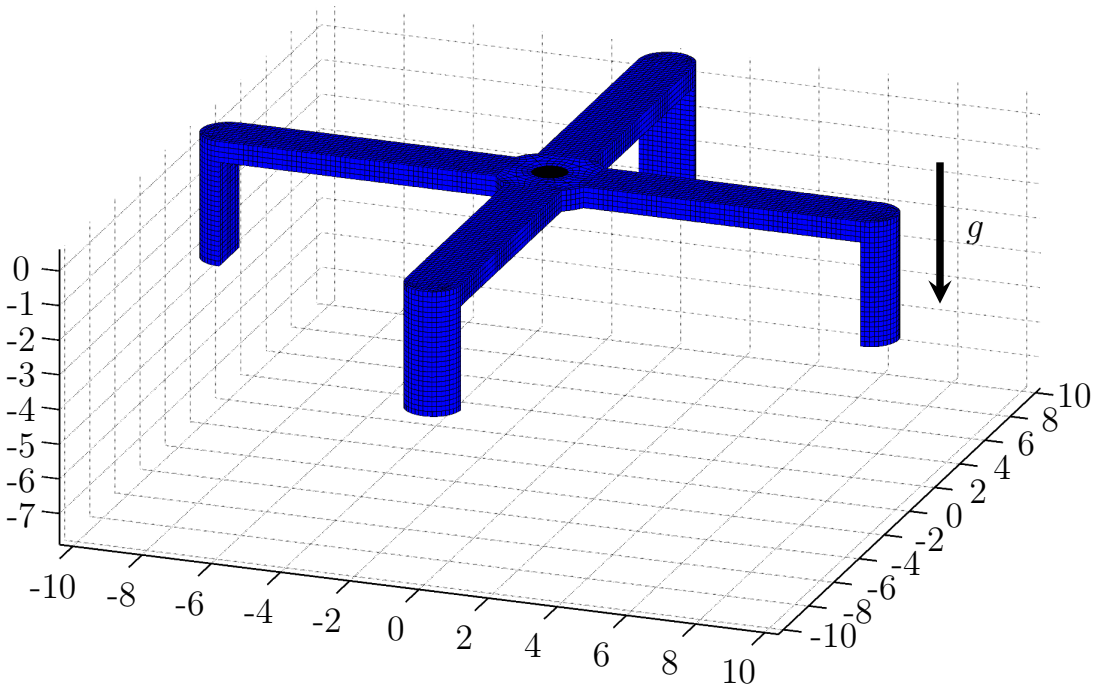


Figure 24: Reference configuration for the hyperelastic cantilever beam for the time $t = 0\,\text{s}$

The Figures 25-30 show the vibration around the static idle position for every blade of the rotor. With the fixed hub, every blade shows the same vibration behavior as the cantilever beam in four directions due to the symmetry. At the time $t = 1.925\,\mathrm{s}$, the blades reach the maximum elongation. After $t = 3.775\,\mathrm{s}$, the vibration of the rotor is again in the initial state.

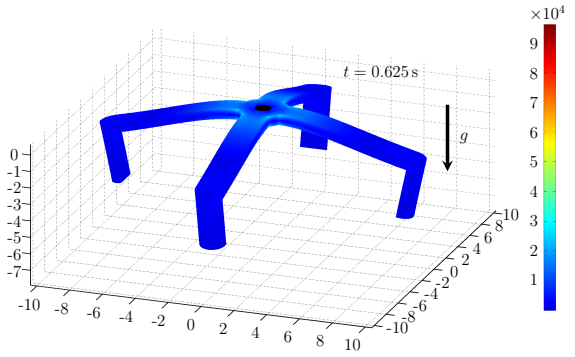


Figure 25: Actual configuration for the hyperelastic cantilever beam for the time $t = 0.625\,\mathrm{s}$

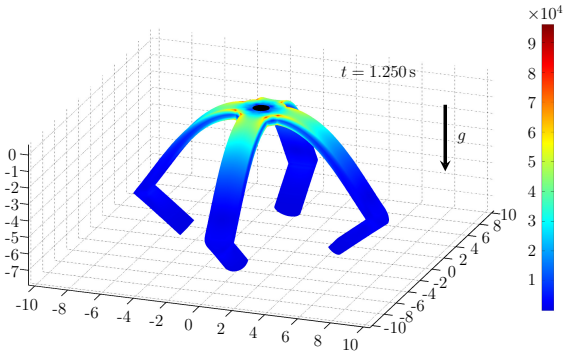


Figure 26: Actual configuration for the hyperelastic cantilever beam for the time $t = 1.25\,\mathrm{s}$

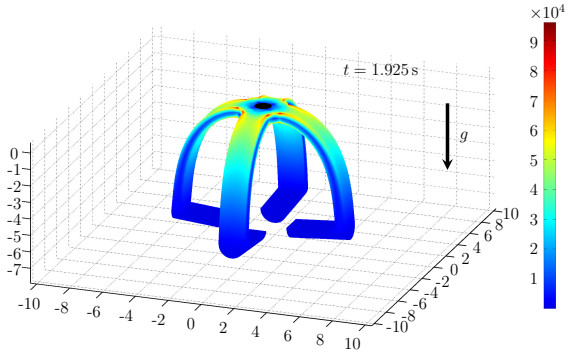


Figure 27: Actual configuration for the hyperelastic cantilever beam for the time $t = 1.925\,\mathrm{s}$

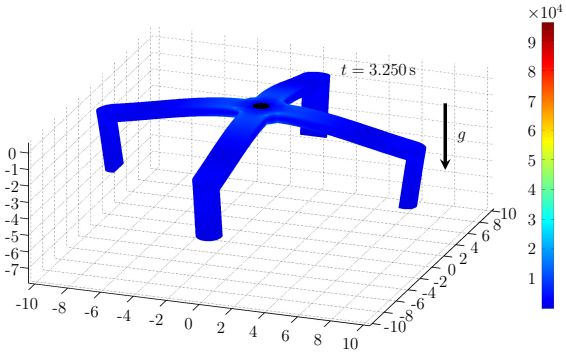


Figure 28: Actual configuration for the hyperelastic cantilever beam for the time $t = 3.25\,\mathrm{s}$

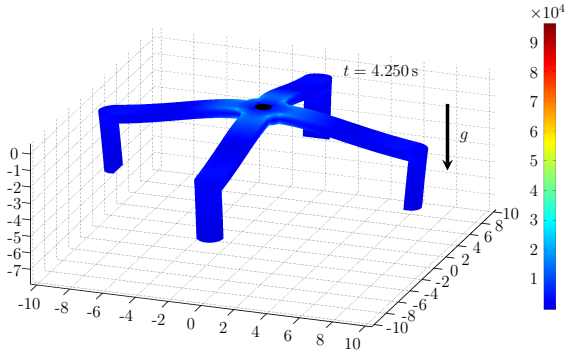


Figure 29: Actual configuration for the hyperelastic cantilever beam for the time $t = 4.25\,\mathrm{s}$

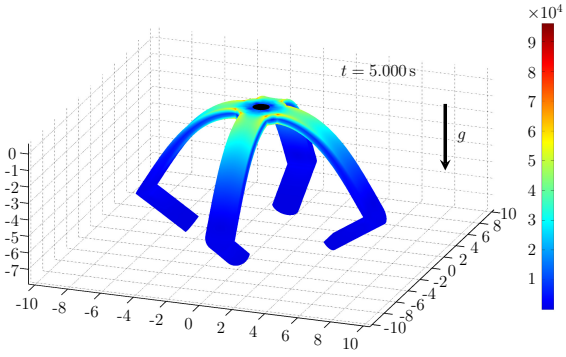


Figure 30: Actual configuration for the hyperelastic cantilever beam for the time $t = 5\,\mathrm{s}$

The balance of total linear momentum is for the variational integrator without a discrete gradient and with the VDG2 preserved. The same behavior is calculated for the balance of total angular momentum. For the calculation of the balance of total linear momentum we use Equation (48) and for the balance of total angular momentum we use once more Equation (50). However, before this we have to calculate the bearing forces with respect to Equation (62). The

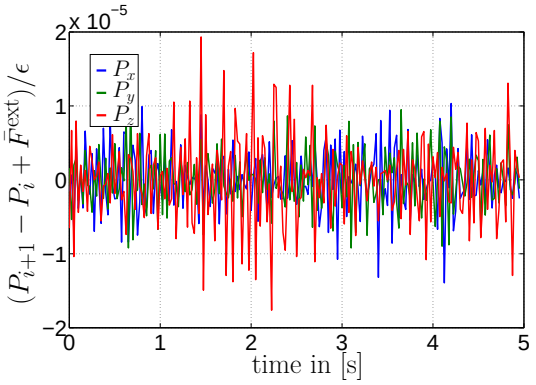


Figure 31: Balance of total linear momentum with VDG2 for hyperelastic cantilever beam with $n_\text{p} = 3$

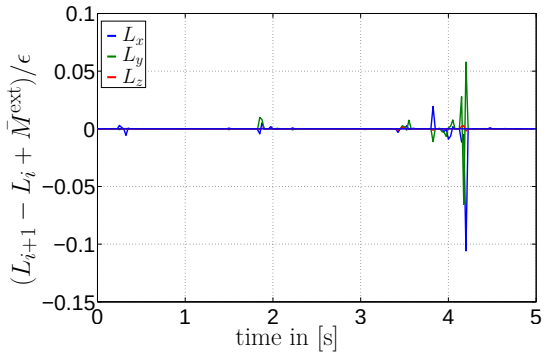


Figure 32: Balance of total angular momentum with VDG2 for hyperelastic cantilever beam with $n_\text{p} = 3$

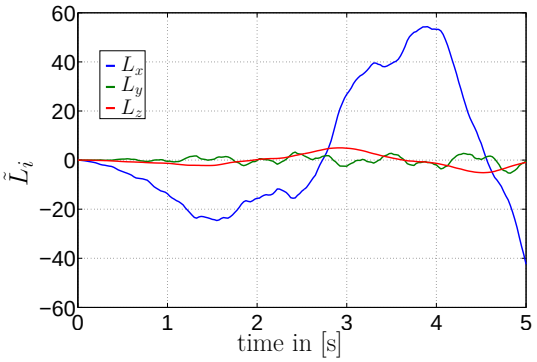


Figure 33: Angular momentum with VDG2 for hyperelastic cantilever beam with $n_\text{p} = 3$

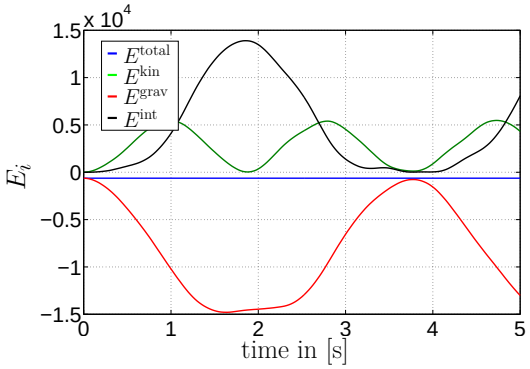


Figure 34: energy with VDG2 for hyperelastic cantilever beam with $n_\text{p} = 3$

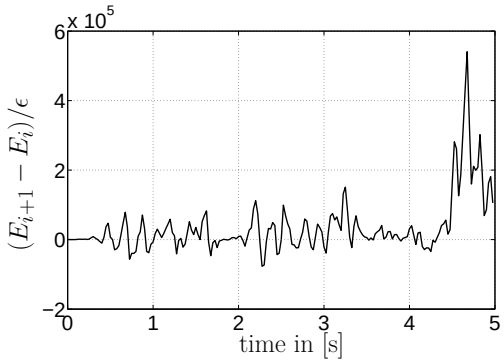


Figure 35: Balance of total energy without discrete gradient for hyperelastic cantilever beam with $n_\text{p} = 3$

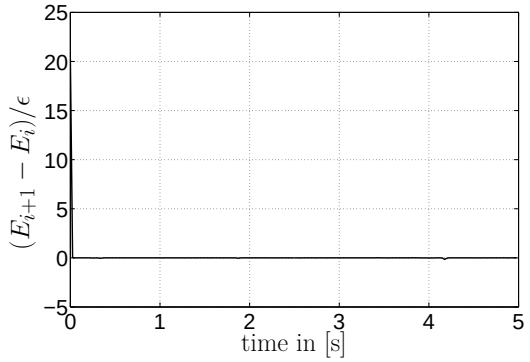


Figure 36: Balance of total energy with VDG2 for hyperelastic cantilever beam with $n_\text{p} = 3$

total angular momentum with the momentum term $\tilde{L}$ is printed in Figure 33. The values of the different energies over the simulation time are shown in Figure 34. Again, the variational integrator without a discrete gradient cannot preserve the balance of total energy, but the VDG2 is able to do it. In Figure 36 the complete balance is met for every time step, but not in the first

time step. Here the simulation starts in the first time step with a very small change in values for the strains. Therefore, for the time step scheme VDG2 it is difficult to handle elements which are nearly undeformed. This means that the NEWTON criterion will not be reached. This brings a little error in the balance of total energy for the first time step. For all the other time steps the balance of total energy is preserved.

5.4 VIBRATION OF HYPERELASTIC CANTILEVER BEAM WITH VDG1

The last numerical example is also the vibration of a cantilever beam. It is almost the same simulation as before, but now we want to use VDG1 time step scheme. We change the number of elements n_{el} to a smaller number and set a smaller time step size $h_n = 0.01\,\text{s}$, see Table 4.

n_{el}	560	n_{dof}	$3156 \cdot n_{\text{p}} = 9468$
h_n	$0.01\,\text{s}$	T_{end}	$5\,\text{s}$
n_{p}	3	ρ_0	8.93
$\rho_0\lambda$	$16.25 \cdot 10^4$	$\rho_0\mu$	$15 \cdot 10^6$
ϵ	$1 \cdot 10^{-7}$	$n_{\text{q,fix}}$	$72 \cdot n_{\text{p}} = 216$

Table 4: Simulation parameters for the vibration of the hyperelastic cantilever beam with VDG1

The reference configuartion is shown in the Figure 37. The single blades of the rotor also swing around their static equilibrium positions, but the vibration do not reach large values because the smaller number of elements makes the material more stiff.

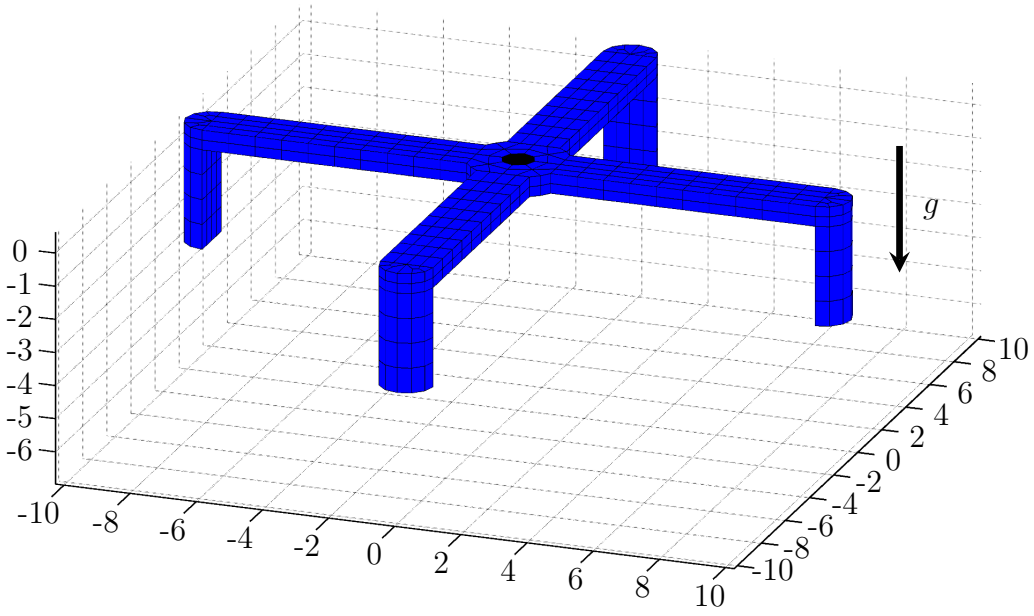


Figure 37: Reference configuration for the hyperelastic cantilever beam with VDG1 for the time $t = 0\,\text{s}$

This behavior can also be seen in the Figures 38-41 that show the vibration of the rotor for different time steps. The smaller values in the VON MISES stress can also be seen in the color bar at the right-hand side of every figure.

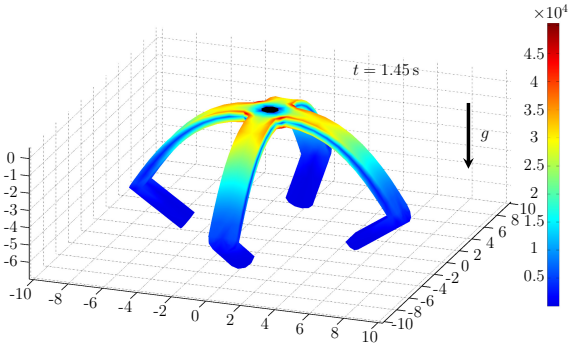


Figure 38: Actual configuration for the hyperelastic cantilever beam with VDG1 for the time $t = 1.45\,\mathrm{s}$

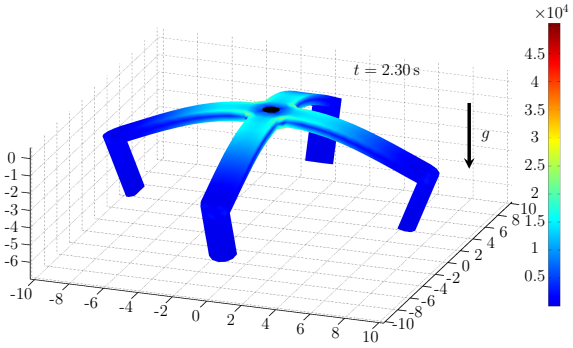


Figure 39: Actual configuration for the hyperelastic cantilever beam with VDG1 for the time $t = 2.3\,\mathrm{s}$

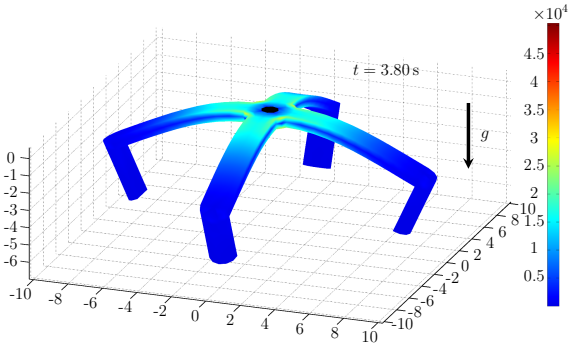


Figure 40: Actual configuration for the hyperelastic cantilever beam with VDG1 for the time $t = 3.8\,\mathrm{s}$

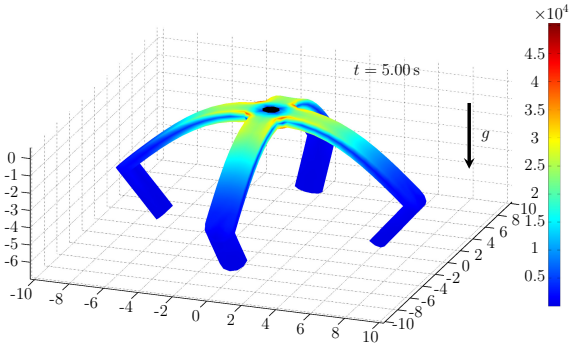


Figure 41: Actual configuration for the hyperelastic cantilever beam with VDG1 for the time $t = 5.0\,\mathrm{s}$

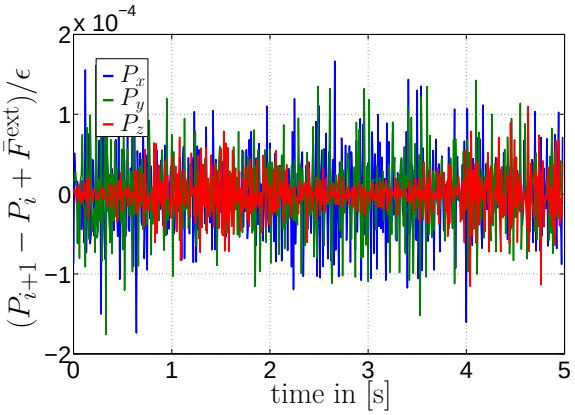


Figure 42: Balance of total linear momentum with VDG1 for hyperelastic cantilever beam with $n_\mathrm{p} = 3$

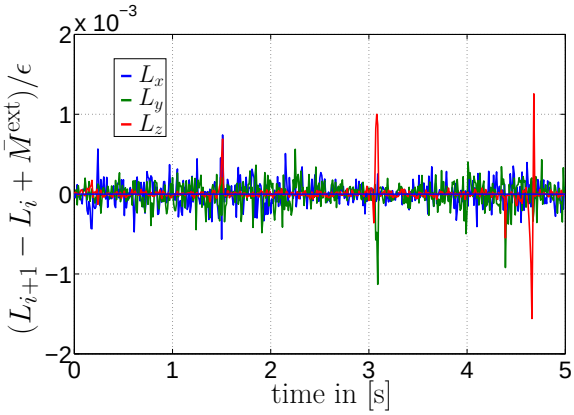


Figure 43: Balance of total angular momentum with VDG1 for hyperelastic cantilever beam with $n_\mathrm{p} = 3$

For these simulation parameters, the balance of total linear momentum and the balance of total angular momentum is met for every time step.
The Figures 44-45 show the different energies and the balance of the total total energy. Almost all values in the balance of total energy are preserved for all time steps. There is only a small number of peaks in which the balance is not met. This behavior can only be shown in the case of a very good choice of simulation parameters because the VDG1 method is numerically very instable, compare [4]. The simulation starts with normal results but, sometimes the NEWTON method cannot find a solution that fulfills the NEWTON criterion for the residual. In this case, the simulation finishes the time step with incorrect values. If this happened too often within the simulation, the entire simulation would diverge.

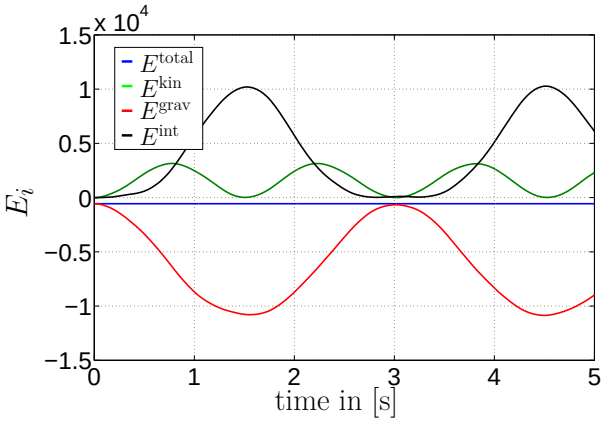


Figure 44: Energy with VDG1 for hyperelastic cantilever beam with $n_{\text{p}} = 3$

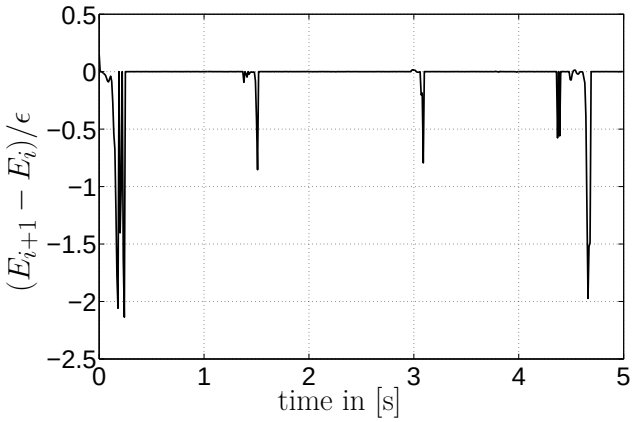


Figure 45: Balance of total energy with VDG1 for hyperelastic cantilever beam with $n_{\text{p}} = 3$

6 CONCLUSION

We have shown how a variational integrator of higher order accuracy can be consistently formulated for a hyperelastic continuum, and that this integrator meets the balance of total linear momentum and total angular momentum. The order of accuracy in position and momentum reaches the theoretical values if using the GAUSSIAN quadrature rule for the approximation of the position vector.
The inclusion of a LAGRANGE multiplier vector is also possible in the variational principle. The application of the NEWTON-COTÊS quadrature allows to decouple the equations in the position-momentum form. We have also shown how the bearing forces can be calculated from the LAGRANGE multiplier vector for the fixed nodes.
The split of the discrete gradient into two terms one in the LAGRANGIAN and one in a D'ALEMBERT term makes it also possible to create an integrator which meets the balance of total energy for every time step in the simulation. It is true that the integrator loses the symplecticity, but achieves the importanted total energy conservation for continuum problems.
However, it should be noted that an approximation of the right CAUCHY-GREEN tensor $\mathbf{C}$ with the position vector $\mathbf{q}$ is numerically instable, see VDG1. But, with an assumed strain approximation in the discrete gradient, the instabilities can be solved.
Further, the ongoing development of the computer hardware opens new possiblities for numerical simulations, like the parallelization of element routines on graphic cards with CUDA. That can save a lot of CPU time and makes time step schemes with better physical properties more attractive for further application.

ACKNOWLEDGEMENTS

The authors thank the 'Deutsche Forschungsgesellschaft (DFG)' for financial support of this work under the grant GR 3297/2-2 and Oliver Klöckner for his contribution about the parallelization of element routines with CUDA.

REFERENCES

[1] M. Groß, P. Betsch, P. Steinmann, Conservation properties of a time FE method. Part IV: Higher order energy and momentum conserving schemes. *Int. J. Numer. Meth. Engng* **63**, 1849–1897, 2005.

[2] P. Betsch, P. Steinmann, Conservation properties of a time FE method–part III: Mechanical systems with holonomic constraints. *Int. J. Numer. Meth. Engng.*, **53**, 2271–2304, 2002.

[3] P. Wriggers, *Nichtlineare Finite-Element-Methoden*, 1th Edition, Springer, Berlin, 2008.

[4] R. Mohr, *Consistent Time-Integration of Finite Elasto-Plasto-Dynamics*. Dissertation, 2008.

[5] M. Krüger, M. Groß, P. Betsch, An energy-entropy-consistent time stepping scheme for nonlinear thermo-viscoelastic continua. *ZAMM*, **96**, 141–178, 2016.

[6] S. Ober-blöbaum, N. Saake, Construction and analysis of higher order Galerkin variational integrators. *Advances in Computational Mathematics*, 1–32, 2014.

[7] J. E. Marsden and M. West, Discrete mechanics and variational integrators. *Acta Numerica*, **10**, 357–514, 2001.

[8] O. Gonzalez, Exact energy and momentum conserving algorithms for general models in nonlinear elasticity. *Computer Methods in Applied Mechanics and Engineering*, **190**, 1763–1783, 2000.

[9] M. Bartelt, M. Groß, Variational integrators of higher order for flexible multibody systems. *Proc. Appl. Math. Mech.*, **15**, 43–44, 2015.

[10] S. Hartmann, P. Ne, Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility. *International Journal of Solids and Structures*, **40**, 2767–2791, 2003.

[11] Nijmeyer H. and van der Schaft A.J., *Nonlinear Dynamical Control Systems*, Springer, New York, 1990.

A SHAPE FUNCTIONS IN TIME

The following matricies are time shape functions for the GAUSSIAN quadrature $\mathbf{M}_\mathrm{g}$ and for the NEWTON-COTÊS quadrature $\mathbf{M}_\mathrm{c}$. $\bar{\mathbf{M}}_\mathrm{g}$ and $\bar{\mathbf{M}}_\mathrm{c}$ are the shape functions one polynomial degree lower but with same quadrature points. $\mathbf{w}_\mathrm{g}$ and $\mathbf{w}_\mathrm{c}$ are the weights of the quadrature rule and $\boldsymbol{\alpha}_\mathrm{g}$ and $\boldsymbol{\alpha}_\mathrm{c}$ are the quadrature points.

$$\mathbf{M}_\mathrm{c} = \begin{bmatrix} M_{\mathrm{c}1,1} & M_{\mathrm{c}1,2} & \dots M_{\mathrm{c}1,n_\mathrm{c}+1} \\ M_{\mathrm{c}2,1} & M_{\mathrm{c}2,2} & \dots M_{\mathrm{c}2,n_\mathrm{c}+1} \\ \vdots & \vdots & \vdots \\ M_{\mathrm{c}n_\mathrm{c}+1,1} & M_{\mathrm{c}n_\mathrm{c}+1,2} & \dots M_{\mathrm{c}n_\mathrm{c}+1,n_\mathrm{c}+1} \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{c} = \begin{bmatrix} \bar{M}_{\mathrm{c}1,1} & \bar{M}_{\mathrm{c}1,2} & \dots \bar{M}_{\mathrm{c}1,n_\mathrm{c}+1} \\ \bar{M}_{\mathrm{c}2,1} & \bar{M}_{\mathrm{c}2,2} & \dots \bar{M}_{\mathrm{c}2,n_\mathrm{c}+1} \\ \vdots & \vdots & \vdots \\ \bar{M}_{\mathrm{c}n_\mathrm{c},1} & \bar{M}_{\mathrm{c}n_\mathrm{c},2} & \dots \bar{M}_{\mathrm{c}n_\mathrm{c},n_\mathrm{c}+1} \end{bmatrix}$$

$$\mathbf{M}_\mathrm{g} = \begin{bmatrix} M_{\mathrm{g}1,1} & M_{\mathrm{g}1,2} & \dots M_{\mathrm{g}1,n_\mathrm{q}+1} \\ M_{\mathrm{g}2,1} & M_{\mathrm{g}2,2} & \dots M_{\mathrm{g}2,n_\mathrm{q}+1} \\ \vdots & \vdots & \vdots \\ M_{\mathrm{g}n_\mathrm{q},1} & M_{\mathrm{g}n_\mathrm{q},2} & \dots M_{\mathrm{g}n_\mathrm{q},n_\mathrm{q}+1} \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{g} = \begin{bmatrix} \bar{M}_{\mathrm{g}1,1} & \bar{M}_{\mathrm{g}1,2} & \dots \bar{M}_{\mathrm{g}1,n_\mathrm{q}} \\ \bar{M}_{\mathrm{g}2,1} & \bar{M}_{\mathrm{g}2,2} & \dots \bar{M}_{\mathrm{g}2,n_\mathrm{q}} \\ \vdots & \vdots & \vdots \\ \bar{M}_{\mathrm{g}n_\mathrm{q},1} & \bar{M}_{\mathrm{g}n_\mathrm{q},2} & \dots \bar{M}_{\mathrm{g}n_\mathrm{q},n_\mathrm{q}} \end{bmatrix}$$

$$\mathbf{M}_\mathrm{g}^1 = \begin{bmatrix} M_{\mathrm{g}1,1} \\ M_{\mathrm{g}2,1} \\ \vdots \\ M_{\mathrm{g}n_\mathrm{q},1} \end{bmatrix} \quad \mathbf{M}_\mathrm{g}^\mathrm{n} = \begin{bmatrix} M_{\mathrm{g}1,2} & \dots M_{\mathrm{g}1,n_\mathrm{q}+1} \\ M_{\mathrm{g}2,2} & \dots M_{\mathrm{g}2,n_\mathrm{q}+1} \\ \vdots & \vdots \\ M_{\mathrm{g}n_\mathrm{q},2} & \dots M_{\mathrm{g}n_\mathrm{q},n_\mathrm{q}+1} \end{bmatrix}$$

$$\mathbf{w}_\mathrm{c} = \begin{bmatrix} w_{c_1} & 0 & \dots 0 \\ 0 & w_{c_2} & \dots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \dots w_{c_{n_\mathrm{q}+1}} \end{bmatrix} \quad \mathbf{w}_\mathrm{g} = \begin{bmatrix} w_{g_1} & 0 & \dots 0 \\ 0 & w_{g_2} & \dots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \dots w_{g_{n_\mathrm{q}}} \end{bmatrix}$$

$n_\mathrm{p} = 1$

$$\mathbf{M}_\mathrm{c} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{c} = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \mathbf{M}_\mathrm{g} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{g} = \begin{bmatrix} 1 \end{bmatrix}$$

$$\mathbf{w}_\mathrm{g} = \begin{bmatrix} 1 \end{bmatrix} \quad \mathbf{w}_\mathrm{c} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \boldsymbol{\alpha}_\mathrm{g} = \begin{bmatrix} 0.5 \end{bmatrix} \quad \boldsymbol{\alpha}_\mathrm{c} = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

$n_\mathrm{p} = 2$

$$\mathbf{M}_\mathrm{c} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{c} = \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 0.5 & 1 \end{bmatrix}$$

$$\mathbf{M}_\mathrm{g} \approx \begin{bmatrix} 0.4553 & 0.6666 & -0.1220 \\ -0.1220 & 0.6666 & 0.4553 \end{bmatrix} \quad \bar{\mathbf{M}}_\mathrm{g} \approx \begin{bmatrix} 0.7887 & 0.2113 \\ 0.2113 & 0.7887 \end{bmatrix}$$

$$\mathbf{w}_\mathrm{g} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \quad \mathbf{w}_\mathrm{c} \approx \begin{bmatrix} 0.1666 & 0 & 0 \\ 0 & 0.6666 & 0 \\ 0 & 0 & 0.1666 \end{bmatrix}$$

$$\boldsymbol{\alpha}_\mathrm{g} \approx \begin{bmatrix} 0.2113 & 0.7887 \end{bmatrix} \quad \boldsymbol{\alpha}_\mathrm{c} = \begin{bmatrix} 0 & 0.5 & 1 \end{bmatrix}$$

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MIXED INTEGRATORS FOR STRUCTURE-PRESERVING SIMULATIONS IN NONLINEAR STRUCTURAL DYNAMICS

Alexander Janz[1], **Peter Betsch**[1] **and Christian Hesch**[1]

[1]Institute of Mechanics
Karlsruhe Institute of Technology (KIT)
Otto-Ammann-Platz 9
76131 Karlsruhe, Germany
{alexander.janz, peter.betsch, christian.hesch}@kit.edu

Keywords: finite element methods; nonlinear elastodynamics; implicit; structure preserving

Abstract. *The present work deals with the design of structure-preserving numerical methods in the field of nonlinear elastodynamics and structural dynamics. Structure-preserving schemes such as energy-momentum consistent (EMC) methods are known to exhibit superior numerical stability and robustness. Most of the previously developed schemes are relying on a displacement-based variational formulation of the underlying mechanical model. In contrast to that we present a mixed variational framework for the systematic design of EMC schemes. The newly proposed mixed approach accomodates high-performance mixed finite elements such as the brick element due to Kasper & Taylor [15]. Accordingly, the proposed approach makes possible the structure-preserving extension to the dynamic regime of mixed high-performance elements. Numerical examples demonstrate the advantageous properties of the newly developed numerical methods resulting from the structure-preserving discretization in space and time.*

1 Introduction

Energy-momentum consistent (EMC) time-stepping schemes and their energy-decaying variants have been primarily developed in the framework of nonlinear structural dynamics. This type of structure-preserving integrators is known to exhibit superior numerical stability and robustness [12]. Previous developments of EMC schemes have been essentially confined to displacement-based finite elements. We refer to [30, 9, 24, 6, 2, 19, 18, 25],[29, 32, 16, 20, 7, 23, 8, 4] and [28, 11, 17, 5, 1, 13] for representative developments in the framework of nonlinear beams, shells and continua, respectively.

In this connection we remark that often applied finite element techniques such as selectively reduced integration, assumed natural strain interpolations and enhanced assumed strain approximations based on incompatible displacement modes essentially do not affect the design of EMC schemes[1]. In fact, second-order accurate EMC integrators typically rely on a mid-point type discretization of the weak form of the balance of linear momentum in conjunction with the introduction of an algorithmic stress formula [22].

To the best of our knowledge, a general framework for the design of EMC schemes for truly mixed nonlinear finite elements has not been accomplished so far. We refer to [3] for a mixed variational formulation for nonlinear structural dynamics which provides a natural framework for the EMC discretization in time. The EMC scheme in [3] is restricted to mixed finite elements, based on a Hu-Washizu type variational formulation in terms of displacements, Green-Lagrangian strains and conjugated stresses.

In the present work we newly present a Hu-Washizu type mixed variational formulation along with its EMC temporal discretization. Our developments start from a 6-field extension of Hamilton's principle that relies on the displacements, velocities, deformation gradient, right Cauchy-Green deformation tensor and the components of both the 1st and 2nd Piola-Kirchhoff stress tensor. The proposed 6-field variational principle can be linked to Liven's theorem [21, Sec. 26.2] and the Hu-Washizu principle [31]. The corresponding Euler-Lagrange equations provide the mixed variational formulation for the EMC discretization in time. The resulting mixed semi-discrete formulation can be used to develop EMC schemes for mixed finite elements relying on independent interpolations as named above.

We apply the newly developed mixed framework for the design of EMC integrators to the high-performance mixed-enhanced element developed by Kasper & Taylor [15]. This element exhibits superior coarse mesh accuracy along with a locking-free response in the incompressible limit. We refer to [14, 15] for further investigations in this direction. Given the commonly observed numerical stability and robustness of EMC

[1] This remark excludes the broader class of assumed strain methods for nearly incompressible problems in nonlinear elastodynamics investigated in [1], see also [11]

integrators, the mixed-enhanced element [15] can be regarded as ideal candidate for the EMC extension to the transient regime. The resulting structure-preserving numerical method can be expected to allow stable simulations on coarser space-time grids compared to standard methods. Furthermore we even obtain physically meaningful results for quasi incompressible material.

An outline of the rest of the paper is as follows. In Section 2 we present the basic notation and introduce our six-field formulation which provides the Euler-Lagrange equations needed for the subsequent discretization process. The design of a structure-preserving time-stepping scheme is treated in Section 3. The present approach to the design of EMC integrators is then further developed in the framework of the mixed-enhanced formulation in Section 4. After the presentation of numerical examples in Section 5, conclusions are drawn in Section 6.

2 Continuum Mechanics

We start our developments by an introduction of the notation of the applied continuum approach. After that, we introduce our mixed framework for the design of EMC integrators. At the end of this section, the conservation properties will be shown.

2.1 Notation

We consider a motion of a continuum body from the reference configuration $\mathcal{B} \in \mathbb{R}^3$ to the current configuration $\boldsymbol{\varphi}(\mathcal{B})$. This motion can be identified as a time-dependent continuous differentiable function $\boldsymbol{\varphi} \in \mathcal{U}$, where

$$\mathcal{U} := \{\boldsymbol{\varphi} : \mathcal{B} \times [t_0, t_e] \to \mathbb{R}^3 : \boldsymbol{\varphi} \in H^1(\mathcal{B}) \text{ and } \boldsymbol{\varphi}\big|_{\partial\mathcal{B}_\varphi} = \bar{\boldsymbol{\varphi}}\} \tag{1}$$

with the time-interval $[t_0, t_e]$ during the motion. We label material points by $\mathbf{X} \in \mathcal{B}$ and the material velocity is labeled by $\mathbf{V} \in \mathcal{V}$, with

$$\mathcal{V} := \{\delta\boldsymbol{\varphi} : \mathcal{B} \times [t_0, t_e] \to \mathbb{R}^3 : \delta\boldsymbol{\varphi} \in H^1(\mathcal{B}) \text{ and } \delta\boldsymbol{\varphi}\big|_{\partial\mathcal{B}_\varphi} = 0\} \tag{2}$$

Furthermore, the smooth boundary of the body can be partitioned into two parts, where

$$\boldsymbol{\varphi} = \bar{\boldsymbol{\varphi}} \quad \text{on} \quad \partial\mathcal{B}_\varphi \qquad \text{and} \qquad \widetilde{\mathbf{P}}\,\mathbf{N} = \bar{\mathbf{t}} \quad \text{on} \quad \partial\mathcal{B}_\sigma \tag{3}$$

with the prescribed motion of the body $\bar{\boldsymbol{\varphi}}$, the first non-symmetric Piola-Kirchhoff stress tensor $\widetilde{\mathbf{P}} \in \mathcal{M}$, the unit outward normal field $\mathbf{N}$ and the traction vector $\bar{\mathbf{t}}$. The space of $\widetilde{\mathbf{P}}$ is defined as

$$\mathcal{M} = \{\widetilde{\mathbf{P}} : \mathcal{B} \times [t_0, t_e] \to \mathbb{R}^{3\times 3} : \widetilde{P}_{iI} \in L^2(\mathcal{B})\} \tag{4}$$

We assume, that

$$\partial\mathcal{B} = \overline{\partial\mathcal{B}_\varphi \cup \partial\mathcal{B}_\sigma} \quad \text{and} \quad \partial\mathcal{B}_\varphi \cap \partial\mathcal{B}_\sigma = \emptyset \tag{5}$$

To describe the deformation of the body, we introduce the deformation gradient

$$\widetilde{\mathbf{F}} = \mathrm{Grad}\,\boldsymbol{\varphi}(\mathbf{X},t) = \frac{\partial \boldsymbol{\varphi}(\mathbf{X},t)}{\mathbf{X}} = D\boldsymbol{\varphi} \tag{6}$$

and its determinant as $j = \det\widetilde{\mathbf{F}} > 0$, where $\widetilde{\mathbf{F}} \in \mathcal{M}$. The right Cauchy-Green deformation tensor defines a measure of strain and can be written as

$$\widetilde{\mathbf{C}} = \widetilde{\mathbf{F}}^{\mathrm{T}}\,\widetilde{\mathbf{F}} \tag{7}$$

where $\widetilde{\mathbf{C}} \in \mathcal{M}$. Next we introduce the stress tensors that are work conjugate to the deformation gradient $\widetilde{\mathbf{F}}$ and the right Cauchy-Green deformation tensor $\widetilde{\mathbf{C}}$. We use the stored energy function $\hat{W}(\widetilde{\mathbf{F}})$ or $\bar{W}(\widetilde{\mathbf{C}})$ to define the stress and strain relations by

$$\widetilde{\mathbf{P}} = \frac{\partial \hat{W}(\widetilde{\mathbf{F}})}{\partial \widetilde{\mathbf{F}}} \quad \text{and} \quad \widetilde{\mathbf{S}} = 2\,\frac{\partial \bar{W}(\widetilde{\mathbf{C}})}{\partial \widetilde{\mathbf{C}}} \tag{8}$$

with the first Piola-Kirchhoff stress tensor $\widetilde{\mathbf{P}}$ as introduced above and the symmetric second Piola-Kirchhoff stress tensor $\widetilde{\mathbf{S}} \in \mathcal{M}$.

The equations of motion can be derived by imposing the condition of stationarity to

$$S(\boldsymbol{\varphi},\mathbf{V}) = \int_{t_0}^{t_e} \left(L(\boldsymbol{\varphi},\mathbf{V}) + (\dot{\boldsymbol{\varphi}} - \mathbf{V}) \cdot \partial_V L(\boldsymbol{\varphi},\mathbf{V})\right)\,\mathrm{d}t \tag{9}$$

for arbitrary variations of the state-space coordinates $(\boldsymbol{\varphi},\mathbf{V}) \in \mathbb{R}^3 \times \mathbb{R}^3$ in the time interval $[t_0,t_e]$. Note that in (9), Hamiltons principle is linked to Livens theorem (see, for example, Pars [21, Sec. **26.2 Liven's theorem**]). The aforementioned Lagrangian is defined as the difference between the kinetic and potential energy as $L(\boldsymbol{\varphi},\mathbf{V}) = T(\mathbf{V}) - V(\boldsymbol{\varphi})$. The kinetic energy of the body is given by

$$T(\mathbf{V}) = \int_{\mathcal{B}} \frac{1}{2}\mathbf{V} \cdot \rho_0 \mathbf{V}\mathrm{d}V \tag{10}$$

where $\rho_0 : \mathcal{B} \to \mathbb{R}$ is the reference density. We admit the decomposition of the potential energies in an internal and external part as

$$V(\boldsymbol{\varphi}) = V_{\mathrm{int}}(\widetilde{\mathbf{C}}(\boldsymbol{\varphi})) + V_{\mathrm{ext}}(\boldsymbol{\varphi}) \tag{11}$$

The external potential energy can be written as

$$V_{\mathrm{ext}} = -\int_{\mathcal{B}} \bar{\mathbf{b}} \cdot \boldsymbol{\varphi}\,\mathrm{d}V - \int_{\partial\mathcal{B}_\sigma} \bar{\mathbf{t}} \cdot \boldsymbol{\varphi}\,\mathrm{d}A \tag{12}$$

where $\bar{\mathbf{b}} : \mathcal{B} \to \mathbb{R}^3$ is the reference body force. In the following we focus on body forces only, so that the second term in (12) will not be considered in the sequel. For a stored energy function $\bar{W}(\widetilde{\mathbf{C}})$, the potential of the internal energy is assumed to be given by

$$V_{\mathrm{int}} = \int_{\mathcal{B}} \bar{W}(\widetilde{\mathbf{C}}(\boldsymbol{\varphi}))\,\mathrm{d}V \tag{13}$$

2.2 Hu-Wahizu type formulation

Instead of a two-field functional (9), we now consider an extension of the original Hu-Washizu variational [31] to a six-field formulation given by

$$\begin{aligned} S_{\mathrm{HW}}(\boldsymbol{\varphi},\mathbf{V},\mathbf{F},\mathbf{P},\mathbf{C},\mathbf{S}) = \int_{t_0}^{t_e} \Big[T(\mathbf{V}) - \big(V_{\mathrm{int}}(\mathbf{C}) + V_{\mathrm{ext}}(\boldsymbol{\varphi})\big) + (\dot{\boldsymbol{\varphi}} - \mathbf{V}) \cdot DT(\mathbf{V}) \\ - \int_{\mathcal{B}} \big(\mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F}) + \mathbf{S} : \mathbf{g}^S(\mathbf{F},\mathbf{C})\big)\,\mathrm{d}V \Big]\,\mathrm{d}t \\ = \int_{t_0}^{t_e} \int_{\mathcal{B}} \Big((\dot{\boldsymbol{\varphi}} - \frac{1}{2}\mathbf{V})\,\rho_0\,\mathbf{V} \Big)\,\mathrm{d}V\,\mathrm{d}t - \int_{t_0}^{t_e} V_{\mathrm{ext}}(\boldsymbol{\varphi})\,\mathrm{d}t \\ - \int_{t_0}^{t_e} \int_{\mathcal{B}} \big(W(\mathbf{C}) + \mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F}) + \mathbf{S} : \mathbf{g}^S(\mathbf{F},\mathbf{C})\big)\,\mathrm{d}V\,\mathrm{d}t \end{aligned} \tag{14}$$

where the energies are defined by (10), (12) and (13), respectively. The algebraic constraints in (14) can be written as

$$\begin{aligned} \mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F}) &= D\boldsymbol{\varphi} - \mathbf{F} = \mathbf{0} \\ \mathbf{g}^S(\mathbf{F},\mathbf{C}) &= \frac{1}{2}(\mathbf{F}^{\mathrm{T}}\mathbf{F} - \mathbf{C}) = \mathbf{0} \end{aligned} \tag{15}$$

Next to the motion $\boldsymbol{\varphi}$ and velocity field $\mathbf{V}$, the independent quantities are the deformation gradient $\mathbf{F} \in \mathcal{M}$, the first Piola-Kirchhoff stress tensor $\mathbf{P} \in \mathcal{M}$, the right Cauchy-Green deformations tensor $\mathbf{C} \in \mathcal{M}$ and the second Piola-Kirchoff stress tensor $\mathbf{S} \in \mathcal{M}$. The stored energy function $W(\mathbf{C})$ depends on $\mathbf{C}$ only. Note that the stress fields $\mathbf{P}$ and $\mathbf{S}$ play the role of Lagrange multipliers for the enforcement of the algebraic constraints in (15). With the condition of stationarity subjected to the end-point conditions, the variational principle in (14) yields the Euler-Lagrange equations. We obtain the stationarity of (14) step-by-step for each independent fields, where the directional derivative is defined by

$$\delta_\bullet S_{\mathrm{HW}} = D_\bullet S_{\mathrm{HW}}(\delta\bullet) \tag{16}$$

For the variation of the velocity field follows

$$\delta_V S_{\mathrm{HW}} = \delta_V \int_{t_0}^{t_e} \int_{\mathcal{B}} \big(\rho_0\,\mathbf{V}\cdot\dot{\boldsymbol{\varphi}} - \frac{1}{2}\rho_0\,\mathbf{V}\cdot\mathbf{V}\big)\,\mathrm{d}V\,\mathrm{d}t = \int_{t_0}^{t_e} \int_{\mathcal{B}} \delta V \cdot \big(\dot{\boldsymbol{\varphi}} - \mathbf{V}\big)\rho_0\,\mathrm{d}V\,\mathrm{d}t \tag{17}$$

The variation of the motion is given by

$$\delta_\varphi S_{\mathrm{HW}} = \delta_\varphi \int_{t_0}^{t_e} \int_{\mathcal{B}} \big(\mathbf{V}\cdot\dot{\boldsymbol{\varphi}}\,\rho_0 - \mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F})\,\mathrm{d}V - V_{\mathrm{ext}}(\boldsymbol{\varphi})\big)\,\mathrm{d}t \tag{18}$$

$$= -\int_{t_0}^{t_e} \int_{\mathcal{B}} \big(\big(\delta\boldsymbol{\varphi}\,\rho_0\cdot\dot{\mathbf{V}} - \mathbf{P} : D\delta\boldsymbol{\varphi}\big)\,\mathrm{d}V - \delta V_{\mathrm{ext}}(\boldsymbol{\varphi})\big)\,\mathrm{d}t + \int_{\mathcal{B}} \big[\mathbf{V}\cdot\delta\boldsymbol{\varphi}\big]_{t_0}^{t_e}\,\mathrm{d}V \tag{19}$$

where we consider

$$\int_{t_0}^{t_e} \int_{\mathcal{B}} \mathbf{V} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \delta\boldsymbol{\varphi}\, \rho_0 \,\mathrm{d}V \,\mathrm{d}t = -\int_{t_0}^{t_e} \int_{\mathcal{B}} \delta\boldsymbol{\varphi} \cdot \rho_0 \,\dot{\mathbf{V}} \,\mathrm{d}V \,\mathrm{d}t + \int_{\mathcal{B}} \left[\mathbf{V} \cdot \delta\boldsymbol{\varphi}\right]_{t_0}^{t_e} \mathrm{d}V \tag{20}$$

The last term vanishes, because $\delta\boldsymbol{\varphi}(t_0) = 0$ and $\delta\boldsymbol{\varphi}(t_e) = 0$. Next, we obtain the variation with respect to the deformation gradient as

$$\begin{aligned} \delta_F S_{\mathrm{HW}} &= \delta_F \int_{t_0}^{t_e} \int_{\mathcal{B}} \left(\mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi}, \mathbf{F}) + \mathbf{S} : \mathbf{g}^S(\mathbf{F}, \mathbf{C})\right) \mathrm{d}V \,\mathrm{d}t \\ &= \int_{t_0}^{t_e} \int_{\mathcal{B}} \left(\mathbf{P} : (-\delta\mathbf{F}) + \mathbf{S} : \frac{1}{2}(\delta\mathbf{F}^{\mathrm{T}}\,\mathbf{F} + \mathbf{F}^{\mathrm{T}}\delta\mathbf{F})\right) \mathrm{d}V \,\mathrm{d}t \\ &= \int_{t_0}^{t_e} \int_{\mathcal{B}} \delta\mathbf{F} : (-\mathbf{P} + \mathbf{F}\,\mathbf{S}) \,\mathrm{d}V \,\mathrm{d}t \end{aligned} \tag{21}$$

The variation with respect to the right Cauchy-Green deformation tensor leads to

$$\begin{aligned} \delta_C S_{\mathrm{HW}} &= -\delta_C \int_{t_0}^{t_e} \int_{\mathcal{B}} \left(\mathbf{S} : \mathbf{g}^S(\mathbf{F}, \mathbf{C}) + W(\mathbf{C})\right) \mathrm{d}V \,\mathrm{d}t \\ &= -\int_{t_0}^{t_e} \int_{\mathcal{B}} \left(DW(\mathbf{C}) : \delta\mathbf{C} - \frac{1}{2}\mathbf{S} : \delta\mathbf{C}\right) \mathrm{d}V \,\mathrm{d}t \\ &= -\int_{t_0}^{t_e} \int_{\mathcal{B}} \delta\mathbf{C} : \left(DW(\mathbf{C}) - \frac{1}{2}\mathbf{S}\right) \mathrm{d}V \,\mathrm{d}t \end{aligned} \tag{22}$$

For the sake of completeness, the variation with respect to the stress tensors are given by

$$\begin{aligned} \delta_P S_{\mathrm{HW}} &= \delta_P \int_{t_0}^{t_e} \int_{\mathcal{B}} \mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi}, \mathbf{F}) \,\mathrm{d}V \,\mathrm{d}t = \int_{t_0}^{t_e} \int_{\mathcal{B}} \delta\mathbf{P} : \mathbf{g}^P(\boldsymbol{\varphi}, \mathbf{F}) \,\mathrm{d}V \,\mathrm{d}t \\ \delta_S S_{\mathrm{HW}} &= \delta_S \int_{t_0}^{t_e} \int_{\mathcal{B}} \mathbf{S} : \mathbf{g}^S(\mathbf{F}, \mathbf{C}) \,\mathrm{d}V \,\mathrm{d}t = \int_{t_0}^{t_e} \int_{\mathcal{B}} \delta\mathbf{S} : \mathbf{g}^S(\mathbf{F}, \mathbf{C}) \,\mathrm{d}V \,\mathrm{d}t \end{aligned} \tag{23}$$

The variations can also be written in the form of

$$\begin{aligned}\delta S_{\mathrm{HW}}(\boldsymbol{\varphi},\mathbf{V},\mathbf{F},\mathbf{C},\mathbf{P},\mathbf{S}) = &-\int_{t_0}^{t_e}\int_{\mathcal{B}}\left(\rho_0\,(\dot{\mathbf{V}}-\bar{\mathbf{b}})\cdot\delta\boldsymbol{\varphi}+\mathbf{P}:D\delta\boldsymbol{\varphi}\right)\mathrm{d}V\,\mathrm{d}t\\ &+\int_{t_0}^{t_e}\int_{\mathcal{B}}(\dot{\boldsymbol{\varphi}}-\mathbf{V})\cdot\rho_0\,\delta\mathbf{V}\,\mathrm{d}V\,\mathrm{d}t\\ &-\int_{t_0}^{t_e}\int_{\mathcal{B}}\left(\delta\mathbf{P}:\mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F})+\delta\mathbf{S}:\mathbf{g}^S(\mathbf{F},\mathbf{C})\right)\mathrm{d}V\,\mathrm{d}t\\ &+\int_{t_0}^{t_e}\int_{\mathcal{B}}(\mathbf{P}-\mathbf{F}\,\mathbf{S}):\delta\mathbf{F}\,\mathrm{d}V\,\mathrm{d}t\\ &+\int_{t_0}^{t_e}\int_{\mathcal{B}}\left(\frac{1}{2}\,\mathbf{S}-DW(\mathbf{C})\right):\delta\mathbf{C}\,\mathrm{d}V\,\mathrm{d}t=0\end{aligned}\tag{24}$$

where the external potential energy includes the body forces only. This procedure yields the corresponding Euler-Lagrange equations which can be written as

$$\begin{aligned}\int_{\mathcal{B}}\delta\mathbf{V}\cdot\left(\dot{\boldsymbol{\varphi}}-\mathbf{V}\right)\rho_0\,\mathrm{d}V&=0\\ \int_{\mathcal{B}}\left(\delta\boldsymbol{\varphi}\cdot\rho_0\,(\dot{\mathbf{V}}-\bar{\mathbf{b}})+\mathbf{P}:D\delta\boldsymbol{\varphi}\right)\mathrm{d}V&=0\\ \int_{\mathcal{B}}\delta\mathbf{F}:\left(\mathbf{P}-\mathbf{F}\,\mathbf{S}\right)\mathrm{d}V&=0\\ \int_{\mathcal{B}}\delta\mathbf{C}:\left(\mathbf{S}-2\,DW(\mathbf{C})\right)\mathrm{d}V&=0\\ \int_{\mathcal{B}}\delta\mathbf{P}:\left(D\boldsymbol{\varphi}-\mathbf{F}\right)\mathrm{d}V&=0\\ \int_{\mathcal{B}}\delta\mathbf{S}:\left(\frac{1}{2}(\mathbf{F}^{\mathrm{T}}\mathbf{F}-\mathbf{C})\right)\mathrm{d}V&=0\end{aligned}\tag{25}$$

for all admissible variations $\delta\boldsymbol{\varphi}\in\mathcal{V},\delta\mathbf{V}\in\mathcal{V},\delta\mathbf{F}\in\mathcal{M},\delta\mathbf{P}\in\mathcal{M},\delta\mathbf{C}\in\mathcal{M}$ and $\delta\mathbf{S}\in\mathcal{M}$. The aforementioned variational equations are supplemented by the initial conditions $\boldsymbol{\varphi}\big|_{t=t_0}=\boldsymbol{\varphi}_0$, $\mathbf{V}\big|_{t=t_0}=\mathbf{V}_0$ and the consistent initial conditions $\mathbf{F}\big|_{t=t_0}=\mathbf{F}_0$, $\mathbf{P}\big|_{t=t_0}=\mathbf{P}_0$, $\mathbf{C}\big|_{t=t_0}=\mathbf{C}_0$ and $\mathbf{S}\big|_{t=t_0}=\mathbf{S}_0$. Differentiating $(25)_{5,6}$ with respect to time, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{B}}\delta\mathbf{P}:\mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F})\,\mathrm{d}V&=\int_{\mathcal{B}}\delta\mathbf{P}:\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{g}^P(\boldsymbol{\varphi},\mathbf{F})\,\mathrm{d}V=0\\ \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{B}}\delta\mathbf{S}:\mathbf{g}^S(\mathbf{F},\mathbf{C})\,\mathrm{d}V&=\int_{\mathcal{B}}\delta\mathbf{S}:\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{g}^S(\mathbf{F},\mathbf{C})\,\mathrm{d}V=0\end{aligned}\tag{26}$$

Thus the consistency conditions of the mechanical system are defined by

$$\begin{aligned} \int_{\mathcal{B}} \delta\mathbf{P} : \left(D\dot{\boldsymbol{\varphi}} - \dot{\mathbf{F}}\right) \mathrm{d}V = 0 \\ \int_{\mathcal{B}} \delta\mathbf{S} : \left(\dot{\mathbf{F}}^{\mathrm{T}}\mathbf{F} - \frac{1}{2}\dot{\mathbf{C}}\right) \mathrm{d}V = 0 \end{aligned} \tag{27}$$

Equation (25) along with initial conditions define the weak form of the initial-boundary value problem (IBVP) governing the motion of the considered body. The present IBVP determines solutions $(\boldsymbol{\varphi}, \mathbf{V}, \mathbf{F}, \mathbf{P}, \mathbf{C}, \mathbf{S}) : [t_0, t_e] \mapsto \mathcal{U} \times \mathcal{V} \times \mathcal{M} \times \mathcal{M} \times \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}$.

2.3 Balance laws of the continuum

In this section we summarize the important conservation laws in the continuous form which should be preserved in the time and space discrete form. Especially if the investigated problem satisfied specific symmetry conditions, the total linear momentum, $\mathbf{L}(\mathbf{V})$, the total angular momentum, $\mathbf{J}(\boldsymbol{\varphi}, \mathbf{V})$, and the total mechanical energy, $E(\boldsymbol{\varphi}, \mathbf{V}, \mathbf{C})$, are first integrals of the motion. The quantities of interest are defined by

$$\begin{aligned} \mathbf{L}(\mathbf{V}) &= \int_{\mathcal{B}} \rho_0 \, \mathbf{V} \, \mathrm{d}V \\ \mathbf{J}(\boldsymbol{\varphi}, \mathbf{V}) &= \int_{\mathcal{B}} \rho_0 \, \boldsymbol{\varphi} \times \mathbf{V} \, \mathrm{d}V \\ E(\boldsymbol{\varphi}, \mathbf{V}, \mathbf{C}) &= \int_{\mathcal{B}} \left(\frac{1}{2}\, \mathbf{V} \cdot \rho_0 \, \mathbf{V} + W(\mathbf{C})\right) \mathrm{d}V + V_{\mathrm{ext}}(\boldsymbol{\varphi}) \end{aligned} \tag{28}$$

Alternatively, the total mechanical energy can be written in the form

$$E(\boldsymbol{\varphi}, \mathbf{V}, \mathbf{C}) = T(\mathbf{V}) + V_{\mathrm{int}}(\boldsymbol{\varphi}, \mathbf{V}, \mathbf{C}) + V_{\mathrm{ext}}(\boldsymbol{\varphi}) \tag{29}$$

where use has been made of (10), (12) and (13).

2.3.1 Balance of energy

At first we proof the balance of energy by choosing $\delta\mathbf{V} = \dot{\mathbf{V}}$ in $(25)_1$ to obtain

$$\int_{\mathcal{B}} \dot{\boldsymbol{\varphi}} \cdot \dot{\mathbf{V}} \, \rho_0 \, dV = \int_{\mathcal{B}} \mathbf{V} \cdot \dot{\mathbf{V}} \, \rho_0 \, dV = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{B}} \frac{1}{2} \mathbf{V} \cdot \mathbf{V} \, \rho_0 \, \mathrm{d}V \tag{30}$$

Next, choose $\delta\boldsymbol{\varphi} = \dot{\boldsymbol{\varphi}}$ in $(25)_2$ and along with (30) we get

$$\begin{aligned} &\int_{\mathcal{B}} \left(\dot{\boldsymbol{\varphi}} \cdot \rho_0 \left(\dot{\mathbf{V}} - \bar{\mathbf{b}}\right) + D\dot{\boldsymbol{\varphi}} : \mathbf{P}\right) \mathrm{d}V \\ &= \int_{\mathcal{B}} \mathbf{V} \cdot \rho_0 \, \dot{\mathbf{V}} \, \mathrm{d}V - \int_{\mathcal{B}} \mathbf{V} \cdot \rho_0 \, \bar{\mathbf{b}} \, \mathrm{d}V + \int_{\mathcal{B}} D\dot{\boldsymbol{\varphi}} : P \, \mathrm{d}V \\ &= \int_{\mathcal{B}} \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left(\rho_0 \, \mathbf{V} \cdot \mathbf{V}\right) \mathrm{d}V - P_{\mathrm{ext}} + \int_{\mathcal{B}} D\dot{\boldsymbol{\varphi}} : \mathbf{P} \, \mathrm{d}V = 0 \end{aligned} \tag{31}$$

The external power can be evaluated by choosing $\delta\mathbf{V} = \bar{\mathbf{b}}$ in $(25)_1$, so that

$$\int_{\mathcal{B}} \bar{\mathbf{b}} \cdot \dot{\boldsymbol{\varphi}}\, \rho_0\, dV = \int_{\mathcal{B}} \bar{\mathbf{b}} \cdot \mathbf{V}\, \rho_0\, dV = \frac{\mathrm{d}}{\mathrm{d}t} V_{\text{ext}} = P_{\text{ext}} \tag{32}$$

Subsequently we set $\delta\mathbf{P} = \mathbf{P}$ in $(27)_1$ to obtain

$$\int_{\mathcal{B}} \mathbf{P} : D\dot{\boldsymbol{\varphi}}\, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{P} : \dot{\mathbf{F}}\, \mathrm{d}V \tag{33}$$

Choosing $\delta\mathbf{F} = \dot{\mathbf{F}}$ in $(25)_3$ leads to

$$\int_{\mathcal{B}} \dot{\mathbf{F}} : \mathbf{P}\, \mathrm{d}V = \int_{\mathcal{B}} \dot{\mathbf{F}} : (\mathbf{F}\,\mathbf{S})\, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{S} : \left(\dot{\mathbf{F}}^{\mathrm{T}}\mathbf{F}\right)\, \mathrm{d}V \tag{34}$$

For the next step, set $\delta\mathbf{S} = \mathbf{S}$ in $(27)_2$ to get

$$\int_{\mathcal{B}} \mathbf{S} : \left(\dot{\mathbf{F}}^{\mathrm{T}}\mathbf{F}\right)\, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{S} : \frac{1}{2}\dot{\mathbf{C}}\, \mathrm{d}V \tag{35}$$

Eventually, for the choice of $\delta\mathbf{C} = \dot{\mathbf{C}}$ in $(25)_4$ we obtain the result

$$\int_{\mathcal{B}} \frac{1}{2}\dot{\mathbf{C}} : \mathbf{S}\, \mathrm{d}V = \int_{\mathcal{B}} \dot{\mathbf{C}} : DW(\mathbf{C})\, \mathrm{d}V = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{B}} W(\mathbf{C})\, \mathrm{d}V \tag{36}$$

and insert the equations (33)-(36) in (31). Accordingly, the total energy is conserved.

2.3.2 Balance of linear momentum

Define $\delta\boldsymbol{\varphi} = \boldsymbol{\zeta}$, where $\boldsymbol{\zeta} \in \mathbb{R}^3$ is a constant vector, so that $D\boldsymbol{\zeta} = \mathbf{0}$. Note that this corresponds to a rigid translation of the body. Next, with $(25)_2$ we obtain

$$\int_{\mathcal{B}} \boldsymbol{\zeta} \cdot \rho_0 \big(\dot{\mathbf{V}} - \bar{\mathbf{b}}\big)\, \mathrm{d}V = 0 \tag{37}$$

The last equation can be recast in the form

$$\boldsymbol{\zeta} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{B}} \rho_0\, \mathbf{V}\, \mathrm{d}V - \boldsymbol{\zeta} \cdot \mathbf{F}_{\text{ext}} = 0 \tag{38}$$

where

$$\mathbf{F}_{\text{ext}} = \int_{\mathcal{B}} \rho_0\, \bar{\mathbf{b}}\, \mathrm{d}V \tag{39}$$

is the external body force. If we consider $(28)_1$, equation (38) yields

$$\boldsymbol{\zeta} \cdot \big(\dot{\mathbf{L}} - \mathbf{F}_{\text{ext}}\big) = 0 \tag{40}$$

This corresponds to the balance law of linear momentum.

2.3.3 Balance of angular momentum

Next, we choose $\delta\boldsymbol{\varphi} = \boldsymbol{\zeta} \times \boldsymbol{\varphi} = \hat{\boldsymbol{\zeta}}\boldsymbol{\varphi}$ in $(25)_2$. This leads to

$$\begin{aligned} &\int_{\mathcal{B}} \left((\boldsymbol{\zeta} \times \boldsymbol{\varphi}) \cdot \rho_0 \left(\dot{\mathbf{V}} - \bar{\mathbf{b}} \right) + (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}) : \mathbf{P} \right) \mathrm{d}V \\ &= \boldsymbol{\zeta} \cdot \Big(\int_{\mathcal{B}} \boldsymbol{\varphi} \times \dot{\mathbf{V}}\, \rho_0 \, \mathrm{d}V - \int_{\mathcal{B}} \boldsymbol{\varphi} \times \bar{\mathbf{b}}\, \rho_0 \, \mathrm{d}V \Big) + \int_{\mathcal{B}} \hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi} : \mathbf{P} \, \mathrm{d}V = 0 \end{aligned} \tag{41}$$

where $\boldsymbol{\zeta} \in \mathbb{R}^3$ is a constant. Setting $\delta\mathbf{V} = \boldsymbol{\zeta} \times \mathbf{V}$ in $(25)_1$, it follows that

$$\int_{\mathcal{B}} (\boldsymbol{\zeta} \times \mathbf{V}) \cdot (\dot{\boldsymbol{\varphi}} - \mathbf{V})\, \rho_0 \, \mathrm{d}V = \boldsymbol{\zeta} \cdot \left(\int_{\mathcal{B}} \mathbf{V} \times \dot{\boldsymbol{\varphi}}\, \rho_0 \, \mathrm{d}V - \int_{\mathcal{B}} \mathbf{V} \times \mathbf{V}\, \rho_0 \, \mathrm{d}V \right) = 0 \tag{42}$$

Furthermore, it can be observed that

$$\boldsymbol{\zeta} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{B}} \boldsymbol{\varphi} \times \mathbf{V}\, \rho_0 \, \mathrm{d}V = \boldsymbol{\zeta} \cdot \int_{\mathcal{B}} (\dot{\boldsymbol{\varphi}} \times \mathbf{V} + \boldsymbol{\varphi} \times \dot{\mathbf{V}})\, \rho_0 \, \mathrm{d}V = \boldsymbol{\zeta} \cdot \int_{\mathcal{B}} \boldsymbol{\varphi} \times \dot{\mathbf{V}} \, \mathrm{d}V \tag{43}$$

holds. Then we set $\delta\mathbf{P} = \hat{\boldsymbol{\zeta}}\, \mathbf{P}$ in $(25)_5$ to get

$$\int_{\mathcal{B}} (\hat{\boldsymbol{\zeta}}\, \mathbf{P}) : D\boldsymbol{\varphi} \, \mathrm{d}V = \int_{\mathcal{B}} (\hat{\boldsymbol{\zeta}}\, \mathbf{P}) : \mathbf{F} \, \mathrm{d}V \quad \text{or} \quad \int_{\mathcal{B}} \mathbf{P} : (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{P} : (\hat{\boldsymbol{\zeta}}\mathbf{F}) \, \mathrm{d}V \tag{44}$$

Eventually, choose $\delta\mathbf{F} = \hat{\boldsymbol{\zeta}}\, \mathbf{F}$ in $(25)_3$ and note that $\hat{\boldsymbol{\zeta}} = -\hat{\boldsymbol{\zeta}}^{\mathrm{T}}$ to obtain

$$\int_{\mathcal{B}} \mathbf{P} : (\hat{\boldsymbol{\zeta}}\, \mathbf{F}) \, \mathrm{d}V = \int_{\mathcal{B}} \hat{\boldsymbol{\zeta}}\, \mathbf{F} : (\mathbf{F}\, \mathbf{S}) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{S} : (\mathbf{F}^{\mathrm{T}} \hat{\boldsymbol{\zeta}}\, \mathbf{F}) \, \mathrm{d}V = 0 \tag{45}$$

As a result, we recast equation $(25)_2$ under consideration of (41)-(45) to

$$\boldsymbol{\zeta} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{B}} \boldsymbol{\varphi} \times \mathbf{V} \rho_0 \, \mathrm{d}V - \boldsymbol{\zeta} \cdot \mathbf{M}_{\mathrm{ext}} = 0 \tag{46}$$

Associated with $(28)_2$, it follows

$$\boldsymbol{\zeta} \cdot \left(\dot{\mathbf{J}} - \mathbf{M}_{\mathrm{ext}} \right) = 0 \tag{47}$$

where the external moment $\mathbf{M}_{\mathrm{ext}}$ depends on the body forces only and can be written as

$$\mathbf{M}_{\mathrm{ext}} = \int_{\mathcal{B}} \boldsymbol{\varphi} \times \bar{\mathbf{b}}\, \rho_0 \, \mathrm{d}V \tag{48}$$

Because (47) holds for all $\boldsymbol{\zeta} \in \mathbb{R}^3$, the angular momentum is a conserved quantity if $\mathbf{M}_{\mathrm{ext}} = \mathbf{0}$.

3 Discretization in time

We next focus on the design of a structure-preserving discretization in time of the underlying variational formulation (25). In particular, we aim at time-stepping scheme which inherit the important conservation properties described above.

3.1 Energy-momentum consistent time-stepping scheme

Consider a representative time interval $[t_0, t_e]$ with time-step size $\Delta t = t_{n+1} - t_n$, given state space coordinates $(\boldsymbol{\varphi}_n, \mathbf{V}_n) \in \mathbb{R}^3 \times \mathbb{R}^3$, consistent deformation states $\mathbf{F}_n \in \mathcal{M}$, $\mathbf{C}_n \in \mathcal{M}$ and stresses $\mathbf{P}_n \in \mathcal{M}$, $\mathbf{S}_n \in \mathcal{M}$ at t_n. We show below, that a mid-point type discretization of the IBVP results in an energy-momentum scheme, which inherits the important balance laws from the continuous case. The energy-momentum consistent scheme may be formulated as

$$\begin{aligned}
\int_{\mathcal{B}} \delta\mathbf{V} \cdot \frac{1}{\Delta t} \left(\boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_n \right) \rho_0 \, \mathrm{d}V &= \int_{\mathcal{B}} \delta\mathbf{V} \cdot \mathbf{V}_{n+\frac{1}{2}} \, \rho_0 \, \mathrm{d}V \\
\int_{\mathcal{B}} \delta\boldsymbol{\varphi} \cdot \frac{1}{\Delta t} \, \rho_0 \left(\mathbf{V}_{n+1} - \mathbf{V}_n \right) \mathrm{d}V &= - \int_{\mathcal{B}} \mathbf{P}_{n+1} : D\delta\boldsymbol{\varphi} \, \mathrm{d}V + \int_{\mathcal{B}} \delta\boldsymbol{\varphi} \cdot \rho_0 \, \bar{\mathbf{b}} \, \mathrm{d}V \\
\int_{\mathcal{B}} \delta\mathbf{F} : \Big(\mathbf{P}_{n+1} - \mathbf{F}_{n+\frac{1}{2}} \, \mathbf{S}_{n+1} \Big) \, \mathrm{d}V &= 0 \\
\int_{\mathcal{B}} \delta\mathbf{C} : \Big(\mathbf{S}_{n+1} - 2 \, DW(\mathbf{C}_{n+\frac{1}{2}}) \Big) \, \mathrm{d}V &= 0 \\
\int_{\mathcal{B}} \delta\mathbf{P} : \left(D\boldsymbol{\varphi}_{n+1} - \mathbf{F}_{n+1} \right) \mathrm{d}V &= 0 \\
\int_{\mathcal{B}} \delta\mathbf{S} : \left(\frac{1}{2} \Big(\mathbf{F}_{n+1}^{\mathrm{T}} \mathbf{F}_{n+1} - \mathbf{C}_{n+1} \Big) \right) \mathrm{d}V &= 0
\end{aligned} \tag{49}$$

for all $\delta\boldsymbol{\varphi}, \delta\mathbf{V} \in \mathcal{V}$, $\delta\mathbf{F} \in \mathcal{M}$, $\delta\mathbf{P} \in \mathcal{M}$, $\delta\mathbf{C} \in \mathcal{M}$, $\delta\mathbf{S} \in \mathcal{M}$. The mid-point value of the quantity $(\bullet)$ is evaluated by $(\bullet) = \frac{1}{2}((\bullet)_n + (\bullet)_{n+1})$. Then the energy-momentum consistent scheme in (49) determines the state-space coordinates $(\boldsymbol{\varphi}_{n+1}, \mathbf{V}_{n+1}) \in \mathbb{R}^3 \times \mathbb{R}^3$, consistent deformation states $\mathbf{F}_{n+1} \in \mathcal{M}$, $\mathbf{C}_{n+1} \in \mathcal{M}$ at t_{n+1} and the consistent stresses $\mathbf{P}_{n+1} \in \mathcal{M}$, $\mathbf{S}_{n+1} \in \mathcal{M}$ in the time interval $[t_0, t_e]$. As mentioned above, our algorithm is based on the mid-point discretization, however, the constraints are evaluated at the end of each time step. The time discrete counterpart of the condition of consistency $(27)_1$ is given by

$$\int_{\mathcal{B}} \delta\mathbf{P} : \left(D\boldsymbol{\varphi}_{n+1} - D\boldsymbol{\varphi}_n \right) \mathrm{d}V = \int_{\mathcal{B}} \delta\mathbf{P} : \left(\mathbf{F}_{n+1} - \mathbf{F}_n \right) \mathrm{d}V \tag{50}$$

and $(27)_2$ can be written in the time-discrete form as

$$\begin{aligned}
&\int_{\mathcal{B}} \delta\mathbf{S} : \left(\frac{1}{2}\Big(\mathbf{F}_{n+1}^{\mathrm{T}}\mathbf{F}_{n+1} - \mathbf{C}_{n+1} - \Big(\mathbf{F}_{n}^{\mathrm{T}}\mathbf{F}_{n} - \mathbf{C}_{n} \Big) \Big) \right) \mathrm{d}V \\
&= \int_{\mathcal{B}} \delta\mathbf{S} : \left(\mathbf{F}_{n+\frac{1}{2}}^{\mathrm{T}} \left(\mathbf{F}_{n+1} - \mathbf{F}_{n} \right) - \frac{1}{2}(\mathbf{C}_{n+1} - \mathbf{C}_{n}) \right) \mathrm{d}V \\
&= \quad 0
\end{aligned} \tag{51}$$

The specific temporal discretization of the IBVP results in the algorithm (49) which defines an energy momentum consistent scheme. Consequently the physical quantities in (28) are consistently approximated independent of the time-step size.

Remark 1 *In the present work we focus on St. Venant-Kirchhoff type material models. It can be easily verified, that*

$$DW(\mathbf{C}_{n+\frac{1}{2}}) : (\mathbf{C}_{n+1} - \mathbf{C}_{n}) = W(\mathbf{C}_{n+1}) - W(\mathbf{C}_{n}) \tag{52}$$

In case of material models suitable for finite elasticity, $DW(\mathbf{C}_{n+\frac{1}{2}})$ *in* (49) *has to be replaced by an appropriate discrete derivative* $\bar{D}W(\mathbf{C}_{n+\frac{1}{2}})$ *in the sense of Gonzalez [10, 11].*

3.2 Discrete balance laws

In this section, we show that the EMC scheme exactly inherited the significant consistency (and conservation) properties from the underlying continuous formulation presented in section 2.3.

3.2.1 Discrete conservation of energy

We start with the proof of the energy conservation by choosing $\delta\mathbf{V} = \mathbf{V}_{n+1} - \mathbf{V}_{n}$ in $(49)_1$ to obtain

$$\begin{aligned}
\int_{\mathcal{B}} (\mathbf{V}_{n+1} - \mathbf{V}_{n}) \cdot \frac{1}{\Delta t}\, \rho_0 \left(\boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_{n} \right) &= \int_{\mathcal{B}} (\mathbf{V}_{n+1} - \mathbf{V}_{n}) \cdot \mathbf{V}_{n+\frac{1}{2}}\, \rho_0 \, \mathrm{d}V \\
&= \frac{1}{2} \int_{\mathcal{B}} (\mathbf{V}_{n+1} \cdot \mathbf{V}_{n+1} - \mathbf{V}_{n} \cdot \mathbf{V}_{n})\, \rho_0 \, \mathrm{d}V \\
&= T(\mathbf{V}_{n+1}) - T(\mathbf{V}_{n})
\end{aligned} \tag{53}$$

Next we choose $\delta\boldsymbol{\varphi} = \boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_{n}$ in $(49)_2$ and observe that

$$\begin{aligned}
&\int_{\mathcal{B}} (\boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_{n}) \cdot \frac{1}{\Delta t}\, \rho_0 \left(\mathbf{V}_{n+1} - \mathbf{V}_{n} \right) \mathrm{d}V \\
&\qquad = - \int_{\mathcal{B}} \mathbf{P}_{n+1} : (D\boldsymbol{\varphi}_{n+1} - D\boldsymbol{\varphi}_{n})\, \mathrm{d}V + \int_{\mathcal{B}} (\boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_{n}) \cdot \rho_0\, \bar{\mathbf{b}}\, \mathrm{d}V
\end{aligned} \tag{54}$$

We notice, that for the discrete form of the external potential energy

$$\int_{\mathcal{B}} (\boldsymbol{\varphi}_{n+1} - \boldsymbol{\varphi}_n) \cdot \rho_0 \, \bar{\mathbf{b}} \, \mathrm{d}V = V_{\mathrm{ext}}(\boldsymbol{\varphi}_{n+1}) - V_{\mathrm{ext}}(\boldsymbol{\varphi}_n) \tag{55}$$

holds. Then we use $\delta\mathbf{P} = \mathbf{P}_{n+1}$ in (50) to get

$$\int_{\mathcal{B}} \mathbf{P}_{n+1} : (D\boldsymbol{\varphi}_{n+1} - D\boldsymbol{\varphi}_n) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{P}_{n+1} : (\mathbf{F}_{n+1} - \mathbf{F}_n) \, \mathrm{d}V \tag{56}$$

After that, set $\delta\mathbf{F} = \mathbf{F}_{n+1} - \mathbf{F}_n$ in $(49)_3$ to obtain

$$\begin{aligned} \int_{\mathcal{B}} \mathbf{P}_{n+1} : (\mathbf{F}_{n+1} - \mathbf{F}_n) \, \mathrm{d}V &= \int_{\mathcal{B}} (\mathbf{F}_{n+1} - \mathbf{F}_n) : \mathbf{F}_{n+\frac{1}{2}} \, \mathbf{S}_{n+1} \, \mathrm{d}V \\ &= \int_{\mathcal{B}} \mathbf{S}_{n+1} : \mathbf{F}^{\mathrm{T}}_{n+\frac{1}{2}} \, (\mathbf{F}_{n+1} - \mathbf{F}_n) \, \mathrm{d}V \end{aligned} \tag{57}$$

For the variation of the second Piola-Kichhoff stress tensor we choose $\delta\mathbf{S} = \mathbf{S}_{n+1}$ in (51) and as a result, we get

$$\int_{\mathcal{B}} \mathbf{S}_{n+1} : \mathbf{F}^{\mathrm{T}}_{n+\frac{1}{2}} \, (\mathbf{F}_{n+1} - \mathbf{F}_n) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{S}_{n+1} : \frac{1}{2} (\mathbf{C}_{n+1} - \mathbf{C}_n) \, \mathrm{d}V \tag{58}$$

Accordingly, choosing $\delta\mathbf{C} = \mathbf{C}_{n+1} - \mathbf{C}_n$ in $(49)_4$ leads to

$$\int_{\mathcal{B}} \mathbf{S}_{n+1} : \frac{1}{2} (\mathbf{C}_{n+1} - \mathbf{C}_n) \, \mathrm{d}V = \int_{\mathcal{B}} DW(\mathbf{C}_{n+\frac{1}{2}}) : (\mathbf{C}_{n+1} - \mathbf{C}_n) \, \mathrm{d}V \tag{59}$$

As a result, for (54) along with (53), (55) and (56)-(59) we obtain

$$T(\mathbf{V}_{n+1}) + V_{\mathrm{int}}(\mathbf{C}_{n+1}) + V_{\mathrm{ext}}(\boldsymbol{\varphi}_{n+1}) = T(\mathbf{V}_n) + V_{\mathrm{int}}(\mathbf{C}_n) + V_{\mathrm{ext}}(\boldsymbol{\varphi}_n) \tag{60}$$

Thus, the EMC scheme reproduces the total mechanical energy exactly, independent of the time-step size. Note that in the last equation the relationship (52) has been used.

3.2.2 Discrete balance of linear momentum

We next choose $\delta\boldsymbol{\varphi} = \boldsymbol{\zeta}$ where $\boldsymbol{\zeta} \in \mathbb{R}^3$ is constant so that, $D\boldsymbol{\zeta} = \mathbf{0}$. Now, its straightforward to show that $(49)_2$ yields the result

$$\int_{\mathcal{B}} \boldsymbol{\zeta} \cdot \rho_0 \left(\frac{1}{\Delta t} (\mathbf{V}_{n+1} - \mathbf{V}_n) - \bar{\mathbf{b}} \right) \mathrm{d}V = 0 \tag{61}$$

which we can recast by $(28)_1$ in the form

$$\boldsymbol{\zeta} \cdot \left(\mathbf{L}(\mathbf{V}_{n+1}) - L(\mathbf{V}_n) + \Delta t \, \mathbf{F}_{\mathrm{ext}} \Big|_{n+\alpha} \right) = 0 \tag{62}$$

where

$$\mathbf{F}_{\text{ext}}\Big|_{n+\alpha} = \int_{\mathcal{B}} \rho_0 \, \bar{\mathbf{b}} \, \mathrm{d}V \tag{63}$$

are the external body forces in the time interval $[t_n, t_{n+1}]$. This is the discrete version of the linear momentum balance corresponding to $(28)_1$. Hence, the linear momentum is a conserved quantity in $[t_n, t_{n+1}]$ for any $\alpha \in [0, 1]$ if no external forces are acting on the body.

3.2.3 Discrete balance of angular momentum

We now choose $\delta\boldsymbol{\varphi} = (\boldsymbol{\zeta} \times \boldsymbol{\varphi}_{n+\frac{1}{2}}) = \hat{\boldsymbol{\zeta}}\, \boldsymbol{\varphi}_{n+\frac{1}{2}}$ in $(49)_2$. We conclude that

$$\begin{aligned}
&\int_{\mathcal{B}} (\boldsymbol{\zeta} \times \boldsymbol{\varphi}_{n+\frac{1}{2}}) \cdot \frac{1}{\Delta t} \, \rho_0 \, (\mathbf{V}_{n+1} - \mathbf{V}_n) \, \mathrm{d}V \\
&\qquad + \int_{\mathcal{B}} \mathbf{P}_{n+1} : (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}_{n+\frac{1}{2}}) \, \mathrm{d}V - \int_{\mathcal{B}} (\boldsymbol{\zeta} \times \boldsymbol{\varphi}_{n+\frac{1}{2}}) \cdot \rho_0 \, \bar{\mathbf{b}} \, \mathrm{d}V \\
&= \boldsymbol{\zeta} \cdot \left(\int_{\mathcal{B}} \boldsymbol{\varphi}_{n+\frac{1}{2}} \times (\mathbf{V}_{n+1} - \mathbf{V}_n) \frac{1}{\Delta t} \, \rho_0 \, \mathrm{d}V - \int_{\mathcal{B}} \boldsymbol{\varphi}_{n+\frac{1}{2}} \times \bar{\mathbf{b}} \, \rho_0 \, \mathrm{d}V \right) \\
&\qquad + \int_{\mathcal{B}} \hat{\boldsymbol{\zeta}}\, \mathbf{P}_{n+1} : D\boldsymbol{\varphi}_{n+\frac{1}{2}} \, \mathrm{d}V \\
&= 0
\end{aligned} \tag{64}$$

Then we choose $\delta\mathbf{P}_{n+1} = \hat{\boldsymbol{\zeta}}\mathbf{P}_{n+1}$ in $(49)_5$ to obtain

$$\int_{\mathcal{B}} (\hat{\boldsymbol{\zeta}}\, \mathbf{P}_{n+1}) : D\boldsymbol{\varphi}_{n+1} \, \mathrm{d}V = \int_{\mathcal{B}} \hat{\boldsymbol{\zeta}}\, \mathbf{P}_{n+1} : \mathbf{F}_{n+1} \, \mathrm{d}V \tag{65}$$

or equally

$$\int_{\mathcal{B}} \mathbf{P}_{n+1} : (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}_{n+1}) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{P}_{n+1} : (\hat{\boldsymbol{\zeta}}\, \mathbf{F}_{n+1}) \, \mathrm{d}V \tag{66}$$

Similarly, for the constraints at t_n with $\delta\mathbf{P} = \hat{\boldsymbol{\zeta}}\, \mathbf{P}_{n+1}$ we get

$$\int_{\mathcal{B}} \mathbf{P}_{n+1} : (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}) \, \mathrm{d}V = \int_{\mathcal{B}} \mathbf{P}_{n+1} (\hat{\boldsymbol{\zeta}}\, \mathbf{F}_n) \, \mathrm{d}V \tag{67}$$

We take the average of the equations from above as $\frac{1}{2}[(66) + (67)]$ to get

$$\begin{aligned}
\int_{\mathcal{B}} \mathbf{P}_{n+1} : (\hat{\boldsymbol{\zeta}}\, D\boldsymbol{\varphi}_{n+\frac{1}{2}}) \, \mathrm{d}V &= \int_{\mathcal{B}} \mathbf{P}_{n+1} : \left(\hat{\boldsymbol{\zeta}}\, \frac{1}{2} (\mathbf{F}_n + \mathbf{F}_{n+1}) \right) \, \mathrm{d}V \\
&= \int_{\mathcal{B}} \mathbf{P}_{n+1} : \left(\hat{\boldsymbol{\zeta}}\, \mathbf{F}_{n+\frac{1}{2}} \right) \mathrm{d}V
\end{aligned} \tag{68}$$

Now let $\delta\mathbf{F} = \hat{\boldsymbol{\zeta}}\,\frac{1}{2}(\mathbf{F_n} + \mathbf{F}_{n+1})$ in $(49)_3$ to verify that

$$\begin{aligned}\int_{\mathcal{B}} \mathbf{P}_{n+1} : \left(\hat{\boldsymbol{\zeta}}\,\frac{1}{2}(\mathbf{F}_n + \mathbf{F}_{n+1})\right) \mathrm{d}V &= \int_{\mathcal{B}} \mathbf{F}_{n+\frac{1}{2}}\,\mathbf{S}_{n+1} : \left(\hat{\boldsymbol{\zeta}}\,\frac{1}{2}(\mathbf{F}_n + \mathbf{F}_{n+1})\right) \mathrm{d}V \\ &= \int_{\mathcal{B}} \mathbf{S}_{n+1} : \mathbf{F}^{\mathrm{T}}_{n+\frac{1}{2}} \left(\hat{\boldsymbol{\zeta}}\,\frac{1}{2}(\mathbf{F}_n + \mathbf{F}_{n+1})\right) \mathrm{d}V \\ &= \quad 0\end{aligned} \tag{69}$$

Consider equation (64) along with (65) - (69) to get:

$$\boldsymbol{\zeta} \cdot \left(\mathbf{J}(\boldsymbol{\varphi}_{\mathbf{n+1}}, \mathbf{V}_{n+1}) - \mathbf{J}(\boldsymbol{\varphi}_n, \mathbf{V}_n) + \Delta t \mathbf{M}_{\mathrm{ext}}\Big|_{n+\frac{1}{2}}\right) = 0 \tag{70}$$

This is the discrete version of the angular momentum balance corresponding to $(28)_2$. Note, that the external moment is evaluated at the mid-point configuration as

$$\mathbf{M}_{\mathrm{ext}}\Big|_{n+\frac{1}{2}} = \int_{\mathcal{B}} \boldsymbol{\varphi}_{n+\frac{1}{2}} \times \rho_0\, \bar{\mathbf{b}}\, \mathrm{d}V \tag{71}$$

4 Discretization in space

In this section, we deal with the finite element method to accomplish a numerical solution for the semi-discrete mechanical system. We start our developments by the approximation of the body in its reference configuration by

$$\mathcal{B} \approx \mathcal{B}^h = \bigcup_{e}^{n_{el}} \Omega_e \tag{72}$$

Accordingly the boundaries are approximated by $\partial\mathcal{B} \approx \partial\mathcal{B}^h = \bigcup_e^{n_{el}} \partial\Omega_e$. Next, we make use of standard isoparametric finite elements for the finite dimensional approximation of $\boldsymbol{\varphi}$, and the velocity $\mathbf{V}$ as

$$\boldsymbol{\varphi}^h(\mathbf{X}, t) = \sum_{A=1}^{n^e_{node}} N_A(\mathbf{X})\, \boldsymbol{\varphi}^A \quad \text{and} \quad \mathbf{V}^h(\mathbf{X}, t) = \sum_{A=1}^{n^e_{node}} N_A(\mathbf{X})\, \mathbf{V}^A \tag{73}$$

where n^e_{node} are the nodes of the element Ω_e and $N_A(\mathbf{X}) : \mathcal{B} \to \mathbb{R}$ are the standard isoparametric shape functions associated with node A. Note that the nodal vectors of an element are denoted by $\boldsymbol{\varphi}^A \in \mathbb{R}^3$ and $\mathbf{V}^A \in \mathbb{R}^3$. In the present work we deal with trilinear shapefunctions with the property that $N_A(\mathbf{X}_B) = \delta^B_A$. The finite element spaces are given due to the approximation in (73) by $\mathcal{U}^h \subset \mathcal{U}$ and $\mathcal{V}^h \subset \mathcal{V}$.

By inserting the spatial interpolations in (6), the approximation of the deformation gradient is given by

$$\widetilde{\mathbf{F}} = \frac{\partial \boldsymbol{\varphi}(\mathbf{X}, t)}{\partial \mathbf{X}} = \sum_{A=1}^{n^e_{node}} \boldsymbol{\varphi}^A \otimes \nabla_X N_A(\mathbf{X}) \tag{74}$$

Next, we define the kinetic energy of an element by (10) with (73) as

$$T = \frac{1}{2} \sum_{A,B=1}^{n^e_{node}} \mathbf{V}^A \cdot \mathbf{M}^{AB}_{\rho} \mathbf{V}^B \quad \text{with} \quad \mathbf{M}^{AB}_{\rho} = \int_{\Omega_e} \rho_0 \, N_A N_B \, \mathrm{d}V \, \mathbf{I} \tag{75}$$

where $\mathbf{I} \in \mathbb{R}^{3\times3}$ represents the identity matrix. The external potential energy, given in (12) can be expresses in the discrete form as

$$V_{ext} = -\sum_{A=1}^{n^e_{node}} \boldsymbol{\varphi}^A \cdot \mathbf{F}^A_{ext} \quad \text{where} \quad \mathbf{F}^A_{ext} = \int_{\Omega_e} N^A \, \bar{\mathbf{b}} \, \mathrm{d}V + \int_{\partial\Omega_e} N^A \, \bar{\mathbf{t}} \, \mathrm{d}A \tag{76}$$

where $\mathbf{F}^A_{ext} \in \mathbb{R}^3$ denotes the external forces associated with node A.

4.1 Mixed approximation

In the present work the discretization of the mixed quantities follows the approach by Kasper & Taylor [15]. They developed a mixed-enhanced finite element with mixed stress field $\mathbf{P}$ and an mixed-enhanced strain field $\mathbf{F}$. The other independent quantities, namely the right Cauchy-Green deformation tensor $\mathbf{C}$ and the second Piola-Kirchhoff stress tensor $\mathbf{S}$ are calculated by the constraints given in $(14)_{5,6}$. Accordingly, we are very flexible in the design of an EMC algorithm as we show below.

Consider a mixed stress and strain field, formulated in the isoparametric space and then use a standard transform to define the approximations in the physical space. The approximation of the deformation gradient is defined by

$$\mathbf{F} = \boldsymbol{\gamma}_0 + \frac{1}{j(\boldsymbol{\xi})} \mathbf{F}_0 \, \mathbf{T} [\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi}, \boldsymbol{\gamma}) - \boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi}, \boldsymbol{\alpha})] \mathbf{T}^{\mathrm{T}} \tag{77}$$

For the approximation of the first Piola-Kirchhoff stress tensor we obtain

$$\mathbf{P} = \boldsymbol{\beta}_0 + \mathbf{F}_0^{-\mathrm{T}} \, \mathbf{T}^{-\mathrm{T}} [\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi}, \boldsymbol{\beta})] \mathbf{T}^{-1} \tag{78}$$

where $\mathbf{F} \in \mathcal{M}^h$, $\mathbf{P} \in \mathcal{M}^h$ with $\mathcal{M}^h \subset \mathcal{M}$. On element level we have 26 additional local parameters for the mixed stress, where $\boldsymbol{\beta}_0 \in \mathbb{R}^{3\times3}$ ensures constant stress states within the element and $\boldsymbol{\beta} \in \mathbb{R}^{15}$ reflects the varying stress field. In contrast, the mixed-enhanced deformation gradient is approximated by 35 local parameters. Similar to the mixed stress we have $\boldsymbol{\gamma}_0 \in \mathbb{R}^{3\times3}$ parameters for the constant part and $\boldsymbol{\gamma} \in \mathbb{R}^{15}$ parameters for the varying part. In addition, the enhanced parameters $\boldsymbol{\alpha} \in \mathbb{R}^9$ were added to improve the element performance. The interpolations $\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi}, \bullet)$ and $\boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi}, \boldsymbol{\alpha})$ can be found in Appendix A.1. The approximation of the mixed quantities contains $\mathbf{T}$ and $\mathbf{F}_0$ to ensure constant states, minimize the order of quadrature and reduce the sensitivity to

Matrix	i	1	2	3	1	2	1	2	3	3
	j	1	2	3	2	3	3	1	2	1
Vector	i	1	2	3	4	5	6	7	8	9

Table 1: Matrix-vector notation

initially distorted elements. Note that $\mathbf{F}_0$ ensures an objective formulation to a superposed rigid body motion. Theses quantities are evaluated at the center of the element as

$$\mathbf{F}_0 = \mathrm{Grad}\boldsymbol{\varphi}^h\big|_{\xi=0} = \sum_{A=1}^{n^e_{node}} \boldsymbol{\varphi}^A \otimes \nabla_0 N_A(\mathbf{X}) \tag{79}$$

and

$$\mathbf{T} = \mathbf{J}\big|_{\xi=0} \quad \text{where} \quad \mathbf{J} = \mathrm{Grad}_\xi \mathbf{X}^h, \quad j(\boldsymbol{\xi}) = \mathrm{det}\mathbf{J} \tag{80}$$

Remark 2 *Alternatively,* $\mathbf{T}$ *and* $\mathbf{F}_0$ *can be calculated as the average over the element by*

$$\mathbf{T} = \frac{1}{\Omega_e}\int_{\Omega_e} \mathbf{J}(\boldsymbol{\xi})\,\mathrm{d}V, \qquad \mathbf{F}_0 = \frac{1}{\Omega_e}\int_{\Omega_e} \widetilde{\mathbf{F}}\,\mathrm{d}V \tag{81}$$

Note that in the two dimensional case both formulations are identical.

4.2 Efficient formulation and variation

We start this section with an matrix-vector transformation to make use of the linearity of the element parameters. Then we apply the variation of the mixed quantities.

4.2.1 Efficient formulation of the approximations

Since the local parameters are linear within the element, an efficient reformulation is possible. Therefore we may write

$$\begin{aligned}
\widetilde{\mathbf{F}} = \sum_{A=1}^{n^e_{node}} \boldsymbol{\varphi}^A \otimes \nabla_X N_A(\mathbf{X}) \;&\xrightarrow{i,j\to i}\; \widetilde{\mathbf{F}}^{\mathrm{v}} = \mathbf{B}\,\boldsymbol{\varphi}^v \\
\mathbf{F} = \boldsymbol{\gamma}_0 + \frac{1}{j(\boldsymbol{\xi})}\mathbf{F}_0\,\mathbf{T}[\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\gamma}) - \boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi},\boldsymbol{\alpha})]\mathbf{T}^{\mathrm{T}} \;&\xrightarrow{i,j\to i}\; \mathbf{F}^{\mathrm{v}} = \boldsymbol{\gamma}_0^v + \mathbf{B}_\gamma(\boldsymbol{\varphi})\,\boldsymbol{\gamma} - \mathbf{B}_\alpha(\boldsymbol{\varphi})\,\boldsymbol{\alpha} \\
\mathbf{P} = \boldsymbol{\beta}_0 + \mathbf{F}_0^{-\mathrm{T}}\,\mathbf{T}^{-\mathrm{T}}[\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\beta})]\mathbf{T}^{-1} \;&\xrightarrow{i,j\to i}\; \mathbf{P}^{\mathrm{v}} = \boldsymbol{\beta}_0^v + \mathbf{B}_\beta(\boldsymbol{\varphi})\,\boldsymbol{\beta}
\end{aligned} \tag{82}$$

where the ordering for the matrix-vector notation is given in Table 1. Due to the formulation in (82), the approximations can be written in a vectorial form as $\widetilde{\mathbf{F}}^v \in \mathbb{R}^9, \mathbf{F}^v \in \mathbb{R}^9$

or $\mathbf{P}^v \in \mathbb{R}^9$, respectively. Then we can calculate them along with the element operator-matrices $\mathbf{B} \in \mathbb{R}^{9\times 24}$, $\mathbf{B}_\gamma(\boldsymbol{\varphi}) \in \mathbb{R}^{9\times 15}$, $\mathbf{B}_\alpha(\boldsymbol{\varphi}) \in \mathbb{R}^{9\times 9}$ or $\mathbf{B}_\beta(\boldsymbol{\varphi}) \in \mathbb{R}^{9\times 15}$ and the local element parameters $\boldsymbol{\gamma}_0^v \in \mathbb{R}^9$ and $\boldsymbol{\beta}_0^v \in \mathbb{R}^9$ in the vectorial form. Note, that the nodal vectors are summarized to $\boldsymbol{\varphi}^v = [\boldsymbol{\varphi}^{1\mathrm{T}} ... \boldsymbol{\varphi}^{8\mathrm{T}}]^{\mathrm{T}}$.

4.2.2 Variations

Consider the variations of the equation given in (82) with respect to both the local element parameters and nodal parameters. For the variation to the one mentioned first we obtain

$$\begin{aligned} \delta_{\gamma_0}\mathbf{F}^{\mathrm{v}} &= \delta\boldsymbol{\gamma}_0^v & \qquad \delta_{\beta_0}\mathbf{P}^{\mathrm{v}} &= \delta\boldsymbol{\beta}_0^v \\ \delta_{\gamma}\mathbf{F}^{\mathrm{v}} &= \mathbf{B}_\gamma(\boldsymbol{\varphi})\,\delta\boldsymbol{\gamma}, & \qquad \delta_{\beta}\mathbf{P}^{\mathrm{v}} &= \mathbf{B}_\beta(\boldsymbol{\varphi})\,\delta\boldsymbol{\beta} \\ \delta_{\alpha}\mathbf{F}^{\mathrm{v}} &= -\mathbf{B}_\alpha(\boldsymbol{\varphi})\,\delta\boldsymbol{\alpha} \end{aligned} \tag{83}$$

For the variation with respect to the nodal parameters we get

$$\begin{aligned} \delta_q\widetilde{\mathbf{F}} &= \sum_{A=1}^{n^e_{node}} \delta\boldsymbol{\varphi}^A \otimes \nabla_X N_A(\mathbf{X}) \\ \delta_q\mathbf{F} &= \Big(\sum_{A=1}^{n^e_{node}} \delta\boldsymbol{\varphi}^A \otimes \nabla_0 N_A\Big)\,\mathbf{T}\,(\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\gamma}) - \boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi},\boldsymbol{\alpha}))\,\mathbf{T}^{\mathrm{T}} \\ \delta_q\mathbf{P} &= -\mathbf{F}_0^{-\mathrm{T}}\Big(\sum_{A=1}^{n^e_{node}} \delta\boldsymbol{\varphi}^A \otimes \nabla_0 N_A\Big)^{\mathrm{T}}\mathbf{F}_0^{-\mathrm{T}}\,\mathbf{T}^{-\mathrm{T}}\,\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\beta})\,\mathbf{T}^{-1} \end{aligned} \tag{84}$$

4.3 Fully discrete system

The present implementation differs from the original approach by Kasper & Taylor since the mixed strain parameters $\boldsymbol{\gamma}_0, \boldsymbol{\gamma}$ are not solved by the nodal parameters directly. Thus, we facilitate the implementation and further improve the clarity of the underlying formulation as we see below. Furthermore, a direct variational stress recovery for post-processing is available. Since a static condensation process is still possible, the number of global unknowns are identical and thus, the computational effort is quiet similar. Next, we insert the approximated quantities in (49) and taking the temporal evaluation

into account to get the fully discrete system given by

$$
\begin{aligned}
\delta\mathbf{V}_A \cdot M^{AB}\Big[(\boldsymbol{\varphi}_{Bn+1} - \boldsymbol{\varphi}_{Bn}) - \Delta t\, \mathbf{V}_{Bn+\frac{1}{2}}\Big] = 0 \\
\delta\boldsymbol{\varphi}_A \cdot \Big[\mathbf{M}_\rho^{AB}\, \frac{1}{\Delta t}(\mathbf{V}_{Bn+1} - \mathbf{V}_{Bn}) - \mathbf{F}^A_{\text{ext}} + \int_{\Omega_e} \mathbf{P}_{n+1}\, \nabla N^A \,\mathrm{d}V \\
+ \int_{\Omega_e} (\mathbf{P}_{n+1} - \mathbf{F}_{n+\frac{1}{2}}\, 2\, DW(\mathbf{C}_{n+\frac{1}{2}}))\, \frac{1}{j(\boldsymbol{\xi})} T(\boldsymbol{\mathcal{E}}(\boldsymbol{\xi}, \boldsymbol{\gamma}_{n+\frac{1}{2}}))^{\mathrm{T}}\, T^{\mathrm{T}}\, \nabla_0 N^A \,\mathrm{d}V \\
- \int_{\Omega_e} (\mathbf{P}_{n+1} - \mathbf{F}_{n+\frac{1}{2}}\, 2\, DW(\mathbf{C}_{n+\frac{1}{2}}))\, \frac{1}{j(\boldsymbol{\xi})} T(\boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi}, \boldsymbol{\alpha}_{n+\frac{1}{2}}))^{\mathrm{T}}\, T^{\mathrm{T}}\, \nabla_0 N^A \,\mathrm{d}V \\
- \int_{\Omega_e} \mathbf{F}_{0\,n+\frac{1}{2}}^{-\mathrm{T}}\, T^{-\mathrm{T}}\, \boldsymbol{\mathcal{E}}(\boldsymbol{\xi}, \boldsymbol{\beta}_{n+1})\, T^{-1}\, (\widetilde{\mathbf{F}}_{n+\frac{1}{2}} - \mathbf{F}_{n+\frac{1}{2}})^{\mathrm{T}}\, \mathbf{F}_{0\,n+\frac{1}{2}}^{-\mathrm{T}}\, \nabla_0 N^A \,\mathrm{d}V\Big] = 0 \\
\delta\boldsymbol{\gamma}_0^v \cdot \Big[\int_{\Omega_e} \mathbf{P}^{\mathrm{v}}_{n+1} - \mathbf{F}^{\mathrm{v}}_{n+\frac{1}{2}}\, 2\, DW(\mathbf{C}^{\mathrm{v}}_{n+\frac{1}{2}})\,\mathrm{d}V\Big] = 0 \\
\delta\boldsymbol{\gamma} \cdot \Big[\int_{\Omega_e} \mathbf{B}_{\gamma n+\frac{1}{2}} \Big(\mathbf{P}^{\mathrm{v}}_{n+1} - \mathbf{F}^{\mathrm{v}}_{n+\frac{1}{2}}\, 2\, DW(\mathbf{C}^{\mathrm{v}}_{n+\frac{1}{2}})\Big)\,\mathrm{d}V\Big] = 0 \\
\delta\boldsymbol{\alpha} \cdot \Big[\int_{\Omega_e} -\mathbf{B}_{\alpha n+\frac{1}{2}} \Big(\mathbf{P}^{\mathrm{v}}_{n+1} - \mathbf{F}^{\mathrm{v}}_{n+\frac{1}{2}}\, 2\, DW(\mathbf{C}^{\mathrm{v}}_{n+\frac{1}{2}})\Big)\,\mathrm{d}V\Big] = 0 \\
\delta\boldsymbol{\beta}_0^v \cdot \Big[\int_{\Omega_e} \big(\widetilde{\mathbf{F}}^{\mathrm{v}}_{n+1} - \mathbf{F}^{\mathrm{v}}_{n+1}\big)\,\mathrm{d}V\Big] = 0 \\
\delta\boldsymbol{\beta} \cdot \Big[\int_{\Omega_e} \mathbf{B}_{\beta n+1} \Big(\widetilde{\mathbf{F}}^{\mathrm{v}}_{n+1} - \mathbf{F}^{\mathrm{v}}_{n+1}\Big)\,\mathrm{d}V\Big] = 0
\end{aligned}
\tag{85}
$$

with

$$
M^{AB} = \int_{\Omega_e} N^A\, N^B \,\mathrm{d}V \tag{86}
$$

Due to the sophisticated element formulation, we have to go into detail with the temporal evaluation of the mixed quantities. As mentioned above, the mixed stress is assumed to be constant within a time-step. In the present formulation the mixed stress depends on the nodal quantities too, since we use the average deformation gradient. To avoid a violation of the angular momentum balance, we have to calculate the mixed stress by

$$
\mathbf{P}_{n+1} = \mathbf{P}(\boldsymbol{\varphi}_{n+\frac{1}{2}}, \boldsymbol{\beta}_{0n+1}, \boldsymbol{\beta}_{n+1}) \tag{87}
$$

The other quantities in (85) can be calculate as expected, such that

$$
\begin{aligned}
\mathbf{F}_{n+1} &= \mathbf{F}(\boldsymbol{\varphi}_{n+1}, \boldsymbol{\gamma}_{0n+1}, \boldsymbol{\gamma}_{n+1}, \boldsymbol{\alpha}_{n+1}) & \mathbf{F}_{0n+\frac{1}{2}} &= \mathbf{F}_0(\boldsymbol{\varphi}_{n+\frac{1}{2}}) \\
\mathbf{F}_{n+\frac{1}{2}} &= \mathbf{F}(\boldsymbol{\varphi}_{n+\frac{1}{2}}, \boldsymbol{\gamma}_{0n+\frac{1}{2}}, \boldsymbol{\gamma}_{n+\frac{1}{2}}, \boldsymbol{\alpha}_{n+\frac{1}{2}}) & \widetilde{\mathbf{F}}_{n+1} &= \widetilde{\mathbf{F}}(\boldsymbol{\varphi}_{n+1}) \\
& & \widetilde{\mathbf{F}}_{n+\frac{1}{2}} &= \widetilde{\mathbf{F}}(\boldsymbol{\varphi}_{n+\frac{1}{2}})
\end{aligned}
\tag{88}
$$

Note that the temporal evaluation of the mixed quantities formulated in tensor notation are similar to (87)-(88).

Subsequently, we linearize the non-linear system of equations (85) in a straightforward way. After a standard assembly operation, a static condensation process is possible. This reduces the computational effort drastically. Further details can be found in [26, 27].

Finally we point out that the conservation properties from the time discrete case are inherited by the present finite element method.

5 Numerical examples

In the following section we would like to point out the performance of the proposed mixed-enhanced finite element along with the newly developed EMC scheme.

5.1 Static investigation

The first numerical example deals with a static benchmark problem for nonlinear finite elements to demonstrate the performance in the nearly incompressible limit. We consider a 3d version of the so called Cook's membrane as depicted in Figure 1. The deformed configuration of the problem is shown in Figure 2. The tapered panel is clamped on the left and loaded by a shear load $\mathbf{p}$ on the right side, where

$$\mathbf{p} = \begin{bmatrix} 0 & 0 & 3 \end{bmatrix}^{\mathrm{T}} \tag{89}$$

A St. Venant-Kirchhoff material model with material parameters $E = 100$ and $\nu = 0.4999$ has been used. To demonstrate the superior performance we compare the Kasper & Taylor element (H1/ME9) with a standard displacement element (H1) and an enhanced element (H1/E9) by Simo & Armero [26]. As illustrated in Figure 3, the convergence with mesh refinement of the three elements under consideration is investigated. We consider the displacement of the top corner versus the number of elements per side. While the H1 elements locks drastically, the H1/E9 and H1/ME9 are in a good agreement. Thus we conclude, that the present formulation performs well, even in the incompressible limit and in bending dominated problems.

Further investigations about objectivity and plasticity and an Eigenvalue analysis can be found in [14, 15].

5.2 Dynamic investigation

The second example can be considered as a classical benchmark for the transient simulation of solids. According to Simo & Tarnow [28] we verify the algorithmic consistency of the developed EMC scheme along with its superior performance as shown in section 5.1.

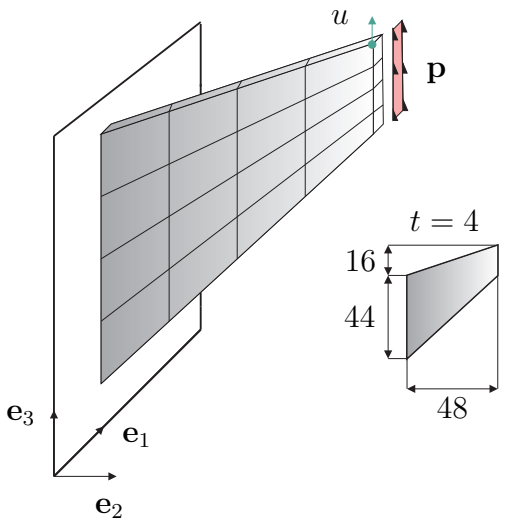


Figure 1: Cooks membrane problem. Geometry and boundary conditions

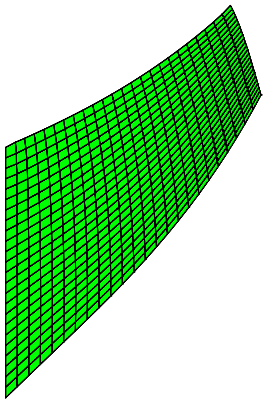


Figure 2: Cooks membrane problem. Deformed configuration

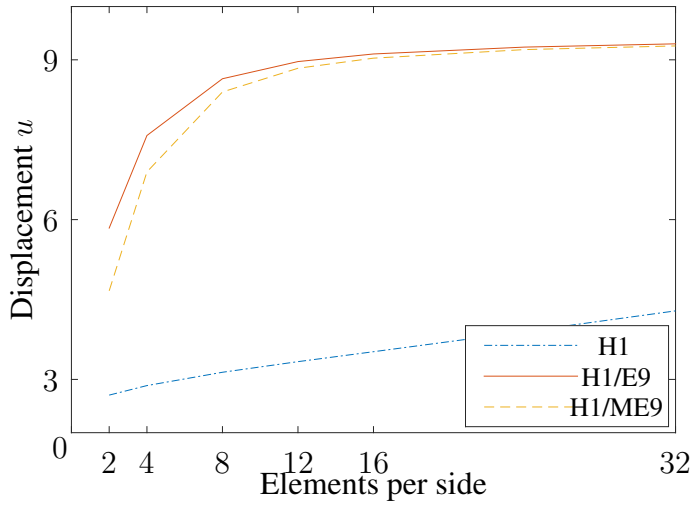


Figure 3: Study of convergence in static case. Quasi incompressible limit

Consider a flying L-shaped block with a constitutive behavior governed by a St. Venant-Kirchhoff material model. The material parameters are the Young Modulus $E = 2100$, the Poisson ratios $\nu = 0.4$ and the density $\rho_0 = 100$. Both, the initial geometry and the finite element mesh are illustrated in Figure 4. The L-shaped block has no Dirichlet boundary and pressure loads are acting on the block as shown in Figure 4. The time history of the external loads is illustrated in Figure 5. In this connection, the nodal

dead loads are given by

$$\begin{aligned} \mathbf{p}_1(t) &= f(t) \times \begin{bmatrix} -50 & -50 & -100 \end{bmatrix} \\ \mathbf{p}_2(t) &= f(t) \times \begin{bmatrix} 50 & 50 & 100 \end{bmatrix} \end{aligned} \tag{90}$$

The results of the EMC scheme are compared to those of the mid-point rule. All numerical simulation have been performed with a time-step size of $\Delta t = 0.25$. Note that after the loading phase the discrete system under consideration can be classified as autonomous Hamiltonian system with symmetry. Correspondingly, after $t \geq 1$, the total linear momentum, angular momentum and energy are conserved quantities. This is correctly reproduced by the EMC integrator (within the computational accuracy), whereas the mid-point rule exhibits numerical instabilities accompanied by an energy blow-up (see Figure 6). As depicted in Figure 7, the EMC scheme conserves the angular momentum exactly.

Eventually, the motion of the L-shape is illustrated in Figure 8 with a sequence of subsequent snapshots.

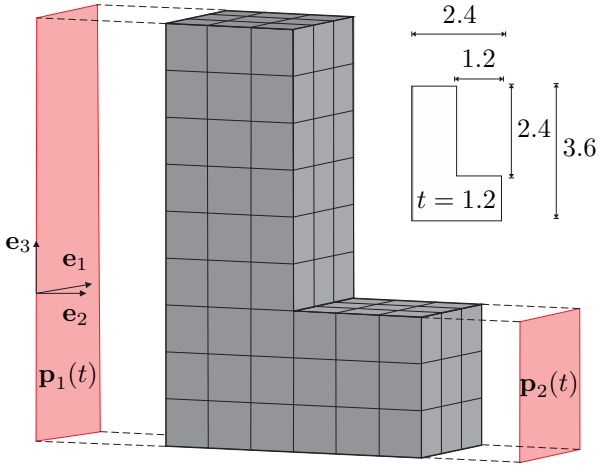


Figure 4: Flying L-Shape problem. Initial mesh configuration and dimensions

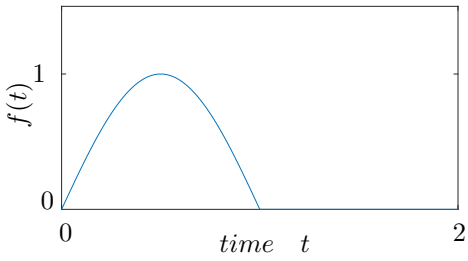


Figure 5: Time history of the external pressure

6 Conclusions

We developed a mixed variational formulation for nonlinear solid and structural dynamics which provides a natural framework for the EMC discretization in time. The underlying mixed variational equations were obtained as Euler-Lagrange equations of a mixed extension of Hamiltons principle. The Hu-Washizu type extension accommodates mixed finite elements based on independent interpolations of the displacements,

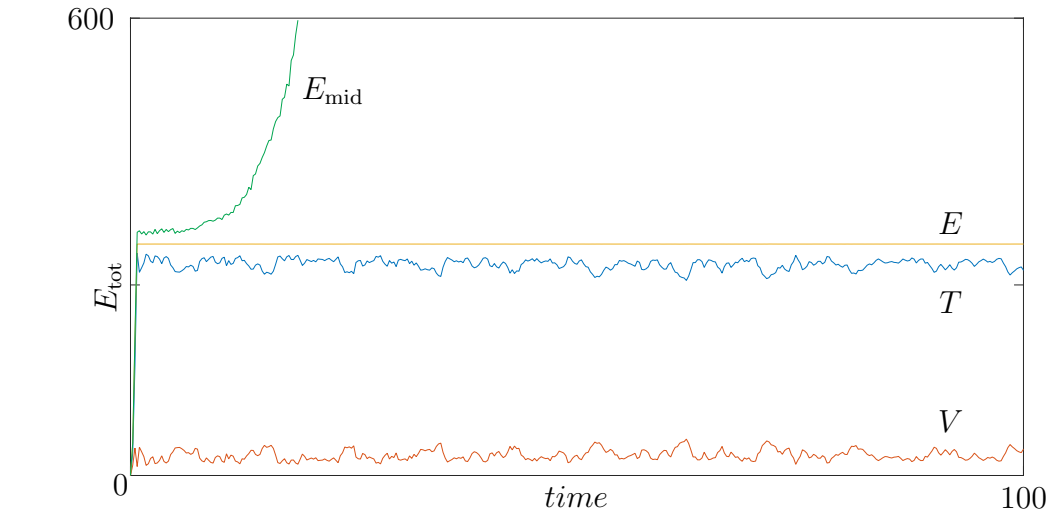


Figure 6: Time history of the total energy

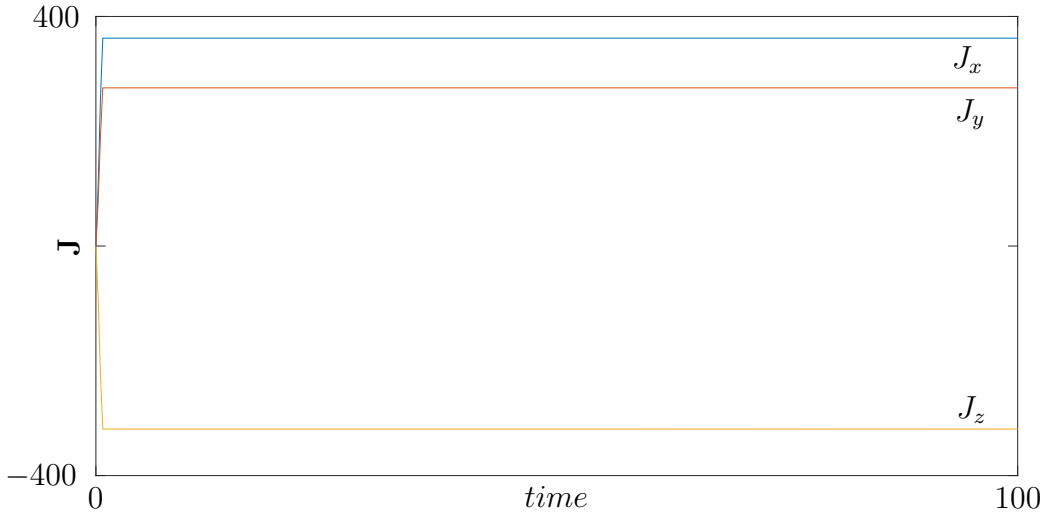


Figure 7: Time history of the angular momentum

the velocities, the deformation gradient, the right Cauchy-Green deformation tensor and both the first and the second Piola-Kirchhoff stresses.

We demonstrated, that the present framework is suitable even for more complicated formulations, like the mixed-enhanced element proposed by [15].

We show in Section 5.1, that the mixed-enhanced element performs well particularly

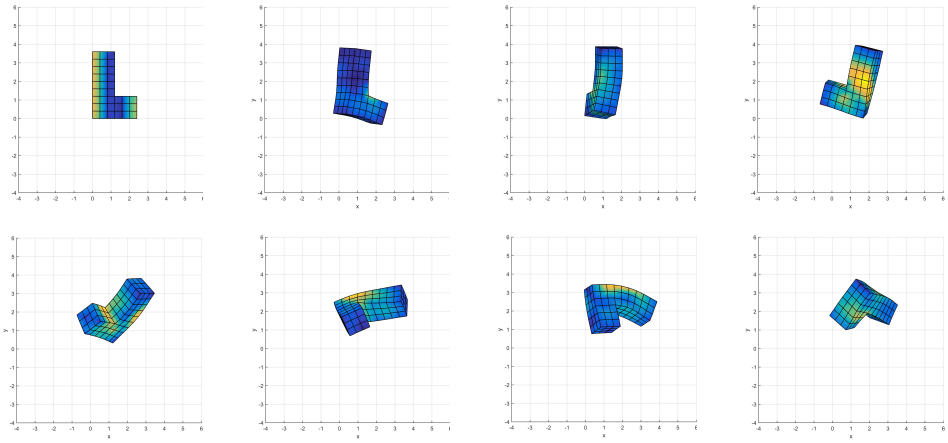


Figure 8: Snapshots of configurations at $t \in 0, 1, 2, 3, 4, 5, 6, 7, 8$

in the quasi-incompressible limit. The second numerical example, presented in section 5.2, shows that the EMC scheme leads to superior numerical stability properties compared to the common mid-point rule. While the conservation properties have been proven in the semi-discrete case and shown numerically, we still require a proof for the fully discrete system.

Accordingly, combining mixed finite element discretization in space with EMC discretization in time shows great promise for the design of numerical methods with superior coarse mesh accuracy in space and time. Thus, we obtain the possibility to use large time steps while still producing physically meaningful results, since the combination of EMC methods and mixed finite elements shows locking free response in non-linear dynamics.

Finally, we emphasize that many semidiscrete formulations of flexible bodies such as nonlinear continua, beams, and shells perfectly fit into the present framework, see [3]. Moreover, the present approach can be directly extended to flexible multibody dynamics as shown in [4].

Acknowledgement
Support for this research was provided by the Deutsche Forschungsgemeinschaft (DFG) under Grant BE 2285/9-1. This support is gratefully acknowledged.

A Appendix

A.1 Interpolation of mixed quantities

The maps for the three dimensional case are given by

$$\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\gamma}) = \begin{bmatrix} \xi_2\gamma_1 + \xi_3\gamma_2 + \xi_2\xi_3\gamma_3 & \xi_3\gamma_{10} & \xi_2\gamma_{14} \\ \xi_3\gamma_{11} & \xi_1\gamma_4 + \xi_3\gamma_5 + \xi_1\xi_3\gamma_6 & \xi_1\gamma_{12} \\ \xi_2\gamma_{15} & \xi_1\gamma_{13} & \xi_1\gamma_7 + \xi_2\gamma_8 + \xi_1\xi_2\gamma_9 \end{bmatrix}$$

$$\boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi},\boldsymbol{\alpha}) = \begin{bmatrix} \xi_1\alpha_1 + \xi_1\xi_2\alpha_2 + \xi_1\xi_3\alpha_3 & 0 & 0 \\ 0 & \xi_2\alpha_4 + \xi_2\xi_3\alpha_5 + \xi_1\xi_2\alpha_6 & 0 \\ 0 & 0 & \xi_3\gamma_7 + \xi_2\xi_3\gamma_8 + \xi_1\xi_3\gamma_9 \end{bmatrix}$$

$$\boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\beta}) = \begin{bmatrix} \xi_2\beta_1 + \xi_3\beta_2 + \xi_2\xi_3\beta_3 & \xi_3\beta_{10} & \xi_2\beta_{14} \\ \xi_3\beta_{11} & \xi_1\beta_4 + \xi_3\beta_5 + \xi_1\xi_3\beta_6 & \xi_1\beta_{12} \\ \xi_2\beta_{15} & \xi_1\beta_{13} & \xi_1\beta_7 + \xi_2\beta_8 + \xi_1\xi_2\beta_9 \end{bmatrix} \tag{91}$$

or equivalent, the interpolations in tensor form

$$\begin{aligned} \boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\gamma}) &\xrightarrow{i,j\to i} \mathbf{E}_1(\boldsymbol{\xi})\,\boldsymbol{\gamma} \\ \boldsymbol{\mathcal{E}}_2(\boldsymbol{\xi},\boldsymbol{\alpha}) &\xrightarrow{i,j\to i} \mathbf{E}_1(\boldsymbol{\xi})\,\boldsymbol{\alpha} \\ \boldsymbol{\mathcal{E}}_1(\boldsymbol{\xi},\boldsymbol{\beta}) &\xrightarrow{i,j\to i} \mathbf{E}_1(\boldsymbol{\xi})\,\boldsymbol{\beta} \end{aligned} \tag{92}$$

$$\mathbf{E}_1(\boldsymbol{\xi}) = \begin{bmatrix} \eta & \zeta & \eta\zeta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \xi & \zeta & \xi\zeta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \xi & \eta & \xi\eta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta \end{bmatrix}$$

$$\mathbf{E}_2(\boldsymbol{\xi}) = \begin{bmatrix} \xi & \xi\eta & \xi\zeta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta & \eta\zeta & \xi\eta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \zeta & \eta\zeta & \xi\zeta \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{93}$$

REFERENCES

[1] F. Armero. Assumed strain finite element methods for conserving temporal integrations in non-linear solid dynamics. *Int. J. Numer. Meth. Engng*, 74(12):1795–1847, 2008.

[2] F. Armero and I. Romero. Energy-dissipative momentum-conserving time-stepping algorithms for the dynamics of nonlinear Cosserat rods. *Computational Mechanics*, 31:3–26, 2003.

[3] P. Betsch and A. Janz. An energy-momentum consistent method for transient simulations with mixed finite elements developed in the framework of geometrically exact shells. Accepted for publication in Int. J. Numer. Meth. Engng.

[4] P. Betsch and N. Sänger. On the use of geometrically exact shells in a conserving framework for flexible multibody dynamics. *Comput. Methods Appl. Mech. Engrg.*, 198:1609–1630, 2009.

[5] P. Betsch and P. Steinmann. A DAE approach to flexible multibody dynamics. *Multibody System Dynamics*, 8:367–391, 2002.

[6] P. Betsch and P. Steinmann. Constrained dynamics of geometrically exact beams. *Computational Mechanics*, 31:49–59, 2003.

[7] C.L. Bottasso, O.A. Bauchau, and J.-Y. Choi. An energy decaying scheme for nonlinear dynamics of shells. *Comput. Methods Appl. Mech. Engrg.*, 191(27-28):3099–3121, 2002.

[8] B. Brank, J. Korelc, and A. Ibrahimbegović. Dynamics and time-stepping schemes for elastic shells undergoing finite rotations. *Computers and Structures*, 81(12):1193–1210, 2003.

[9] U. Galvanetto and M.A. Crisfield. An energy-conserving co-rotational procedure for the dynamics of planar beam structures. *Int. J. Numer. Meth. Engng*, 39:2265–2282, 1996.

[10] O. Gonzalez. Time integration and discrete Hamiltonian systems. *J. Nonlinear Sci.*, 6:449–467, 1996.

[11] O. Gonzalez. Exact energy and momentum conserving algorithms for general models in nonlinear elasticity. *Comput. Methods Appl. Mech. Engrg.*, 190(13-14):1763–1783, 2000.

[12] O. Gonzalez and J.C. Simo. On the stability of symplectic and energy-momentum algorithms for non-linear Hamiltonian systems with symmetry. *Comput. Methods Appl. Mech. Engrg.*, 134:197–222, 1996.

[13] C. Hesch and P. Betsch. Energy-momentum consistent algorithms for dynamic thermomechanical problems – application to mortar domain decomposition problems. *Int. J. Numer. Meth. Engng*, 86(11):1277–1302, 2011.

[14] E.P. Kasper and R.L. Taylor. A mixed-enhanced strain method. Part I: Geometrically linear problems. *Computers and Structures*, 75(3):237–250, 2000.

[15] E.P. Kasper and R.L. Taylor. A mixed-enhanced strain method. Part II: Geometrically nonlinear problems. *Computers and Structures*, 75(3):251–260, 2000.

[16] D. Kuhl and E. Ramm. Time integration in the context of energy control and locking free finite elements. *Archives of Computational Methods in Engineering. State of the Art Reviews*, 7(3):299–332, 2000.

[17] T.A. Laursen and X.N. Meng. A new solution procedure for application of energy-conserving algorithms to general constitutive models in nonlinear elastodynamics. *Comput. Methods Appl. Mech. Engrg.*, 190:6309–6322, 2001.

[18] E.V. Lens and A. Cardona. A nonlinear beam element formulation in the framework of an energy preserving time integration scheme for constrained multibody systems dynamics. *Computers and Structures*, 86(1-2):47–63, 2008.

[19] S. Leyendecker, P. Betsch, and P. Steinmann. Objective energy-momentum conserving integration for the constrained dynamics of geometrically exact beams. *Comput. Methods Appl. Mech. Engrg.*, 195:2313–2333, 2006.

[20] C. Miehe and J. Schröder. Energy and momentum conserving elastodynamics of a non-linear brick-type mixed finite shell element. *Int. J. Numer. Meth. Engng*, 50:1801–1823, 2001.

[21] L.A. Pars. *A treatise on analytical dynamics*. Heinemann, 1965.

[22] I. Romero. An analysis of the stress formula for energy-momentum methods in nonlinear elastodynamics. *Computational Mechanics*, 50:603–610, 2012.

[23] I. Romero and F. Armero. Numerical integration of the stiff dynamics of geometrically exact shells: an energy-dissipative momentum-conserving scheme. *Int. J. Numer. Meth. Engng*, 54:1043–1086, 2002.

[24] I. Romero and F. Armero. An objective finite element approximation of the kinematics of geometrically exact rods and its use in the formulation of an energy-momentum conserving scheme in dynamics. *Int. J. Numer. Meth. Engng*, 54:1683–1716, 2002.

[25] C. Sansour, T.L. Nguyen, and M. Hjiaj. An energy-momentum method for in-plane geometrically exact Euler-Bernoulli beam dynamics. *Int. J. Numer. Meth. Engng*, 102:99–134, 2015.

[26] J.C. Simo and F. Armero. Geometrically nonlinear enhanced strain mixed methods and the method of incompatible modes. *Int. J. Numer. Meth. Engng*, 33:1413–1449, 1992.

[27] J.C. Simo, F. Armero, and R.L. Taylor. Improved versions of assumed enhanced strain tri-linear elements for 3d finite deformation problems. *Comput. Methods Appl. Mech. Engrg.*, 110:359–386, 1993.

[28] J.C. Simo and N. Tarnow. The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *Z. angew. Math. Phys. (ZAMP)*, 43:757–792, 1992.

[29] J.C. Simo and N. Tarnow. A new energy and momentum conserving algorithm for the nonlinear dynamics of shells. *Int. J. Num. Meth. Engng*, 37:2527–2549, 1994.

[30] J.C. Simo, N. Tarnow, and M. Doblare. Non-linear dynamics of three-dimensional rods: Exact energy and momentum conserving algorithms. *Int. J. Numer. Meth. Engng*, 38:1431–1473, 1995.

[31] K. Washizu. *Variational Methods in Elasticity & Plasticity*. Pergamon Press, 3rd edition, 1982.

[32] H.G. Zhong and M.A. Crisfield. An energy-conserving co-rotational procedure for the dynamics of shell structures. *Eng. Comp.*, 15:552–576, 1998.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ENERGY AND MOMENTUM CONSERVING VARIATIONAL BASED TIME INTEGRATION OF ANISOTROPIC HYPERELASTIC CONTINUA

Michael Groß[1], Rajesh Ramesh[2] and Julian Dietzsch[3]

Technische Universität Chemnitz, Professorship of applied mechanics and dynamics
Reichenhainer Straße 70, D-09126 Chemnitz
[1] michael.gross@mb.tu-chemnitz.de, [2] rajesh.ramesh@mb.tu-chemnitz.de, [3] tmd@mb.tu-chemnitz.de

Keywords: Time stepping schemes, finite elasticity, anisotropy, fiber-reinforced continua.

Abstract. *For many years, the importance of fiber-reinforced polymers is steadily increasing in mechanical engineering. According to the high strength in fiber direction, these composites replace more and more traditional homogeneous materials, especially in lightweight structures. Fiber-reinforced material parts are often manufactured from carbon fibers as pure attachment parts, or from steel for transmitting forces. Whereas attachment parts are mostly subjected to small deformations, force transmission parts usually suffer large deformations in at least one direction. For the latter, a geometrically non-linear formulation of these anisotropic continua is indispensable [1]. A familar example is a rotor blade, in which the fibers possess the function of stabilizing the structure in order to counteract large centrifugal forces. For long-run numerical analyses of rotor blade motions, we have to apply numerically stable and robust time integration schemes for anisotropic continua.*

This paper is an extension of Reference [2], which is in turn an extension of Reference [3] to a special anisotropic material class, namely a transversely isotropic hyperelastic material based on the wellknown concept of structural tensors. In Reference [3], higher-order accurate time-stepping schemes are developed systematically with the focus on numerical stability and robustness in the presence of stiffness combined with large rotations for computing large motions. In the former work, these advantages over conventional time stepping schemes are combined with highly non-linear anisotropic material formulated with polyconvex free energy density functions [4]. The corresponding time integrators preserve all conservation laws of a free motion of a hyperelastic continuum, which means the total linear and the total angular momentum conservation law as well as the total energy conservation law. Both are numerically advantageous, because it guarantees that the discrete configuration vector is embedded in the physically consistent solution space. In order to guarantee the preservation of the total energy, the transient approximation of the anisotropic stress tensor is superimposed with an algorithmic stress field based on an assumed 'strain' field.

The presented numerical examples show the behaviour of the non-linear anisotropic material in Reference [4] under static and transient loads, their conservation laws and the higher-order accuracy of the variational based time approximation.

1 INTRODUCTION

We begin by summarizing the kinematical aspects of the considered transversely isotropic continuum. In Fig. 1 on the right-hand side, we show the reference configuration $\mathscr{B}_0$ of the considered fiber-reinforced continuum body $\mathscr{B}$. The configuration $\mathscr{B}_0 = \mathscr{B}_0^M \cup \mathscr{B}_0^F$ is defined

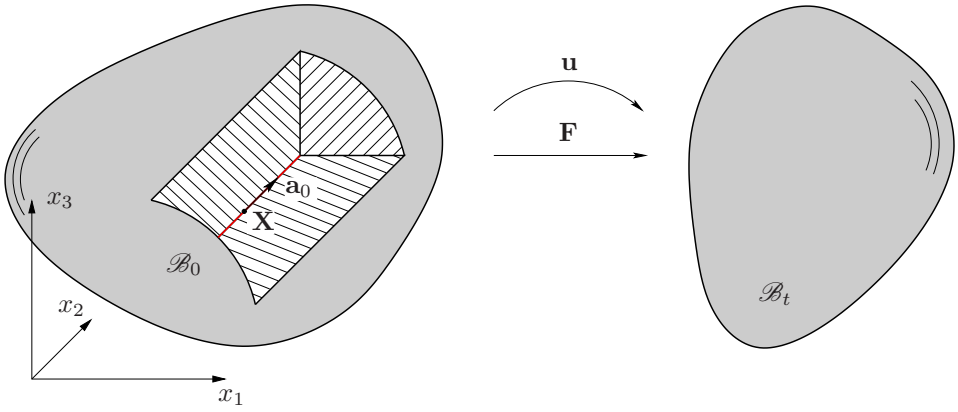


Figure 1: Reference and current configuration of a transversely isotropic continuum.

as the *homogenized* union of the set $\mathscr{B}_0^M$ for the matrix and the set $\mathscr{B}_0^F$ for the fibers. The imaginary fiber at *any* point $\mathbf{X} \in \mathscr{B}_0$ is directed along the *normalized vector* $\mathbf{a}_0$. Since we assume that both subsets are perfectly connected, the corresponding stretched vector $\mathbf{a}$ in the deformed configuration $\mathscr{B}_t$ is given by

$$\mathbf{a} = \mathbf{F}\,\mathbf{a}_0 \tag{1}$$

where

$$\mathbf{F} := \nabla\mathbf{u} + \mathbf{I} \tag{2}$$

denotes the deformation gradient of $\mathscr{B}_0$ and $\mathbf{u}$ the displacement vector field. The tensor $\mathbf{I}$ designates the second-order unit tensor. The symbol ∇ denotes the partial derivative with respect to the material point $\mathbf{X} \in \mathscr{B}_0$. The deformation gradient $\mathbf{F}_F$ of the fiber continuum $\mathscr{B}_0^F$ then takes the form

$$\mathbf{F}_F := \mathbf{a} \otimes \mathbf{a}_0 = \mathbf{F}\,\mathbf{a}_0 \otimes \mathbf{a}_0 = \mathbf{F}\,\mathbf{A}_0 \tag{3}$$

where

$$\mathbf{A}_0 := \mathbf{a}_0 \otimes \mathbf{a}_0 \tag{4}$$

designates the structural tensor of the fiber reference configuration $\mathscr{B}_0^F$. The corresponding right CAUCHY-GREEN tensor $\mathbf{C}_F$ then reads

$$\mathbf{C}_F := \mathbf{F}_F^T\,\mathbf{F}_F = [\mathbf{F}\,\mathbf{a}_0 \otimes \mathbf{a}_0]^T\,[\mathbf{F}\,\mathbf{a}_0 \otimes \mathbf{a}_0] = [\mathbf{a}_0 \otimes \mathbf{a}_0]\,\mathbf{C}\,[\mathbf{a}_0 \otimes \mathbf{a}_0] = \mathbf{A}_0\,\mathbf{C}\,\mathbf{A}_0 = [\mathbf{C} : \mathbf{A}_0]\,\mathbf{A}_0 \tag{5}$$

where $\mathbf{C} := \mathbf{F}^T\,\mathbf{F}$ denotes the right CAUCHY-GREEN tensor of $\mathscr{B}_0$. Based on these deformation measures, we consider the strain energy function W of the considered transversely isotropic elastic continuum on the one hand (i) as the *unpartitioned* function $W(\mathbf{C}; \mathbf{A}_0, \boldsymbol{\kappa}_0)$, where the semicolon in the argument separates the *parameter* $\mathbf{A}_0$ and $\boldsymbol{\kappa}_0$, *acting at any* $\mathbf{X} \in \mathscr{B}_0$, from the variable $\mathbf{C}$, and on the other hand (ii) as the *partitioned* function

$$W(\mathbf{C}; \mathbf{A}_0, \boldsymbol{\kappa}_0) = W_M(\mathbf{C}; \boldsymbol{\kappa}_{0_M}) + W_F(\mathbf{C}_F; \boldsymbol{\kappa}_{0_F}) \tag{6}$$

The parameter $\boldsymbol{\kappa}_0$, $\boldsymbol{\kappa}_{0_M}$ and $\boldsymbol{\kappa}_{0_F}$ are vectors including material constants with respect to $\mathscr{B}_0$. The second PIOLA-KIRCHHOFF stress tensor $\mathbf{S}$ corresponding to Eq. (6) is given by

$$\begin{aligned}
\mathbf{S} \equiv 2\,\frac{\partial W(\mathbf{C};\mathbf{A}_0,\boldsymbol{\kappa}_0)}{\partial \mathbf{C}} &= \mathbf{S}_M + \mathbf{S}_F && (7)\\
&= 2\,\mathrm{D}W_M(\mathbf{C};\boldsymbol{\kappa}_{0_M}) + 2\,\mathrm{D}W_F(\mathbf{C}_F;\boldsymbol{\kappa}_{0_M}) : \frac{\partial \mathbf{C}_F}{\partial \mathbf{C}} \\
&= 2\,\mathrm{D}W_M(\mathbf{C};\boldsymbol{\kappa}_{0_M}) + 2\,\mathbf{A}_0\,\mathrm{D}W_F(\mathbf{C}_F;\boldsymbol{\kappa}_{0_M})\,\mathbf{A}_0 && (8)\\
\mathbf{S} &= 2\,\mathrm{D}W_M(\mathbf{C};\boldsymbol{\kappa}_{0_M}) + 2\,\left[\mathrm{D}W_F(\mathbf{C}_F;\boldsymbol{\kappa}_{0_M}) : \mathbf{A}_0\right]\mathbf{A}_0 && (9)
\end{aligned}$$

The notation $\mathrm{D}(\bullet)$ denotes the FRÉCHET derivative of a volume density with respect to its argument. The strain energy functions W of $\mathscr{B}_0$, W_M of $\mathscr{B}_0^M$ or W_F of $\mathscr{B}_0^F$, respectively, directly depends on the *invariants* of the corresponding right CAUCHY-GREEN tensors. We assume (i) the unpartitioned case

$$W(\mathbf{C};\mathbf{A}_0,\boldsymbol{\kappa}_0) = \hat{W}(I_1, I_2, I_3, I_4;\boldsymbol{\kappa}_0) \tag{10}$$

and (ii) the partitioned case

$$W_M = \hat{W}_M(I_1, I_2, I_3;\boldsymbol{\kappa}_{0_M}) \qquad\qquad W_F = \hat{W}_F(I_4;\boldsymbol{\kappa}_{0_F}) \tag{11}$$

where

$$I_1 := \mathbf{C} : \mathbf{I} \qquad\qquad I_2 := \tfrac{1}{2}\left[(I_1)^2 - \mathbf{C}^2 : \mathbf{I}\right] \qquad\qquad I_3 := \det \mathbf{C} \tag{12}$$

denotes the tensor invariants of the right CAUCHY-GREEN tensors $\mathbf{C}$, and

$$I_4 \equiv \mathbf{C}_F : \mathbf{A}_0 = \mathbf{A}_0\,\mathbf{C}\,\mathbf{A}_0 : \mathbf{A}_0 = [\mathbf{a}_0 \cdot \mathbf{C}\,\mathbf{a}_0]\,\mathbf{a}_0 \cdot \mathbf{a}_0 = \mathbf{a}_0 \cdot \mathbf{C}\,\mathbf{a}_0 = \mathbf{C} : \mathbf{A}_0 = \mathbf{a} \cdot \mathbf{a} =: C_F \tag{13}$$

the squared fiber stretch $C_F \equiv \lambda_F^2$. Using the fourth invariant $I_4 =: C_F$, the right CAUCHY-GREEN tensor $\mathbf{C}_F$ and the second PIOLA-KIRCHHOFF stress tensor $\mathbf{S}_F$ of the partitioned strain energy function, respectively, can be simply written as

$$\mathbf{C}_F = C_F\,\mathbf{A}_0 \qquad\qquad \mathbf{S}_F = 2\,\left[\mathrm{D}\hat{W}_F(C_F;\boldsymbol{\kappa}_{0_F})\,\frac{\partial C_F}{\partial \mathbf{C}_F} : \mathbf{A}_0\right]\mathbf{A}_0 = 2\,\mathrm{D}\hat{W}_F(C_F;\boldsymbol{\kappa}_{0_F})\,\mathbf{A}_0 \tag{14}$$

Hence, the directions of the fiber deformation tensor $\mathbf{C}_F$ and fiber stress tensor $\mathbf{S}_F$ are uniquely prescribed by the structure tensor $\mathbf{A}_0$, as expected.

2 EULER-LAGRANGE EQUATIONS

With regard to the numerical time integration, we now introduce *variationally consistent*

- temporally continuous assumed 'strains' $\tilde{\mathbf{C}}$ and $\tilde{C}_F$, as well as
- temporally discontinuous superimposed stresses $\tilde{\mathbf{S}}$ and $\tilde{S}_F$, respectively.

The former are necessary for an exact analytical time integration of *approximated* strain energy functions [6], and the later for their exact numerical time integration [8]. Hence, the superimposed stresses $\tilde{\mathbf{S}}$ and $\tilde{S}_F$ are responsible for the energy consistency of the discrete EULER-LAGRANGE equations, but they have to vanish identically for guaranteeing energy consistency of the continuous EULER-LAGRANGE equations. We consider the strain energy function

1. $W(\mathbf{C};\mathbf{A}_0,\boldsymbol{\kappa}_0)$ on $\mathscr{B}_0$ (unpartitioned strain energy), or
2. $W_M(\mathbf{C};\boldsymbol{\kappa}_{0_M})$ and $W_F(\mathbf{C}_F;\boldsymbol{\kappa}_{0_F})$ separately (partitioned strain energy).

We derive the continuous equations of motion by using a mixed *principle of virtual power* or *principle of Jourdain*, respectively, a differential variational principle [9]. The modivation for applying this principle from the outset is to satisfy the *total energy balance* in both the continuous as well as the discrete setting. Denoting by a superimposed dot the partial time derivative, in the unpartitioned case, this balance takes the form

$$\dot{\mathcal{T}}(\dot{\mathbf{u}},\dot{\mathbf{v}},\dot{\mathbf{p}};\rho_0)+\dot{\Pi}^{\text{int}}(\dot{\mathbf{u}},\dot{\tilde{\mathbf{C}}},\mathbf{S};\mathbf{A}_0,\boldsymbol{\kappa}_0,\tilde{\mathbf{S}})=\int_{\mathscr{B}_0}\rho_0\mathbf{b}\cdot\dot{\mathbf{u}}\,\mathrm{d}V+\int_{\partial_t\mathscr{B}_0}\mathbf{t}\cdot\dot{\mathbf{u}}\,\mathrm{d}A+\int_{\partial_u\mathscr{B}_0}\mathbf{h}\cdot\left(\dot{\mathbf{u}}-\dot{\bar{\mathbf{u}}}\right)\mathrm{d}A \tag{15}$$

with the kinetic power

$$\dot{\mathcal{T}}(\dot{\mathbf{u}},\dot{\mathbf{v}},\dot{\mathbf{p}};\rho_0):=\int_{\mathscr{B}_0}[\rho_0\mathbf{v}-\mathbf{p}]\cdot\dot{\mathbf{v}}\,\mathrm{d}V-\int_{\mathscr{B}_0}\dot{\mathbf{p}}\cdot[\mathbf{v}-\dot{\mathbf{u}}]\,\mathrm{d}V+\int_{\mathscr{B}_0}\mathbf{p}\cdot\ddot{\mathbf{u}}\,\mathrm{d}V \tag{16}$$

where

$$\mathbf{p}=\rho_0\mathbf{v}\qquad\text{and}\qquad\mathbf{v}=\dot{\mathbf{u}} \tag{17}$$

denotes the linear momentum vector and the material velocity vector, respectively. The scalar ρ_0 denotes the mass density field in $\mathscr{B}_0$. The stress power $\dot{\Pi}^{\text{int}}$ is written in dependence on the second PIOLA-KIRCHHOFF stress tensor $\mathbf{S}$, the superimposed stress tensor $\tilde{\mathbf{S}}$ and the assumed 'strain' tensor $\tilde{\mathbf{C}}$ as

$$\begin{aligned}\dot{\Pi}^{\text{int}}(\dot{\mathbf{u}},\dot{\tilde{\mathbf{C}}},\mathbf{S};\mathbf{A}_0,\boldsymbol{\kappa}_0,\tilde{\mathbf{S}}):=\quad&\frac{1}{2}\int_{\mathscr{B}_0}\left[2\,\mathrm{D}W(\tilde{\mathbf{C}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}-\mathbf{S}\right]:\dot{\tilde{\mathbf{C}}}\,\mathrm{d}V\\-\quad&\frac{1}{2}\int_{\mathscr{B}_0}\dot{\mathbf{S}}:\left[\tilde{\mathbf{C}}-\mathbf{C}(\mathbf{u})\right]\mathrm{d}V+\frac{1}{2}\int_{\mathscr{B}_0}\mathbf{S}:\dot{\mathbf{C}}(\dot{\mathbf{u}})\,\mathrm{d}V\\=\quad&\int_{\mathscr{B}_0}\dot{W}\,\mathrm{d}V\end{aligned} \tag{18}$$

where

$$\mathbf{S}=2\,\mathrm{D}W(\tilde{\mathbf{C}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}\qquad\qquad\tilde{\mathbf{C}}=\mathbf{C}(\mathbf{u}):=(\nabla\mathbf{u}+\mathbf{I})^T(\nabla\mathbf{u}+\mathbf{I}) \tag{19}$$

The superimposed stress tensor $\tilde{\mathbf{S}}=\mathbf{O}$, with the zero tensor $\mathbf{O}$, has to vanish for energy consistency. On the right-hand side of Eq. (15), there is the external power depending on the body force vector $\mathbf{b}$ per unit mass on $\mathscr{B}_0$, the traction force vector $\mathbf{t}$ per unit area on the NEUMANN boundary $\partial_t\mathscr{B}_0$ and the LAGRANGE multiplier vector $\mathbf{h}$ enforcing the constraint

$$\mathbf{u}-\bar{\mathbf{u}}=\mathbf{0}\quad\text{on}\quad\partial_u\mathscr{B}_0 \tag{20}$$

of a prescribed displacement $\bar{\mathbf{u}}$ on the DIRICHLET boundary $\partial_u\mathscr{B}_0$. Both boundary sets satisfy the conditions $\partial\mathscr{B}_0=\partial_t\mathscr{B}_0\cup\partial_u\mathscr{B}_0$ and $\partial_t\mathscr{B}_0\cap\partial_u\mathscr{B}_0=\emptyset$, where $\partial\mathscr{B}_0$ designates the boundary of the reference configuration.

2.1 The unpartitioned strain energy function

In the unpartitioned case, we introduce the assumed 'strain' field $\tilde{\mathbf{C}}$ and the superimposed stress field $\tilde{\mathbf{S}}$ for the entire reference configuration variationally consistent by considering the virtual power principle

$$\begin{aligned}&\delta_*\dot{\mathcal{H}}(\dot{\mathbf{u}},\dot{\mathbf{v}},\dot{\mathbf{p}},\dot{\tilde{\mathbf{C}}},\mathbf{S};\rho_0,\mathbf{A}_0,\boldsymbol{\kappa}_0,\mathbf{b},\mathbf{t},\dot{\tilde{\mathbf{u}}},\tilde{\mathbf{S}}) :=\\&\delta_*\dot{\mathcal{T}}(\dot{\mathbf{u}},\dot{\mathbf{v}},\dot{\mathbf{p}};\rho_0)+\delta_*\dot{\Pi}^{\mathrm{int}}(\dot{\mathbf{u}},\dot{\tilde{\mathbf{C}}},\mathbf{S};\mathbf{A}_0,\boldsymbol{\kappa}_0,\tilde{\mathbf{S}})+\delta_*\dot{\Pi}^{\mathrm{ext}}(\dot{\mathbf{u}};\rho_0,\mathbf{b},\mathbf{t},\dot{\tilde{\mathbf{u}}})=0\end{aligned}\tag{21}$$

with the virtual kinetic power

$$\delta_*\dot{\mathcal{T}}(\dot{\mathbf{u}},\dot{\mathbf{v}},\dot{\mathbf{p}};\rho_0) := \int_{\mathscr{B}_0}[\rho_0\mathbf{v}-\mathbf{p}]\cdot\delta_*\dot{\mathbf{v}}\,\mathrm{d}V-\int_{\mathscr{B}_0}\delta_*\dot{\mathbf{p}}\cdot[\mathbf{v}-\dot{\mathbf{u}}]\,\mathrm{d}V+\int_{\mathscr{B}_0}\dot{\mathbf{p}}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}V\tag{22}$$

the virtual external power

$$\delta_*\dot{\Pi}^{\mathrm{ext}}(\dot{\mathbf{u}};\rho_0,\mathbf{b},\mathbf{t},\dot{\tilde{\mathbf{u}}}) := -\int_{\mathscr{B}_0}\rho_0\,\mathbf{b}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}V-\int_{\partial_t\mathscr{B}_0}\mathbf{t}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}A-\int_{\partial_u\mathscr{B}_0}\mathbf{h}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}A\tag{23}$$

and the virtual internal power

$$\begin{aligned}\delta_*\dot{\Pi}^{\mathrm{int}}(\dot{\mathbf{u}},\dot{\tilde{\mathbf{C}}},\mathbf{S};\mathbf{A}_0,\boldsymbol{\kappa}_0,\tilde{\mathbf{S}}) := \quad&\frac{1}{2}\int_{\mathscr{B}_0}\Big[2\mathrm{D}W(\tilde{\mathbf{C}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}-\mathbf{S}\Big]:\delta_*\dot{\tilde{\mathbf{C}}}\,\mathrm{d}V\\&-\frac{1}{2}\int_{\mathscr{B}_0}\delta_*\mathbf{S}:\Big[\dot{\tilde{\mathbf{C}}}-\dot{\mathbf{C}}(\dot{\mathbf{u}})\Big]\,\mathrm{d}V+\frac{1}{2}\int_{\mathscr{B}_0}\mathbf{S}:\delta_*\dot{\mathbf{C}}(\dot{\mathbf{u}})\,\mathrm{d}V\end{aligned}\tag{24}$$

The symbol δ_* denotes the variation with respect to the variables (not the parameter behind the semicolon) in the function argument. Integration by parts in the last term of Eq. (24) furnishes

$$\frac{1}{2}\int_{\mathscr{B}_0}\mathbf{S}:\delta_*\dot{\mathbf{C}}(\dot{\mathbf{u}})\,\mathrm{d}V=\int_{\mathscr{B}_0}\mathbf{FS}:\nabla(\delta_*\dot{\mathbf{u}})\,\mathrm{d}V=\int_{\partial_t\mathscr{B}_0}\mathbf{FSN}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}A-\int_{\mathscr{B}_0}\mathrm{DIV}[\mathbf{FS}]\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}V\tag{25}$$

The vector $\mathbf{N}$ denotes the normal field on the NEUMANN boundary $\partial_t\mathscr{B}_0$, and $\mathrm{DIV}[\bullet]$ the divergence operator with respect to $\mathbf{X}\in\mathscr{B}_0$. Rearranging termes in Eq. (21) according to the variations $\delta_*\dot{\mathbf{p}}$, $\delta_*\dot{\mathbf{v}}$, $\delta_*\mathbf{S}$, $\delta_*\dot{\tilde{\mathbf{C}}}$ and $\delta_*\dot{\mathbf{u}}$, we obtain the variational form

$$\begin{aligned}0 \;=\;&\int_{\mathscr{B}_0}[\rho_0\mathbf{v}-\mathbf{p}]\cdot\delta_*\dot{\mathbf{v}}\,\mathrm{d}V-\int_{\mathscr{B}_0}\delta_*\dot{\mathbf{p}}\cdot[\mathbf{v}-\dot{\mathbf{u}}]\,\mathrm{d}V-\int_{\mathscr{B}_0}[\mathrm{DIV}[\mathbf{FS}]+\rho_0\mathbf{b}-\dot{\mathbf{p}}]\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}V\\&-\frac{1}{2}\int_{\mathscr{B}_0}\Big[\mathbf{S}-2\,\mathrm{D}W(\tilde{\mathbf{C}};\mathbf{A}_0,\boldsymbol{\kappa}_0)-\tilde{\mathbf{S}}\Big]:\delta_*\dot{\tilde{\mathbf{C}}}\,\mathrm{d}V-\frac{1}{2}\int_{\mathscr{B}_0}\delta_*\mathbf{S}:\Big[\dot{\tilde{\mathbf{C}}}-\dot{\mathbf{C}}(\dot{\mathbf{u}})\Big]\,\mathrm{d}V\\&-\int_{\partial_t\mathscr{B}_0}[\mathbf{t}-\mathbf{FSN}]\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}A-\int_{\partial_u\mathscr{B}_0}\mathbf{h}\cdot\delta_*\dot{\mathbf{u}}\,\mathrm{d}A\end{aligned}\tag{26}$$

Owing to the fundamental theorem of variational calculus, the corresponding continuous EULER-LAGRANGE equations read

$$\mathbf{v}=\dot{\mathbf{u}}\quad\text{with}\quad\mathbf{u}(t_0)=\mathbf{u}_0\tag{27}$$

$$\rho_0\mathbf{v}=\mathbf{p}\qquad\forall t\geqslant t_0\tag{28}$$

$$\mathrm{DIV}[\mathbf{FS}]+\rho_0\mathbf{b}=\dot{\mathbf{p}}\quad\text{with}\quad\mathbf{p}(t_0)=\mathbf{p}_0\equiv\rho_0\mathbf{v}_0\tag{29}$$

$$\dot{\mathbf{C}}(\dot{\mathbf{u}})=\dot{\tilde{\mathbf{C}}}\quad\text{with}\quad\tilde{\mathbf{C}}(t_0)=\mathbf{C}(\mathbf{u}_0)\equiv(\nabla\mathbf{u}_0+\mathbf{I})^T(\nabla\mathbf{u}_0+\mathbf{I})\tag{30}$$

$$2\,\mathrm{D}W(\tilde{\mathbf{C}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}=\mathbf{S}\qquad\forall t\geqslant t_0\tag{31}$$

with the corresponding inital conditions in $\mathscr{B}_0$ as well as the boundary conditions

$$\mathbf{FSN} = \mathbf{t} \qquad \forall t \geqslant t_0 \quad \text{on } \partial_t \mathscr{B}_0 \tag{32}$$

$$\delta_* \dot{\mathbf{u}} = \mathbf{0} \Longleftrightarrow \dot{\mathbf{u}} = \dot{\bar{\mathbf{u}}} \text{ with } \mathbf{u}(t_0) = \bar{\mathbf{u}}(t_0) \quad \text{on } \partial_u \mathscr{B}_0 \tag{33}$$

Consequently, the PIOLA-KIRCHHOFF stress tensor field $\mathbf{S}$ is temporally discontinuous, but the displacement vector field $\mathbf{u}$, the material velocity field $\mathbf{v}$, the linear momentum field $\mathbf{p}$ *as well as* the assumed 'strain' field $\tilde{\mathbf{C}}$ are temporally continuous. The superimposed stress tensor $\tilde{\mathbf{S}}$ has to vanish in these EULER-LAGRANGE equations for satisfying the total energy balance in Eq. (15).

2.2 The partitioned strain energy function

In the partitioned formulation, we introduce the assumed 'strain' tensor $\tilde{\mathbf{C}}$ and the superimposed stress tensor $\tilde{\mathbf{S}}_M$ on $\mathscr{B}_0^M$ and the assumed 'strain' $\tilde{C}_F$ and the superimposed fiber stress $\tilde{S}_F$ on $\mathscr{B}_0^F$ by considering the virtual power principle

$$\begin{aligned} &\delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}, \dot{\mathbf{v}}, \dot{\mathbf{p}}, \dot{\tilde{\mathbf{C}}}, \mathbf{S}_M, \dot{\tilde{C}}_F, \mathbf{S}_F; \rho_0, \boldsymbol{\kappa}_{0_M}, \boldsymbol{\kappa}_{0_F}, \mathbf{A}_0, \mathbf{b}, \mathbf{t}, \dot{\tilde{\mathbf{u}}}, \tilde{\mathbf{S}}_M, \tilde{S}_F) := \\ &\delta_* \dot{\mathcal{T}}(\dot{\mathbf{u}}, \dot{\mathbf{v}}, \dot{\mathbf{p}}; \rho_0) + \delta_* \dot{\Pi}^{\text{ext}}(\dot{\mathbf{u}}; \rho_0, \mathbf{b}, \mathbf{t}, \dot{\tilde{\mathbf{u}}}) + \delta_* \dot{\Pi}^{\text{int}}(\dot{\mathbf{u}}, \dot{\tilde{\mathbf{C}}}, \mathbf{S}_M, \dot{\tilde{C}}_F, \mathbf{S}_F; \boldsymbol{\kappa}_{0_M}, \boldsymbol{\kappa}_{0_F}, \mathbf{A}_0, \tilde{\mathbf{S}}_M, \tilde{S}_F) = 0 \end{aligned} \tag{34}$$

The virtual kinetic power is identical to Eq. (22) and the virtual external power is identical to Eq. (23). But the virtual internal power is now given by

$$\begin{aligned} &\delta_* \dot{\Pi}^{\text{int}}(\dot{\mathbf{u}}, \dot{\tilde{\mathbf{C}}}, \mathbf{S}_M, \dot{\tilde{C}}_F, \mathbf{S}_F; \boldsymbol{\kappa}_{0_M}, \boldsymbol{\kappa}_{0_F}, \mathbf{A}_0, \tilde{\mathbf{S}}_M, \tilde{S}_F) := \\ &\quad \frac{1}{2} \int_{\mathscr{B}_0} \Big[2\, \mathrm{D} W_M(\tilde{\mathbf{C}}; \boldsymbol{\kappa}_{0_M}) + \tilde{\mathbf{S}}_M - \mathbf{S}_M \Big] : \delta_* \dot{\tilde{\mathbf{C}}} \, \mathrm{d}V \\ &\quad \frac{1}{2} \int_{\mathscr{B}_0} \Big[2\, \mathrm{D} \hat{W}_F(\tilde{C}_F; \boldsymbol{\kappa}_{0_F}) + \tilde{S}_F - \mathbf{S}_F : \mathbf{A}_0 \Big] \, \delta_* \dot{\tilde{C}}_F \, \mathrm{d}V \\ &- \frac{1}{2} \int_{\mathscr{B}_0} \delta_* \mathbf{S}_M : \Big[\dot{\tilde{\mathbf{C}}} - \dot{\mathbf{C}}(\dot{\mathbf{u}}) \Big] \mathrm{d}V + \frac{1}{2} \int_{\mathscr{B}_0} \mathbf{S}_M : \delta_* \dot{\mathbf{C}}(\dot{\mathbf{u}}) \, \mathrm{d}V \\ &- \frac{1}{2} \int_{\mathscr{B}_0} \delta_* \mathbf{S}_F : \Big[\dot{\tilde{C}}_F \, \mathbf{A}_0 - \dot{\mathbf{C}}_F(\dot{\mathbf{u}}) \Big] \, \mathrm{d}V + \frac{1}{2} \int_{\mathscr{B}_0} \mathbf{S}_F : \delta_* \dot{\mathbf{C}}_F(\dot{\mathbf{u}}) \, \mathrm{d}V \end{aligned} \tag{35}$$

with

$$\mathbf{C}_F(\mathbf{u}) := \left[(\nabla \mathbf{u} + \mathbf{I}) \, \mathbf{A}_0 \right]^T \left[\nabla \mathbf{u} + \mathbf{I} \right] \, \mathbf{A}_0 = \mathbf{A}_0 (\nabla \mathbf{u} + \mathbf{I})^T (\nabla \mathbf{u} + \mathbf{I}) \, \mathbf{A}_0 \tag{36}$$

Bearing in mind the identity

$$\frac{1}{2} \mathbf{S}_F : \delta_* \dot{\mathbf{C}}_F(\dot{\mathbf{u}}) = \frac{1}{2} \mathbf{S}_F : \mathbf{A}_0 \, \delta_* \dot{\mathbf{C}}(\dot{\mathbf{u}}) \, \mathbf{A}_0 = \frac{1}{2} \mathbf{F} \, \mathbf{A}_0 \, \mathbf{S}_F \, \mathbf{A}_0 : \nabla \dot{\mathbf{u}} = \mathbf{F}_F \left[\mathbf{S}_F : \mathbf{A}_0 \right] : \nabla \dot{\mathbf{u}} \tag{37}$$

integration by parts leads to

$$\frac{1}{2} \int_{\mathscr{B}_0} \mathbf{S}_F : \delta_* \dot{\mathbf{C}}_F(\dot{\mathbf{u}}) \, \mathrm{d}V = \int_{\partial_t \mathscr{B}_0} \mathbf{F}_F \left[\mathbf{S}_F : \mathbf{A}_0 \right] \mathbf{N} \cdot \delta_* \dot{\mathbf{u}} \, \mathrm{d}A - \int_{\mathscr{B}_0} \mathrm{DIV} \left[\mathbf{F}_F \left(\mathbf{S}_F : \mathbf{A}_0 \right) \right] \cdot \delta_* \dot{\mathbf{u}} \, \mathrm{d}V \tag{38}$$

Again, by rearranging the terms in Eq. (34) due to the independent variations $\delta_*\dot{\mathbf{p}}$, $\delta_*\dot{\mathbf{v}}$, $\delta_*\mathbf{S}_M$, $\delta_*\dot{\tilde{\mathbf{C}}}$, $\delta_*\mathbf{S}_F$, $\delta_*\dot{\tilde{C}}_F$ and $\delta_*\dot{\mathbf{u}}$, we obtain the variational form

$$
\begin{aligned}
0 = &\int_{\mathscr{B}_0} [\rho_0\mathbf{v} - \mathbf{p}] \cdot \delta_*\dot{\mathbf{v}} \,\mathrm{d}V - \int_{\mathscr{B}_0} \delta_*\dot{\mathbf{p}} \cdot [\mathbf{v} - \dot{\mathbf{u}}] \,\mathrm{d}V \\
&- \int_{\mathscr{B}_0} [\mathrm{DIV}[\mathbf{F}\mathbf{S}_M + \mathbf{F}_F\,(\mathbf{S}_F : \mathbf{A}_0)] + \rho_0\mathbf{b} - \dot{\mathbf{p}}] \cdot \delta_*\dot{\mathbf{u}} \,\mathrm{d}V \\
&- \int_{\partial_t\mathscr{B}_0} [\mathbf{t} - (\mathbf{F}\mathbf{S}_M + \mathbf{F}_F\,(\mathbf{S}_F : \mathbf{A}_0))\,\mathbf{N}] \cdot \delta_*\dot{\mathbf{u}} \,\mathrm{d}A - \int_{\partial_u\mathscr{B}_0} \mathbf{h} \cdot \delta_*\dot{\mathbf{u}} \,\mathrm{d}A \\
&- \frac{1}{2}\int_{\mathscr{B}_0} \Big[\mathbf{S}_M - 2\,\mathrm{D}W_M(\tilde{\mathbf{C}}; \boldsymbol{\kappa}_{0_M}) - \tilde{\mathbf{S}}_M\Big] : \delta_*\dot{\tilde{\mathbf{C}}} \,\mathrm{d}V - \frac{1}{2}\int_{\mathscr{B}_0} \Big[\dot{\tilde{\mathbf{C}}} - \dot{\mathbf{C}}(\dot{\mathbf{u}})\Big] : \delta_*\mathbf{S}_M \,\mathrm{d}V \\
&- \int_{\mathscr{B}_0} \Big[\mathbf{S}_F : \mathbf{A}_0 - 2\,\mathrm{D}\hat{W}_F(\tilde{C}_F; \boldsymbol{\kappa}_{0_F}) - \tilde{S}_F\Big] \frac{\delta_*\dot{\tilde{C}}_F}{2} \,\mathrm{d}V - \frac{1}{2}\int_{\mathscr{B}_0} \Big[\dot{\tilde{C}}_F\,\mathbf{A}_0 - \dot{\mathbf{C}}_F(\dot{\mathbf{u}})\Big] : \delta_*\mathbf{S}_F \,\mathrm{d}V
\end{aligned}
\tag{39}
$$

Taking the fundamental theorem of variational calculus into account, we arrive at the EULER-LAGRANGE equations

$$
\mathbf{v} = \dot{\mathbf{u}} \quad \text{with} \quad \mathbf{u}(t_0) = \mathbf{u}_0 \tag{40}
$$

$$
\rho_0\mathbf{v} = \mathbf{p} \qquad \forall t \geqslant t_0 \tag{41}
$$

$$
\mathrm{DIV}\,[\mathbf{F}\mathbf{S}_M + \mathbf{F}_F\,(\mathbf{S}_F : \mathbf{A}_0)] + \rho_0\mathbf{b} = \dot{\mathbf{p}} \quad \text{with} \quad \mathbf{p}(t_0) = \mathbf{p}_0 \equiv \rho_0\mathbf{v}_0 \tag{42}
$$

$$
\dot{\mathbf{C}}(\dot{\mathbf{u}}) = \dot{\tilde{\mathbf{C}}} \quad \text{with} \quad \tilde{\mathbf{C}}(t_0) = \mathbf{C}(\mathbf{u}_0) \tag{43}
$$

$$
\dot{\mathbf{C}}_F(\dot{\mathbf{u}}) : \mathbf{A}_0 = \dot{\tilde{C}}_F \quad \text{with} \quad \tilde{C}_F(t_0) = \mathbf{C}_F(\mathbf{u}_0) : \mathbf{A}_0 \tag{44}
$$

$$
2\,\mathrm{D}W_M(\tilde{\mathbf{C}}; \boldsymbol{\kappa}_{0_M}) + \tilde{\mathbf{S}}_M = \mathbf{S}_M \qquad \forall t \geqslant t_0 \tag{45}
$$

$$
\Big[2\,\mathrm{D}\hat{W}_F(\tilde{C}_F; \boldsymbol{\kappa}_{0_F}) + \tilde{S}_F\Big]\,\mathbf{A}_0 = \mathbf{S}_F \qquad \forall t \geqslant t_0 \tag{46}
$$

with $\tilde{\mathbf{S}}_M := \mathbf{O}$ and $\tilde{S}_F := 0$ for satisfying the total energy balance, and the boundary conditions

$$
[\mathbf{F}\mathbf{S}_M + \mathbf{F}_F\,(\mathbf{S}_F : \mathbf{A}_0)]\,\mathbf{N} = \mathbf{t} \qquad \forall t \geqslant t_0 \quad \text{on} \quad \partial_t\mathscr{B}_0 \tag{47}
$$

$$
\delta_*\dot{\mathbf{u}} = \mathbf{0} \Longleftrightarrow \dot{\mathbf{u}} = \dot{\bar{\mathbf{u}}} \;\text{ with }\; \mathbf{u}(t_0) = \bar{\mathbf{u}}(t_0) \quad \text{on} \quad \partial_u\mathscr{B}_0 \tag{48}
$$

Consequently, the PIOLA-KIRCHHOFF stress fields $\mathbf{S}_M$ and $\mathbf{S}_F$ are temporally discontinuous, but the assumed 'strain' fields $\tilde{\mathbf{C}}$ and $\tilde{C}_F$ are again temporally continuous.

3 FULLY-DISCRETE WEAK FORMULATION

Next, we derive the temporally and spatially discrete weak variational formulation. In this section, we restrict ourselves to a linear piecewise continuous time approximation in $\mathbf{u}$, $\mathbf{v}$, $\mathbf{p}$ (compare Reference [5]) as well as $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{C}}_F$, in order to demonstrate the consistency of the variational derivation of the assumed 'strain' approximations and the superimposed stress tensors $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{S}}_F$ with Reference [8]. But note that in the unpartitioned case both approximations are new for *higher-order time approximations*.

3.1 The unpartitioned strain energy function

The time integrator presented here follows from considering *the virtual power principle at collocation points* ξ_i of the time interval $[t_0, t_N]$ of interest, introduced by a time integration

$$
\begin{aligned}
&\int_{t_0}^{t_N} \delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}(t), \dot{\mathbf{v}}(t), \dot{\mathbf{p}}(t), \dot{\tilde{\mathbf{C}}}(t), \mathbf{S}(t); \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}(t), \mathbf{t}(t), \dot{\bar{\mathbf{u}}}(t), \tilde{\mathbf{S}}(t))\, \mathrm{d}t = \\
&\sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}^n(t), \dot{\mathbf{v}}^n(t), \dot{\mathbf{p}}^n(t), \dot{\tilde{\mathbf{C}}}^n(t), \mathbf{S}^n(t); \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}^n(t), \mathbf{t}^n(t), \dot{\bar{\mathbf{u}}}^n(t), \tilde{\mathbf{S}}^n(t))\, \mathrm{d}t \approx \\
&\sum_{n=0}^{N-1} \int_0^1 \delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}_h^n(\alpha), \dot{\mathbf{v}}_h^n(\alpha), \dot{\mathbf{p}}_h^n(\alpha), \dot{\tilde{\mathbf{C}}}_h^n(\alpha), \mathbf{S}_h^n(\alpha); \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}_h^n(\alpha), \mathbf{t}_h^n(\alpha), \dot{\bar{\mathbf{u}}}_h^n(\alpha), \tilde{\mathbf{S}}_h^n(\alpha)) h_n \mathrm{d}\alpha \approx \\
&\sum_{n=0}^{N-1} \delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}_h^n(\xi_1), \dot{\mathbf{v}}_h^n(\xi_1), \dot{\mathbf{p}}_h^n(\xi_1), \dot{\tilde{\mathbf{C}}}_h^n(\xi_1), \mathbf{S}_h^n(\xi_1); \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}_h^n(\xi_1), \mathbf{t}_h^n(\xi_1), \dot{\bar{\mathbf{u}}}_h^n(\xi_1), \tilde{\mathbf{S}}_h^n(\xi_1))\, h_n \\
&\sum_{n=0}^{N-1} \delta_* \dot{\mathcal{H}}_d(\mathbf{u}_{n+1}, \mathbf{v}_{n+1}, \mathbf{p}_{n+1}, \tilde{\mathbf{C}}_{n+1}, \mathbf{S}_{n+\frac{1}{2}}; \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}_{n+\frac{1}{2}}, \mathbf{t}_{n+\frac{1}{2}}, \bar{\mathbf{u}}_{n+1}, \tilde{\mathbf{S}}_{n+\frac{1}{2}})\, h_n \doteq 0
\end{aligned}
\tag{49}
$$

and the normalized time $\alpha \in [0, 1]$ via the linear transformation

$$
\tau : [t_n, t_{n+1}] \ni t \mapsto t_n + \alpha\,(t_{n+1} - t_n) = t_n + \alpha\, h_n \tag{50}
$$

with respect to the time step size h_n, and after applying the midpoint rule with the one Gauss point $\xi_1 = \frac{1}{2}$ to the mentioned piecewise linear time approximations

$$
\mathbf{u}_h^n(\alpha) := \mathbf{u}_n + \alpha\,(\mathbf{u}_{n+1} - \mathbf{u}_n) \qquad \mathbf{v}_h^n(\alpha) := \mathbf{v}_n + \alpha\,(\mathbf{v}_{n+1} - \mathbf{v}_n) \tag{51}
$$

$$
\mathbf{p}_h^n(\alpha) := \mathbf{p}_n + \alpha\,\big(\mathbf{p}_{n+1} - \mathbf{p}_n\big) \qquad \tilde{\mathbf{C}}_h^n(\alpha) := \mathbf{C}_n + \alpha\,(\mathbf{C}_{n+1} - \mathbf{C}_n) \tag{52}
$$

In the following, we use the common finite difference notation $(\bullet)_{n+\frac{1}{2}}$ for symbols $(\bullet)_h^n(\frac{1}{2})$. Without integrating by parts but rearranging termes in Eq. (49) according to the independent variations, we obtain the semi-discrete variational form

$$
\begin{aligned}
0 \;=\; & \int_{\mathscr{B}_0} \left[\rho_0 \mathbf{v}_{n+\frac{1}{2}} - \mathbf{p}_{n+\frac{1}{2}}\right] \cdot \delta_* \mathbf{v}_{n+1}\, \mathrm{d}V - \int_{\mathscr{B}_0} \delta_* \mathbf{p}_{n+1} \cdot \left[\mathbf{v}_{n+\frac{1}{2}} - \frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{h_n}\right] \mathrm{d}V \\
& - \frac{1}{2} \int_{\mathscr{B}_0} \left[\mathbf{S}_{n+\frac{1}{2}} - 2\,\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) - \tilde{\mathbf{S}}_{n+\frac{1}{2}}\right] : \delta_* \mathbf{C}_{n+1}\, \mathrm{d}V \\
& - \frac{1}{2} \int_{\mathscr{B}_0} \left[\mathbf{C}_{n+1} - \mathbf{C}_n - (\mathbf{F}_{n+1} + \mathbf{F}_n)^T (\mathbf{F}_{n+1} - \mathbf{F}_n)\right] : \delta_* \mathbf{S}_{n+\frac{1}{2}}\, \mathrm{d}V \\
& + \int_{\mathscr{B}_0} \left[\frac{\mathbf{p}_{n+1} - \mathbf{p}_n}{h_n} + \mathbf{B}_{n+\frac{1}{2}}^T \mathbf{S}_{n+\frac{1}{2}} - \rho_0 \mathbf{b}_{n+\frac{1}{2}}\right] \cdot \delta_* \mathbf{u}_{n+1}\, \mathrm{d}V \\
& - \int_{\partial_t \mathscr{B}_0} \mathbf{t}_{n+\frac{1}{2}} \cdot \delta_* \mathbf{u}_{n+1}\, \mathrm{d}A - \int_{\partial_u \mathscr{B}_0} \mathbf{h}_{n+\frac{1}{2}} \cdot \delta_* \mathbf{u}_{n+1}\, \mathrm{d}A
\end{aligned}
\tag{53}
$$

with the linearized strain operator $\mathbf{B}_{n+\frac{1}{2}}$ defined by [8]

$$
2\,\mathbf{B}_{n+\frac{1}{2}}\, \delta_* \mathbf{u}_{n+1} := \mathbf{F}_{n+\frac{1}{2}}^T \nabla(\delta_* \mathbf{u}_{n+1}) + \nabla(\delta_* \mathbf{u}_{n+1})^T \mathbf{F}_{n+\frac{1}{2}} \tag{54}
$$

and bearing in mind the differentiation rule

$$\overset{\cdot}{\overline{(\bullet)}} = \frac{\mathrm{d}(\bullet)}{\mathrm{d}\alpha}\frac{\mathrm{d}\alpha}{\mathrm{d}t} = \frac{\mathrm{d}(\bullet)}{\mathrm{d}\alpha}\,\frac{1}{h_n} \tag{55}$$

as well as the *vanishing* variations $\delta_*\mathbf{u}_n$, $\delta_*\mathbf{v}_n$, $\delta_*\mathbf{p}_n$ and $\delta_*\mathbf{C}_n$ due to the initial conditions

$$\mathbf{u}(t_0) = \mathbf{u}_0 \qquad\qquad \mathbf{p}(t_0) = \rho_0\mathbf{v}_0 \tag{56}$$

$$\mathbf{v}(t_0) = \mathbf{v}_0 \qquad\qquad \tilde{\mathbf{C}}(t_0) = (\nabla\mathbf{u}_0 + \mathbf{I})^T(\nabla\mathbf{u}_0 + \mathbf{I}) \tag{57}$$

in the first time step $[t_0, t_1]$. At this point, we are able to derive spatially local relations of tensor fields at each point $\mathbf{X} \in \mathscr{B}_0$, which can be used to eliminate variables in the discrete system of equations of motion without taking spatial integrals. On the other hand, we may keep all the variables and solve a multifield formulation if we are interested in these variables for post-processing purposes, or we may eliminate the variables after the spatial integrals have been taken. The first line of Eq. (53) leads to

$$\mathbf{p}_n = \rho_0\mathbf{v}_n \qquad\qquad \mathbf{p}_{n+1} = \rho_0\mathbf{v}_{n+1} \qquad\qquad h_n\,\mathbf{v}_{n+\frac{1}{2}} = \mathbf{u}_{n+1} - \mathbf{u}_n \tag{58}$$

where Eq. (58.1) is obvious from the initial condition in Eq. (56.2), and Eq. (58.3) can be seen as first equation of motion [2]. The second line of Eq. (58) furnishes the discrete constitutive relation

$$\mathbf{S}_{n+\frac{1}{2}} = 2\,\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) + \tilde{\mathbf{S}}_{n+\frac{1}{2}} \tag{59}$$

But note that the superimposed second PIOLA-KIRCHHOFF stress tensor $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ *must not vanish* for energy consistency as in the continuous setting (compare Reference [7]). We derive it below in a separate variational problem. In the third line of Eq. (53), we take into account the symmetry of $\delta_*\mathbf{S}_{n+\frac{1}{2}}$, leading to the identity

$$\left[\mathbf{F}_{n+1}^T\mathbf{F}_n - \mathbf{F}_n^T\mathbf{F}_{n+1}\right] : \delta_*\mathbf{S}_{n+\frac{1}{2}} = 0 \tag{60}$$

and consequently to the local relation

$$\mathbf{C}_{n+1} - \mathbf{C}_n = \mathbf{F}_{n+1}^T\mathbf{F}_{n+1} - \mathbf{F}_n^T\mathbf{F}_n \Longleftrightarrow \mathbf{C}_{n+1} = \mathbf{F}_{n+1}^T\mathbf{F}_{n+1} \tag{61}$$

according to the initial condition in Eq. (57.2). Hence, we arrive at the following *variationally consistent assumed 'strain' approximation*, which is proposed for energy consistent time stepping schemes at least since the publication of Reference [6]:

$$\tilde{\mathbf{C}}_{n+\frac{1}{2}} := \frac{1}{2}\left[\mathbf{C}_n + \mathbf{C}_{n+1}\right] \tag{62}$$

The spatial approximation in the variational formulation is based on trilinear shape functions for an eight-node brick element for the volume and bilinear shape functions for an four-node quadrilateral element for the boundaries, which approximate the geometry in $\mathscr{B}_0$, the displacement vector $\mathbf{u}$ and the material velocity vector $\mathbf{v}$ at the considered discrete time points t_n. Hence, following the notation in Reference [10], we apply the approximations

$$\mathsf{u} = \mathsf{N}\,\mathsf{u} \qquad\qquad \delta_*\mathsf{u} = \mathsf{N}\,\delta_*\mathsf{u} \tag{63}$$

$$\mathsf{v} = \mathsf{N}\,\mathsf{v} \qquad\qquad \delta_*\mathsf{v} = \mathsf{N}\,\delta_*\mathsf{v} \tag{64}$$

where $\mathbf{N}$ denotes the matrix of the trilinear shape functions and the vector $\mathbf{u}$ combines the nodal displacements. An analogous notation is used for the nodal velocities and nodal variations, respectively. On the boundary $\partial\mathscr{B}_0$, we apply the approximations

$$\mathbf{u} = \bar{\mathbf{N}}\,\mathbf{u} \qquad\qquad \delta_*\mathbf{u} = \bar{\mathbf{N}}\,\delta_*\mathbf{u} \tag{65}$$

where $\bar{\mathbf{N}}$ denotes the matrix of the bilinear shape functions. The last two lines of Eq. (53) then leads to the discrete variational formulation

$$\delta_*\mathbf{u}_{n+1}^T\left[\mathbf{M}\,\frac{\mathbf{v}_{n+1}-\mathbf{v}_n}{h_n} + \int_{\mathscr{B}_0}\mathbf{B}_{n+\frac{1}{2}}^T\mathbf{S}_{n+\frac{1}{2}}\,\mathrm{d}V\right] = \delta_*\mathbf{u}_{n+1}^T\left[\mathbf{H}_t\,\mathbf{t}_{n+\frac{1}{2}} + \mathbf{H}_u\,\mathbf{h}_{n+\frac{1}{2}} + \mathbf{M}\,\mathbf{b}_{n+\frac{1}{2}}\right] \tag{66}$$

with the system matrices

$$\mathbf{M} := \int_{\mathscr{B}_0}\rho_0\,\mathbf{N}^T\mathbf{N}\,\mathrm{d}V \qquad \mathbf{H}_t := \int_{\partial_t\mathscr{B}_0}\bar{\mathbf{N}}^T\bar{\mathbf{N}}\,\mathrm{d}V \qquad \mathbf{H}_u := \int_{\partial_u\mathscr{B}_0}\bar{\mathbf{N}}^T\bar{\mathbf{N}}\,\mathrm{d}V \tag{67}$$

as well as the matrix representations $\mathbf{B}_{n+\frac{1}{2}}$ and $\mathbf{S}_{n+\frac{1}{2}}$ of the linearized strain operator and the second PIOLA-KIRCHHOFF stress tensor, respectively. Finally, we apply the fundamental theorem of variational calculus and arrive at the discrete system of equations of motion

$$\mathbf{M}\,\frac{\mathbf{v}_{n+1}-\mathbf{v}_n}{h_n} + \int_{\mathscr{B}_0}\mathbf{B}_{n+\frac{1}{2}}^T\mathbf{S}_{n+\frac{1}{2}} = \mathbf{H}_t\,\mathbf{t}_{n+\frac{1}{2}} + \mathbf{H}_u\,\mathbf{h}_{n+\frac{1}{2}} + \mathbf{M}\,\mathbf{b}_{n+\frac{1}{2}} \tag{68}$$

If we now multiply Eq. (68) on both sides from the left by the velocity vector

$$\mathbf{v}_{n+\frac{1}{2}} = \frac{1}{2}\left[\mathbf{v}_{n+1}+\mathbf{v}_n\right] \tag{69}$$

the first term on the left hand side takes the form

$$\begin{aligned}\mathbf{v}_{n+\frac{1}{2}}^T\mathbf{M}\,\frac{\mathbf{v}_{n+1}-\mathbf{v}_n}{h_n} &= \frac{1}{2\,h_n}\left[\mathbf{v}_{n+1}^T\mathbf{M}\,\mathbf{v}_{n+1} - \mathbf{v}_n^T\mathbf{M}\,\mathbf{v}_n\right]\\ &= \frac{\mathcal{T}_{n+1}-\mathcal{T}_n}{h_n}\end{aligned} \tag{70}$$

which denotes the discrete time derivative of the total kinetic energy. On the righthand side of Eq. (68), we obtain the discrete external power

$$\begin{aligned}\mathbf{v}_{n+\frac{1}{2}}^T\left[\mathbf{H}_t\,\mathbf{t}_{n+\frac{1}{2}} + \mathbf{H}_u\,\mathbf{h}_{n+\frac{1}{2}} + \mathbf{M}\,\mathbf{b}_{n+\frac{1}{2}}\right] &= \frac{\mathbf{u}_{n+1}^T-\mathbf{u}_n^T}{h_n}\left[\mathbf{H}_t\,\mathbf{t}_{n+\frac{1}{2}} + \mathbf{H}_u\,\mathbf{h}_{n+\frac{1}{2}} + \mathbf{M}\,\mathbf{b}_{n+\frac{1}{2}}\right]\\ &= -\frac{\Pi_{n+1}^{\mathrm{ext}}-\Pi_n^{\mathrm{ext}}}{h_n}\end{aligned} \tag{71}$$

where Eq. (58.3) have been taken into account. Accordingly, we arrive at the discrete total energy balance *if the relation*

$$\begin{aligned}\int_{\mathscr{B}_0}\mathbf{v}_{n+\frac{1}{2}}\cdot\mathbf{B}_{n+\frac{1}{2}}^T\left[2\,\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}_{n+\frac{1}{2}}\right]\mathrm{d}V &=\\ \int_{\mathscr{B}_0}\frac{\mathbf{u}_{n+1}-\mathbf{u}_n}{h_n}\cdot\mathbf{B}_{n+\frac{1}{2}}^T\left[2\,\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\mathbf{A}_0,\boldsymbol{\kappa}_0)+\tilde{\mathbf{S}}_{n+\frac{1}{2}}\right]\mathrm{d}V &= \frac{\Pi_{n+1}^{\mathrm{int}}-\Pi_n^{\mathrm{int}}}{h_n}\end{aligned} \tag{72}$$

or

$$W_{n+1} - W_n = \left[\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) + \frac{1}{2}\,\tilde{\mathbf{S}}_{n+\frac{1}{2}} \right] : 2\,\mathbf{B}_{n+\frac{1}{2}}\,[\mathbf{u}_{n+1} - \mathbf{u}_n] \tag{73}$$

is fulfilled. On the other hand, employing the assumed 'strain' tensor in Eq. (61) in the definition of the linearized strain operator in Eq. (54), we obtain the identity

$$\begin{aligned} 2\,\mathbf{B}_{n+\frac{1}{2}}\,[\mathbf{u}_{n+1} - \mathbf{u}_n] &= \mathbf{F}^T_{n+\frac{1}{2}}\,[\mathbf{F}_{n+1} - \mathbf{F}_n] + \mathbf{F}_{n+\frac{1}{2}}\,[\mathbf{F}_{n+1} - \mathbf{F}_n]^T \\ &= \frac{1}{2}\left[\mathbf{C}_{n+1} - \mathbf{C}_n + \mathbf{C}^T_{n+1} - \mathbf{C}^T_n\right] \\ &= \mathbf{C}_{n+1} - \mathbf{C}_n \end{aligned} \tag{74}$$

which in the end leads to the scalar-valued constraint $\mathcal{G}(\tilde{\mathbf{S}}_{n+\frac{1}{2}})$ on the superimposed stress field $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ at each point $\mathbf{X} \in \mathscr{B}_0$, given by

$$\mathcal{G}(\tilde{\mathbf{S}}_{n+\frac{1}{2}}) := W_{n+1} - W_n - \left[\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) + \frac{1}{2}\,\tilde{\mathbf{S}}_{n+\frac{1}{2}} \right] : [\mathbf{C}_{n+1} - \mathbf{C}_n] = 0 \tag{75}$$

As the superimposed stress tensor is symmetric, $n_{\mathrm{dim}}(n_{\mathrm{dim}}+1)/2$ components of the tensor $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ has to be uniquely determined such that the scalar-valued constraint in Eq. (75) is satisfied. Therefore, we solve the *separate constraint variational problem*

$$\delta_* \mathcal{L}(\mu, \tilde{\mathbf{S}}_{n+\frac{1}{2}}) = 0 \tag{76}$$

with

$$\mathcal{L}(\mu, \tilde{\mathbf{S}}_{n+\frac{1}{2}}) := \frac{1}{2}\,\tilde{\mathbf{C}}_{n+\frac{1}{2}}\,\tilde{\mathbf{S}}_{n+\frac{1}{2}} : \tilde{\mathbf{S}}_{n+\frac{1}{2}}\,\tilde{\mathbf{C}}_{n+\frac{1}{2}} + \mu\,\mathcal{G}(\tilde{\mathbf{S}}_{n+\frac{1}{2}}) \tag{77}$$

using the corresponding discrete EULER-LAGRANGE equations

$$\frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{S}}_{n+\frac{1}{2}}} \equiv \tilde{\mathbf{C}}_{n+\frac{1}{2}}\,\tilde{\mathbf{S}}_{n+\frac{1}{2}}\,\tilde{\mathbf{C}}_{n+\frac{1}{2}} - \frac{\mu}{2}\,[\mathbf{C}_{n+1} - \mathbf{C}_n] = \mathbf{O} \qquad \frac{\partial \mathcal{L}}{\partial \mu} \equiv \mathcal{G}(\tilde{\mathbf{S}}_{n+\frac{1}{2}}) = 0 \tag{78}$$

Note that in Eq. (77) the right CAUCHY-GREEN tensor $\tilde{\mathbf{C}}_{n+\frac{1}{2}}$ operates as metric tensor as in the physically consistent deviator stress in Reference [12]. Therefore, this constraint variational problem could be also pushed forward to the current configuration $\mathscr{B}_t$, and formulated with the KIRCHHOFF stress tensor $\boldsymbol{\tau}$ and the metric $\mathbf{g}$ in $\mathscr{B}_t$. After inserting Eq. (78.1) in Eq. (78.2), we arrive at the two spatially local discrete EULER-LAGRANGE equations for the superimposed stress field $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ and the scaling factor μ at each point $\mathbf{X} \in \mathscr{B}_0$, given by

$$\tilde{\mathbf{S}}_{n+\frac{1}{2}} = \frac{\mu}{2}\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}\,[\mathbf{C}_{n+1} - \mathbf{C}_n]\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \tag{79}$$

$$2\,\mathcal{G}(\mathbf{O}) = \frac{\mu}{2}\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}\,[\mathbf{C}_{n+1} - \mathbf{C}_n] : [\mathbf{C}_{n+1} - \mathbf{C}_n]\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \tag{80}$$

We are able to search numerically for the LAGRANGE multiplier μ, but usually it is eliminated analytically. This leads to the stress tensor [8]

$$\tilde{\mathbf{S}}_{n+\frac{1}{2}} = 2\,\frac{\mathcal{G}(\mathbf{O})}{\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}\,[\mathbf{C}_{n+1} - \mathbf{C}_n] : [\mathbf{C}_{n+1} - \mathbf{C}_n]\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}}\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}\,[\mathbf{C}_{n+1} - \mathbf{C}_n]\,\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \tag{81}$$

Consequently, using the superimposed discrete stress tensor in Eq. (81), the discrete equation of motion in Eq. (68) leads, by design, to the discrete energy balance

$$\mathcal{T}_{n+1} - \mathcal{T}_n + \Pi^{\text{int}}_{n+1} - \Pi^{\text{int}}_n + \Pi^{\text{ext}}_{n+1} - \Pi^{\text{ext}}_n = 0 \iff \mathcal{H}_{n+1} = \mathcal{H}_n \tag{82}$$

which indicates exact algorithmic total energy conservation. But note that in a numerical implementation, the exact algorithmic total energy conservation is indicated by

$$|\mathcal{H}_{n+1} - \mathcal{H}_n| < \mathsf{tol} \tag{83}$$

where tol denotes the tolerance of the applied NEWTON-RAPHSON method for solving the nonlinear discrete EULER-LAGRANGE equations [11]. Further, you should be careful with solution steps where $\mathbf{C}_{n+1} \approx \mathbf{C}_n$ when applying the stress formula in Eq. (81). You should generally implement a prestep formula for the displacements $\mathbf{u}_{n+1}$ taking into account all applied loads and initial conditions [13]. If the tensor $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ is *neglected* then a TAYLOR series expansion of the strain energies W_n and W_{n+1} at the assumed 'strain' tensor $\tilde{\mathbf{C}}_{n+\frac{1}{2}}$, given by

$$\mathrm{D}W(\tilde{\mathbf{C}}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) : [\mathbf{C}_{n+1} - \mathbf{C}_n] = W_{n+1} - W_n + \mathcal{O}\left(\|\mathbf{C}_{n+1} - \mathbf{C}_n\|^3\right) \tag{84}$$

shows that Eq. (83) can be guaranteed only for $\|\mathbf{C}_{n+1} - \mathbf{C}_n\| < \mathsf{tol}$. Therefore, we conclude that energy consistency of the discrete EULER-LAGRANGE equations is only given if the discrete superimposed stress tensor $\tilde{\mathbf{S}}_{n+\frac{1}{2}}$ is non-vanishing.

3.2 The partitioned strain energy function

The time stepping scheme for the partitioned strain energy also follows from a time integration of the corresponding virtual power principle on the time interval $[t_0, t_N]$ of interest. Thus, we obtain a discrete variational condition at the collocation point ξ_1, given by

$$\begin{aligned}\delta_* \dot{\mathcal{H}}_d(&\mathbf{u}_{n+1}, \mathbf{v}_{n+1}, \mathbf{p}_{n+1}, \tilde{\mathbf{C}}_{n+1}, \mathbf{S}_{M_{n+\frac{1}{2}}}, \tilde{C}_{F_{n+1}}, \mathbf{S}_{F_{n+\frac{1}{2}}}; \\ &\rho_0, \boldsymbol{\kappa}_{0_F}, \boldsymbol{\kappa}_{0_M}, \mathbf{A}_0, \mathbf{b}_{n+\frac{1}{2}}, \mathbf{t}_{n+\frac{1}{2}}, \bar{\mathbf{u}}_{n+1}, \tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}}, \tilde{S}_{F_{n+\frac{1}{2}}}) \, h_n = 0\end{aligned} \tag{85}$$

The assumed squared fiber stretch $\tilde{C}_F$ is also linear piecewise continuous approximated by

$$\tilde{C}^n_{F_h}(\alpha) := C_{F_n} + \alpha \left(C_{F_{n+1}} - C_{F_n}\right) \tag{86}$$

with the 'initial' condition on the time step $[t_n, t_{n+1}]$, given by

$$C_{F_n} = \mathbf{C}_{F_n} : \mathbf{A}_0 = \mathbf{F}^T_{F_n} \mathbf{F}_{F_n} : \mathbf{A}_0 \tag{87}$$

Rearranging termes in Eq. (85) according to the independent variations $\delta_*\mathbf{p}_{n+1}$, $\delta_*\mathbf{v}_{n+1}$, $\delta_*\mathbf{u}_{n+1}$, $\delta_*\mathbf{S}_{M_{n+\frac{1}{2}}}$, $\delta_*\mathbf{C}_{n+1}$, $\delta_*\mathbf{S}_{F_{n+\frac{1}{2}}}$ and $\delta_* C_{F_{n+1}}$, we obtain the semi-discrete variational form

$$
\begin{aligned}
0 \;=\; & \int_{\mathscr{B}_0} \left[\rho_0 \mathbf{v}_{n+\frac{1}{2}} - \mathbf{p}_{n+\frac{1}{2}}\right] \cdot \delta_*\mathbf{v}_{n+1}\,\mathrm{d}V - \int_{\mathscr{B}_0} \delta_*\mathbf{p}_{n+1} \cdot \left[\mathbf{v}_{n+\frac{1}{2}} - \frac{\mathbf{u}_{n+1}-\mathbf{u}_n}{h_n}\right] \mathrm{d}V \\
& - \frac{1}{2}\int_{\mathscr{B}_0} \left[\mathbf{S}_{M_{n+\frac{1}{2}}} - 2\,\mathrm{D}W_M(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\boldsymbol{\kappa}_{0_M}) - \tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}}\right] : \delta_*\mathbf{C}_{n+1}\,\mathrm{d}V \\
& - \frac{1}{2}\int_{\mathscr{B}_0} \left[\mathbf{S}_{F_{n+\frac{1}{2}}} : \mathbf{A}_0 - 2\,\mathrm{D}\hat{W}_F(\tilde{C}_{F_{n+\frac{1}{2}}};\boldsymbol{\kappa}_{0_F}) - \tilde{S}_{F_{n+\frac{1}{2}}}\right] : \delta_* C_{F_{n+1}}\,\mathrm{d}V \\
& - \frac{1}{2}\int_{\mathscr{B}_0} \left[\mathbf{C}_{n+1} - \mathbf{C}_n - (\mathbf{F}_{n+1}+\mathbf{F}_n)^T(\mathbf{F}_{n+1}-\mathbf{F}_n)\right] : \delta_*\mathbf{S}_{n+\frac{1}{2}}\,\mathrm{d}V \\
& - \frac{1}{2}\int_{\mathscr{B}_0} \left[C_{F_{n+1}}\mathbf{A}_0 - C_{F_n}\mathbf{A}_0 - \left(\mathbf{F}_{F_{n+1}}+\mathbf{F}_{F_n}\right)^T\left(\mathbf{F}_{F_{n+1}}-\mathbf{F}_{F_n}\right)\right] : \delta_*\mathbf{S}_{F_{n+\frac{1}{2}}}\,\mathrm{d}V \\
& + \int_{\mathscr{B}_0} \left[\frac{\mathbf{p}_{n+1}-\mathbf{p}_n}{h_n} + \mathbf{B}^T_{n+\frac{1}{2}}\left[\mathbf{S}_{M_{n+\frac{1}{2}}} + \left(\mathbf{S}_{F_{n+\frac{1}{2}}} : \mathbf{A}_0\right)\mathbf{A}_0\right] - \rho_0\mathbf{b}_{n+\frac{1}{2}}\right] \cdot \delta_*\mathbf{u}_{n+1}\,\mathrm{d}V \\
& - \int_{\partial_t\mathscr{B}_0} \mathbf{t}_{n+\frac{1}{2}} \cdot \delta_*\mathbf{u}_{n+1}\,\mathrm{d}A - \int_{\partial_u\mathscr{B}_0} \mathbf{h}_{n+\frac{1}{2}} \cdot \delta_*\mathbf{u}_{n+1}\,\mathrm{d}A
\end{aligned}
\tag{88}
$$

The first line of Eq. (88) also furnishes the Eqs. (58). The second and third line leads to the discrete constitutive stress relations

$$
\mathbf{S}_{M_{n+\frac{1}{2}}} \;=\; 2\,\mathrm{D}W_M(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\boldsymbol{\kappa}_{0_M}) + \tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}} \tag{89}
$$

$$
\mathbf{S}_{F_{n+\frac{1}{2}}} \;=\; \left[2\,\mathrm{D}\hat{W}_F(\tilde{C}_{F_{n+\frac{1}{2}}};\boldsymbol{\kappa}_{0_F}) + \tilde{S}_{F_{n+\frac{1}{2}}}\right]\mathbf{A}_0 \tag{90}
$$

The fourth and fivth line of Eq. (88) determines the right CAUCHY-GREEN 'strains' at the time point t_{n+1} by the equations

$$
\mathbf{C}_{n+1} = \mathbf{F}^T_{n+1}\mathbf{F}_{n+1} \qquad\qquad C_{F_{n+1}} = \mathbf{C}_{F_{n+1}} : \mathbf{A}_0 = \mathbf{F}^T_{F_{n+1}}\mathbf{F}_{F_{n+1}} : \mathbf{A}_0 \tag{91}
$$

leading to the approximations

$$
\tilde{\mathbf{C}}_{n+\frac{1}{2}} := \frac{1}{2}\left[\mathbf{C}_n + \mathbf{C}_{n+1}\right] \qquad\qquad \tilde{C}_{F_{n+\frac{1}{2}}} := \frac{1}{2}\left[C_{F_n} + C_{F_{n+1}}\right] \tag{92}
$$

Hence, the full symmetric tensor $\mathbf{C}_F$ has not to be stored at the midpoint $t_{n+\frac{1}{2}}$, but merely the scalar 'strain' $\tilde{C}_{F_{n+\frac{1}{2}}}$. Analogous to the unpartitioned case, the last lines of Eq. (88) gives the discrete system of equations of motion

$$
\begin{aligned}
\mathbf{M}\,\frac{\mathbf{v}_{n+1}-\mathbf{v}_n}{h_n} + \int_{\mathscr{B}_0} \mathbf{B}^T_{n+\frac{1}{2}} \Big[& 2\,\mathrm{D}W_M(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\boldsymbol{\kappa}_{0_M}) + 2\,\mathrm{D}\hat{W}_F(\tilde{C}_{F_{n+\frac{1}{2}}};\boldsymbol{\kappa}_{0_F})\,\mathbf{A}_0 \\
& + \tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}} + \tilde{S}_{F_{n+\frac{1}{2}}}\,\mathbf{A}_0\Big]\,\mathrm{d}V = \mathbf{H}_t\,\mathbf{t}_{n+\frac{1}{2}} + \mathbf{H}_u\,\mathbf{h}_{n+\frac{1}{2}} + \mathbf{M}\,\mathbf{b}_{n+\frac{1}{2}}
\end{aligned}
\tag{93}
$$

by taking into account the Eq. (89) and (90). Accordingly, we arrive at exact algorithmic total energy conservation in the sense of Eq. (82), if for the matrix continuum $\mathscr{B}^M_0$ the constraint

$$
\mathcal{G}_M(\tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}}) := W_{M_{n+1}} - W_{M_n} - \left[\mathrm{D}W_M(\tilde{\mathbf{C}}_{n+\frac{1}{2}};\boldsymbol{\kappa}_{0_M}) + \frac{1}{2}\,\tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}}\right] : \left[\mathbf{C}_{n+1} - \mathbf{C}_n\right] = 0 \tag{94}
$$

is fulfilled, and *if the relation*

$$\begin{aligned}
\int_{\mathscr{B}_0} \mathbf{v}_{n+\frac{1}{2}} \cdot \mathbf{B}^T_{n+\frac{1}{2}} \left[2\,\mathrm{D}\hat{W}_F(\tilde{C}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_{0_F}) + \tilde{S}_{n+\frac{1}{2}} \right] \mathbf{A}_0 \,\mathrm{d}V &= \\
\int_{\mathscr{B}_0} \frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{h_n} \cdot \mathbf{B}^T_{n+\frac{1}{2}} \left[2\,\mathrm{D}\hat{W}_F(\tilde{C}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_{0_F}) + \tilde{S}_{n+\frac{1}{2}} \right] \mathbf{A}_0 \,\mathrm{d}V &= \\
\int_{\mathscr{B}_0} \left[\mathrm{D}\hat{W}_F(\tilde{C}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_{0_F}) + \frac{\tilde{S}_{n+\frac{1}{2}}}{2} \right] \mathbf{A}_0 : 2\,\mathbf{B}_{n+\frac{1}{2}} \left[\mathbf{u}_{n+1} - \mathbf{u}_n \right] \mathrm{d}V &= \int_{\mathscr{B}_0} \left[\hat{W}_{F_{n+1}} - \hat{W}_{F_n} \right] \mathrm{d}V \\
\left[\mathrm{D}\hat{W}_F(\tilde{C}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_{0_F}) + \frac{\tilde{S}_{n+\frac{1}{2}}}{2} \right] \mathbf{A}_0 : \left[\mathbf{C}_{n+1} - \mathbf{C}_n \right] &= \hat{W}_{F_{n+1}} - \hat{W}_{F_n} \qquad (95)
\end{aligned}$$

or

$$\mathcal{G}_F(\tilde{S}_{F_{n+\frac{1}{2}}}) := \hat{W}_{F_{n+1}} - \hat{W}_{F_n} - \left[\mathrm{D}\hat{W}_F(\tilde{C}_{F_{n+\frac{1}{2}}}; \boldsymbol{\kappa}_{0_F}) + \frac{\tilde{S}_{F_{n+\frac{1}{2}}}}{2} \right] \left[C_{F_{n+1}} - C_{F_n} \right] = 0 \qquad (96)$$

is fulfilled. According to Eq. (81), the superimposed stress tensor $\mathbf{S}_M$ for the matrix continuum is given by

$$\tilde{\mathbf{S}}_{M_{n+\frac{1}{2}}} = 2\,\frac{\mathcal{G}_M(\mathbf{O})}{\tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \left[\mathbf{C}_{n+1} - \mathbf{C}_n \right] : \left[\mathbf{C}_{n+1} - \mathbf{C}_n \right] \tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}}} \tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \left[\mathbf{C}_{n+1} - \mathbf{C}_n \right] \tilde{\mathbf{C}}^{-1}_{n+\frac{1}{2}} \qquad (97)$$

The superimposed scalar stress field $\tilde{S}_F$ in the equation of motion is defined such that we have to take into account the identity

$$\frac{\hat{W}_{F_{n+1}} - \hat{W}_{F_n}}{C_{F_{n+1}} - C_{F_n}} \mathbf{A}_0 = \left[\mathrm{D}\hat{W}_F(\tilde{C}_{F_{n+\frac{1}{2}}}; \boldsymbol{\kappa}_{0_F}) + \frac{\tilde{S}_{F_{n+\frac{1}{2}}}}{2} \right] \mathbf{A}_0 \qquad (98)$$

which eliminates completely a FRÉCHET derivative $\mathrm{D}\hat{W}_F$ of the strain energy $\hat{W}_F$ in the equation of motion.

4 NUMERICAL EXAMPLE

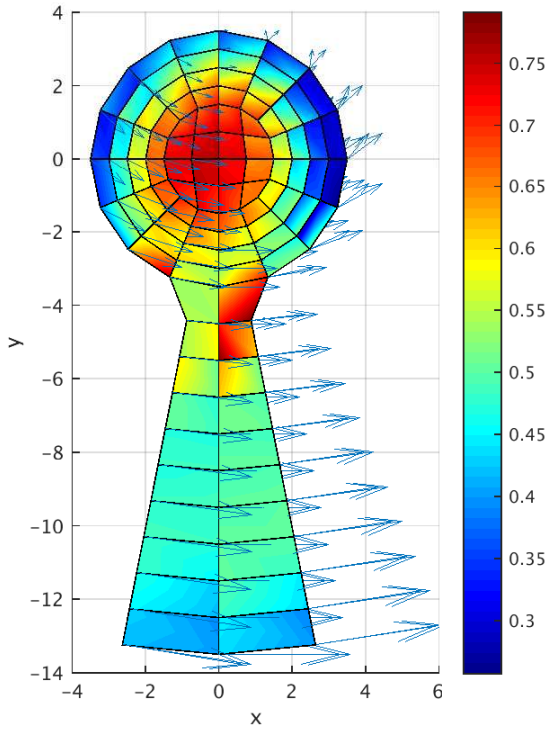
As numerical example, we consider a transversely isotropic blade discretized in space by eight-node brick elements. In the initial configuration, the center of the blade's hub is positioned in the origin of the three-dimensional EUCLIDEAN space (see Fig. 2). The material is described by the *unpartitioned* strain energy function

$$\hat{W}(I_1, I_2, I_3, I_4; \boldsymbol{\kappa}_0) = \hat{W}^{\mathrm{isotr}}(I_1, I_2, I_3; c_1, c_2, c_3) + \hat{W}^{\mathrm{aniso}}(I_3, I_4; c_3, c_4) \qquad (99)$$

with the functions

$$\hat{W}^{\mathrm{isotr}}(I_1, I_2, I_3; c_1, c_2, c_3) = \frac{c_1}{2}(I_3^{-\frac{1}{3}} I_1 - 3) + c_2(I_3 - 1)^2 \qquad (100)$$

$$\hat{W}^{\mathrm{aniso}}(I_3, I_4; c_3, c_4) = \frac{c_3}{2c_4} \exp\left[c_4 \left(I_3^{-1/3} I_4 - 1 \right)^2 - 1 \right] \qquad (101)$$



Simulation parameter		
spatial mesh	Eight-node bricks	
element number	n_{el}	100
node number	n_{no}	238
mass density	ρ_0	1
strain energy	$W = W^{\text{isotr}} + W^{\text{aniso}}$	
soft material	c_1	300
	c_2	100
	c_3	240
	c_4	80
stiff material	c_1	3000
	c_2	1000
	c_3	2400
	c_4	800
fiber vector	$\mathbf{a}_0$	$[1,1,1]^T/\sqrt{3}$
initial velocity	$\mathbf{v}_0^A = \mathbf{v}_T + \boldsymbol{\omega}_0 \times \mathbf{q}_0^A$	
velocity vector	$\mathbf{v}_T$	$[2,0,-0.1]^T$
angular velocity	$\boldsymbol{\omega}_0$	$[0,0.7,0.7]^T$
NEWTON tolerance	tol	10^{-8}

Figure 2: Left: Initital configuration $\mathcal{B}_0$ of the fiber-reinforced blade. The colours indicate the VON MISES stress at the temporal Gauss point $t_{0+1/2}$ determined by the cG(1) method in the non-stiff case. The arrows show the initial velocity field of the free flight. Right: Simulation parameter of the motion and of the algorithm.

the parameters c_1, c_2, c_3 and the dimensionless parameter c_4 (compare Reference [14]). The applied material parameter values are summarized in Fig. 2 on the right. We distinguish between soft and stiff material. The blade are in free flight due to its initial translational velocity field and its initial angular velocity field (see Fig. 2). We compare two numerical methods:

(i) the variational consistent discrete method presented above, referred to as eG(1) method in the following, and

(ii) the continuous Galerkin cG(1) method or midpoint rule, respectively, given by Eq. (68) and the corresponding second PIOLA-KIRCHHOFF stress tensor

$$\mathbf{S}^{\text{mid}}_{n+\frac{1}{2}} = 2\,\mathrm{D}W(\mathbf{F}^T_{n+\frac{1}{2}}\mathbf{F}_{n+\frac{1}{2}}; \mathbf{A}_0, \boldsymbol{\kappa}_0) \tag{102}$$

based on a *temporally discontinuous 'strain' approximation*.

Considering soft material, both methods show similar current configurations for a moderate constant time step size. Therefore, we show only the motion of the cG(1) method in Fig. 3. But, by changing the time step size during the simulation, the NEWTON-RAPHSON method in the time loop of the cG(1) method aborts after some time steps. This can be shown by plotting the total energy of the blade versus time (see Fig. 4 on the left). In contrast to the eG(1) method, the cG(1) method shows an oscillating total energy with an energy blow-up after the time step size change. Considering stiff material, no time step size change is necessary for illustrating the unstable behaviour of the cG(1) method in contrast to the eG(1) method (see Fig. 4 on the right).

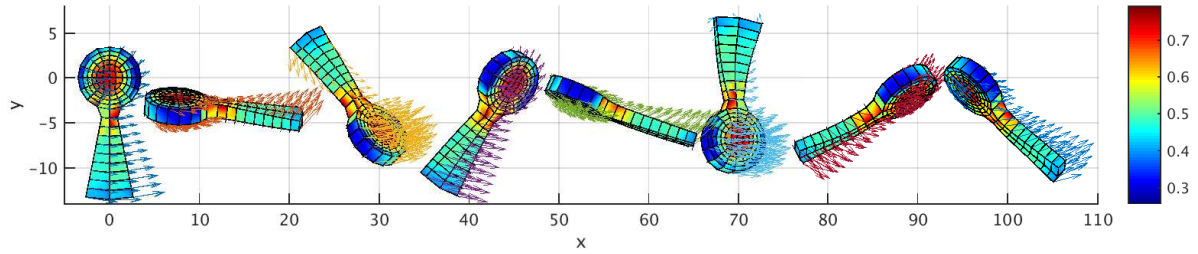


Figure 3: Current configurations $\mathcal{B}_{t_n}$ of the blade with the non-stiff material determined by the cG(1) method, starting at $t_0 = 0$ on the left. The colour indicates the VON MISES stress at the temporal Gauss point $t_{n+1/2}$. The arrows shows the current Lagrangian velocity field.

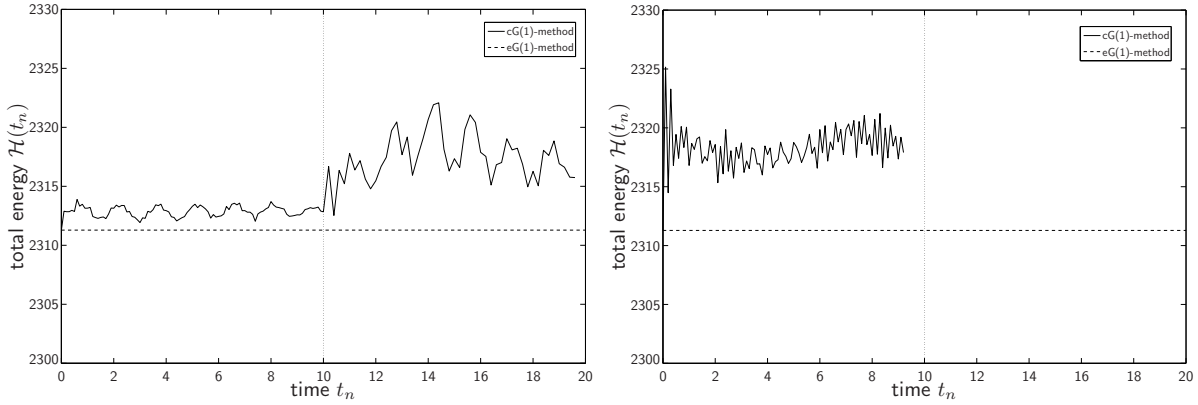


Figure 4: Comparison of the total energy $\mathcal{H}_n$ of the cG(1)-method and the eG(1)-method, respectively, using the soft material (left) and stiff material (right). The time step size is 0.1 for $t \leq 10$ and 0.2 for $t > 10$.

5 SUMMARY

In this paper, we consider transversely isotropic materials from two perspectives. We examine

1. the general case of formally one free energy function with no separation of tensor invariants (*unpartitioned free energy function*), and

2. a partition of the free energy function into two separate terms corresponding to isotropic and anisotropic invariants, respectively (*partitioned free energy function*).

The reason is that the fundamental theorem of calculus corresponding to partitioned free energy functions can be split into separate equations as it is well-known from the kinetic and potential energy of natural systems with respect to inertial reference frames. As the fundamental theorem of calculus serves as a design criterion for energy consistent time stepping algorithms, a separation into two criteria therefore allows to modify the algorithm in a more targeted manner. Further, already implemented energy consistent time stepping algorithms for isotropic materials could be extented rather than modified to transversely isotropic materials or composite materials with more than one family of fibers.

In this work, we start the design of energy consistent time stepping algorithms by *discretizing a mixed variational principle*, because we aim at a unified design procedure for these important algorithms. Such an unified procedure already exists for momentum consistent time stepping

algorithms leading to the so-called variational integrators [5]. In this variational framework, the consistency with the momentum balances does not depend on the numerical quadrature as in usual finite difference or GALERKIN-based schemes. Energy consistent time stepping algorithms, however, are hitherto designed for specific mechanical problems [2, 3, 8, 11, 13], although the corresponding design procedures exhibit many common features. In each of these references, the discrete total energy balance is satisfied by introducing

- temporally *continuous* approximations of the independent argument tensors of the total energy, together with
- temporally discontinuous superimposed work conjugate tensor fields emanating from spatially local formulations of the fundamental theorem of calculus in time.

But, the application of this concept for designing higher-order accurate time stepping schemes as generalisation of existing second-order accurate schemes raises questions in the details [2, 3], starting with the temporally continuous approximation of the independent argument tensors of the *strain energy*. In these references, a mixture of 'strain' tensor approximations has to be used for satisfying energy consistency, which is not obvious from a physical perspective. These problems and the need of a unified framework have led to the herewith presented idea of discretizing a mixed variational principle, providing

1. a *proof of existing adhoc time approximations*, and
2. *new higher-order accurate* energy consistent time approximations (see Appendix).

The first adhoc time approximation is the midpoint evaluation of the GREEN-LAGRANGE strain tensor in Reference [6], or later called *assumed 'strain' approximation* in Reference [15], respectively. This approximation is often used as a physically modivated assumption (frame-invariance of discrete strains), or as an inherent part of energy consistent discrete gradients of strain energy functions in finite difference schemes [8]. In Reference [6], this temporally continuous approximation of the GREEN-LAGRANGE strain tensor is modivated by the exact quadrature of the approximated strain energy function pertaining to the quadratic SAINT-VENANT KIRCHHOFF model. The discrete total energy balance corresponding to this strain energy function is therefore fulfilled without a superimposed stress field, or in other words, a superimposed stress field vanish for this strain energy function as in the temporally continuous equations of motion. Hence, inspired by three-field variational functionals of EAS methods [16], we here introduce a temporally continuous strain tensor with the corresponding natural ordinary differential equation in time by means of a mixed variational principle. In this way, we actually prove the *variational consistency of the assumed 'strain' approximation* for well-known second-order accurate methods, and derive a *new* assumed 'strain' approximation for higher-order accurate schemes which avoids unphysical approximation mixtures as in Reference [3] (see Appendix).

The second adhoc time approximation is the superimposed stress tensor in References [2, 3], based on the well-known discrete gradient in Reference [17]. This superimposed stress tensor is derived from a constraint variational problem at each point $\mathbf{X} \in \mathscr{B}_0$ in References [3], which therefore provides a proof of the uniqueness of this superimposed stress tensor. However, this variational problem is *not physically modivated* and therefore not invariant with respect to a push-forward in the spatial configuration $\mathcal{B}_t$. This problem has been caused by not using a coordinate free and metric independent geometric formulation of continuum mechanics. In fact, the Euclidean metric δ_{AB} is assumed in the reference configuration from the outset. Therefore,

in this paper, we arrive at the right CAUCHY-GREEN tensor as metric tensor by using a covariant tensor formulation. The obtained constraint variational problem is therefore invariant under a push-forward in the spatial configuration $\mathcal{B}_t$, and leads to an equivalent variational problem with respect to the KIRCHHOFF stress tensor. As special case for a linear approximation in time, we obtain the superimposed stress field in Reference [8]. Note, however, that a further improvement of the computational performance in comparison to the superimposed stress tensor in References [3] is not recognized by considering free flights of stiff materials. *But if the algorithm has to be pushed forward in a computational more efficient spatial setting, the new superimposed stress field is necessary.* Further note that the constraint variational problems are still separate variational problems of parameters of the mixed variational principle.

A APPENDIX

In this appendix, we show an interesting consequence of the above theory for the new assumed 'strain' approximation of higher-order accurate time integration schemes, i.e. schemes which take into account inner time points $t_{n+\alpha_i}$, $\alpha_i \in]0,1[$, beside the time step boundary points t_n and t_{n+1}. This inner time points are usually equidistant distributed over the time step (see Table 1). Here, we have to start in the unpartitioned case with the discrete principle

$$\sum_{n=0}^{N-1}\sum_{i=1}^{k} \delta_* \dot{\mathcal{H}}(\dot{\mathbf{u}}_h^n(\xi_i), \dot{\mathbf{v}}_h^n(\xi_i), \dot{\mathbf{p}}_h^n(\xi_i), \dot{\tilde{\mathbf{C}}}_h^n(\xi_i), \mathbf{S}_h^n(\xi_i); \\ \rho_0, \mathbf{A}_0, \boldsymbol{\kappa}_0, \mathbf{b}_h^n(\xi_i), \mathbf{t}_h^n(\xi_i), \dot{\tilde{\mathbf{u}}}_h^n(\xi_i), \tilde{\mathbf{S}}_h^n(\xi_i))\, w_i\, h_n = 0 \tag{103}$$

where ξ_i, w_i, $i = 1, \ldots, k$, denote the quadrature points and weights, respectively, and k the degree of the shape functions $M_j(\alpha)$, $j = 1, \ldots, k+1$ in time. Usually, the Lagrangian shape functions and the Gaussian quadrature rules are used (see Table 1 and Table 2, respectively). According to this principle, we obtain the weak equation

$$\sum_{i=1}^{k} \int_{\mathcal{B}_0} \delta_* \mathbf{S}_h^n(\xi_i) : \left[\frac{\mathrm{d}\tilde{\mathbf{C}}_h^n(\xi_i)}{\mathrm{d}\alpha} - \overset{\circ}{\mathbf{C}}(\overset{\circ}{\mathbf{u}}{}_h^n(\xi_i)) \right] w_i \, \mathrm{d}V = 0 \tag{104}$$

with the assumed 'strain' approximation

$$\frac{\mathrm{d}\tilde{\mathbf{C}}_h^n(\alpha)}{\mathrm{d}\alpha} = \sum_{j=1}^{k+1} \overset{\circ}{M}_{j+1}(\alpha)\, \mathbf{C}_j^n \equiv \sum_{i=1}^{k} \tilde{M}_i(\alpha)\, \tilde{\mathbf{C}}_i^n \tag{105}$$

and the shorthand notation $\overset{\circ}{(\bullet)}$ for the differentiation with respect to α. The tensors $\mathbf{C}_j^n$ and $\tilde{\mathbf{C}}_i^n$ designate the independend nodal values of the assumed 'strain' approximation and its derivative, respectively, at the corresponding time points α_i and $\tilde{\alpha}_i$ (see Table 1). Having again an elimination of the assumed 'strain' field in mind, we arrive at the spatially local relation

$$\sum_{j=1}^{k+1} \overset{\circ}{M}_{j+1}(\xi_i)\, \mathbf{C}_j^n - \overset{\circ}{\mathbf{C}}(\overset{\circ}{\mathbf{u}}{}_h^n(\xi_i)) = \mathbf{O} \qquad (i = 1, \ldots, k) \tag{106}$$

After further algebraic transformations, the unkonwn nodal values $\mathbf{C}_l^n$, $l = 2, \ldots, k$, becomes

$$\mathbf{C}_l^n := \sum_{i=1}^{k} m_{li}\, \overset{\circ}{\mathbf{C}}(\overset{\circ}{\mathbf{u}}{}_h^n(\xi_i)) + \mathbf{C}_1^n \tag{107}$$

by taking into account the initial condition $\mathbf{C}_1^n := (\mathbf{F}_1^n)^T\mathbf{F}_1^n \equiv \mathbf{C}_n$. The coefficients A_{li} are the components of the $k \times k$ matrix

$$\mathbf{m} = \begin{bmatrix} \overset{\circ}{M}_2(\xi_1) & \dots & \overset{\circ}{M}_{k+1}(\xi_1) \\ \vdots & \dots & \vdots \\ \overset{\circ}{M}_2(\xi_k) & \dots & \overset{\circ}{M}_{k+1}(\xi_k) \end{bmatrix}^{-1} \tag{108}$$

In the case of linear time approximations ($k = 1$), these relations lead to the nodal values

$$\mathbf{C}_n \equiv \mathbf{C}_1^n = (\mathbf{F}_1^n)^T\mathbf{F}_1^n \qquad\qquad \mathbf{C}_{n+1} \equiv \mathbf{C}_2^n = (\mathbf{F}_2^n)^T\mathbf{F}_2^n \tag{109}$$

(compare Eq. (61). For quadratic time approximations ($k = 2$), we arrive at the nodal values

$$\mathbf{C}_n \equiv \mathbf{C}_1^n := (\mathbf{F}_1^n)^T\mathbf{F}_1^n \tag{110}$$

$$\mathbf{C}_2^n := \frac{1}{3}\left[\frac{\mathbf{F}_1^n + \mathbf{F}_3^n}{2} - \mathbf{F}_2^n\right]^T\left[\frac{\mathbf{F}_1^n + \mathbf{F}_3^n}{2} - \mathbf{F}_2^n\right] + (\mathbf{F}_2^n)^T\mathbf{F}_2^n \tag{111}$$

$$\mathbf{C}_{n+1} \equiv \mathbf{C}_3^n := (\mathbf{F}_3^n)^T\mathbf{F}_3^n \tag{112}$$

For higher degrees of shape functions, we obtain analogous results. Accordingly, the nodal values at the time step boundaries t_n and t_{n+1} are solely determined by $\mathbf{u}_n$ and $\mathbf{u}_{n+1}$, respectively, but the nodal values at the inner time points depends on the *displacements of all time points*. *This is in contrast* to the (frame-indifferent) assumed 'strain' approximation $\bar{\mathbf{C}}_h^n(\alpha)$ defined in Reference [3] by the extrapolation

$$\bar{\mathbf{C}}_h^n(\alpha) = \sum_{j=1}^{k+1} M_j(\alpha)\,(\mathbf{F}_j^n)^T\mathbf{F}_j^n \tag{113}$$

of the formula $\tilde{\mathbf{C}}_h^n(\alpha)$ for linear time approximations ($k = 1$) known from Reference [15]. As the first term in Eq. (111) is unknowingly neglected in Eq. (113), the authors of Reference [3] have to introduce a mixture of time approximations in the superimposed stress tensor for energy consistency.

On the other hand, looking at Eq. (111), we may recognize a possibility to simplify the relations for the nodal values $\mathbf{C}_2, \dots \mathbf{C}_{k-1}$ by introducing an *assumed deformation gradient field*, which for $k = 2$ takes the form

$$\tilde{\mathbf{F}}_h^n(\alpha) = M_1(\alpha)\,\mathbf{F}_1^n + M_2(\alpha)\,\hat{\mathbf{F}}_2^n + M_3(\alpha)\,\mathbf{F}_3^n \tag{114}$$

with the nodal value

$$\hat{\mathbf{F}}_2^n := \frac{\mathbf{F}_1^n + \mathbf{F}_3^n}{2} \neq \nabla\mathbf{u}_2^n + \mathbf{I} \equiv \nabla\mathbf{u}_{n+\frac{1}{2}} + \mathbf{I} \tag{115}$$

Thus, the approximated displacement field $\mathbf{u}_h^n(\alpha)$ is here connected to the deformation gradient field only at the boundaries of the time step $[t_n, t_{n+1}]$ with the linear approximation

$$\tilde{\mathbf{F}}_h^n(\alpha) = \tilde{M}_1(\alpha)\,\mathbf{F}_1^n + \tilde{M}_2(\alpha)\,\mathbf{F}_3^n \tag{116}$$

The corresponding assumed 'strain' field reads

$$\tilde{\mathbf{C}}_h^n(\alpha) = M_1(\alpha)\,(\mathbf{F}_1^n)^T\mathbf{F}_1^n + M_2(\alpha)\,(\hat{\mathbf{F}}_2^n)^T\hat{\mathbf{F}}_2^n + M_3(\alpha)\,(\mathbf{F}_3^n)^T\mathbf{F}_3^n \tag{117}$$

$$= (\alpha-1)^2\,(\mathbf{F}_1^n)^T\mathbf{F}_1^n + \alpha(\alpha-1)\left[(\mathbf{F}_1^n)^T\mathbf{F}_3^n + (\mathbf{F}_3^n)^T\mathbf{F}_1^n\right] + \alpha^2\,(\mathbf{F}_3^n)^T\mathbf{F}_3^n \tag{118}$$

However, we have to examine this possibility with respect to the accuracy order of the resulting time integration scheme. Further, a linear time approximation of the deformation gradient field and a quadratic time approximation of the displacement field seems to be inconsistent, in contrast to an analogous space approximation [14].

Nevertheless, the accuracy order of the time integration scheme corresponding to the approximation $\tilde{\mathbf{C}}_h^n(\alpha)$ with the nodal values in Eq. (107) is guaranteed for shape functions of degree up to four ($k = 4$). We have even implemented this approximation in the finite element code associated with the thermo-mechanical problem in Reference [13], and obtained the numerical results in Fig. 5. This approximation is therefore even recommended for mechanically coupled problems. The higher-order approximation of the superimposed stress tensor of the unpartitioned strain energy function at the temporal Gauss point is given by

$$\tilde{\mathbf{S}}_h^n(\xi_i) := 2\,\frac{\mathcal{G}(\mathbf{O})}{\sum_{l=1}^{k}[\tilde{\mathbf{C}}_h^n(\xi_l)]^{-1}\,\overset{\circ}{\overline{\tilde{\mathbf{C}}_h^n}}(\xi_l) : \overset{\circ}{\overline{\tilde{\mathbf{C}}_h^n}}(\xi_l)\,[\tilde{\mathbf{C}}_h^n(\xi_l)]^{-1}\,w_l}\,[\tilde{\mathbf{C}}_h^n(\xi_i)]^{-1}\,\overset{\circ}{\overline{\tilde{\mathbf{C}}_h^n}}(\xi_i)\,[\tilde{\mathbf{C}}_h^n(\xi_i)]^{-1} \tag{119}$$

with

$$\mathcal{G}(\mathbf{O}) := W_{n+1} - W_n - \sum_{l=1}^{k} \mathrm{D}W(\tilde{\mathbf{C}}_h^n(\xi_l); \mathbf{A}_0, \boldsymbol{\kappa}_0) : \overset{\circ}{\overline{\tilde{\mathbf{C}}_h^n}}(\xi_l)\,w_l = 0 \tag{120}$$

The superimposed stress $\tilde{S}_F$ corresponding to the partitioned strain energy is due to the scalar-valued argument $\tilde{C}_F$ analogous to the dynamical problem of a particle system in Reference [3] (compare Eq. (98) with the case $k = 1$). Hence, we obtain the relation

$$\tilde{S}_{F_h}^n(\xi_i) := 2\,\frac{\mathcal{G}_F(0)}{\sum_{l=1}^{k}\overset{\circ}{\overline{\tilde{C}_{F_h}^n}}(\xi_l)\;\overset{\circ}{\overline{\tilde{C}_{F_h}^n}}(\xi_l)\,w_l}\;\overset{\circ}{\overline{\tilde{C}_{F_h}^n}}(\xi_i) \tag{121}$$

with

$$\mathcal{G}_F(0) := \hat{W}_{F_{n+1}} - \hat{W}_{F_n} - \sum_{l=1}^{k} \mathrm{D}\hat{W}_F(\tilde{C}_{F_h}^n(\xi_l); \boldsymbol{\kappa}_{0_F})\;\overset{\circ}{\overline{\tilde{C}_{F_h}^n}}(\xi_l)\,w_l = 0 \tag{122}$$

and the fiber assumed strain approximation

$$\tilde{C}_{F_h}^n(\alpha) = \sum_{j=1}^{k+1} M_{j+1}(\alpha)\,C_{F_j}^n \tag{123}$$

where the nodal values $C_{F_l}^n$, $l = 2, \ldots, k$, take the form

$$C_{F_l}^n := \sum_{i=1}^{k} A_{li}\;\overset{\circ}{\mathbf{C}}(\overset{\circ}{\mathbf{u}}_h^n(\xi_i)) : \mathbf{A}_0 + \mathbf{C}_1^n : \mathbf{A}_0 \tag{124}$$

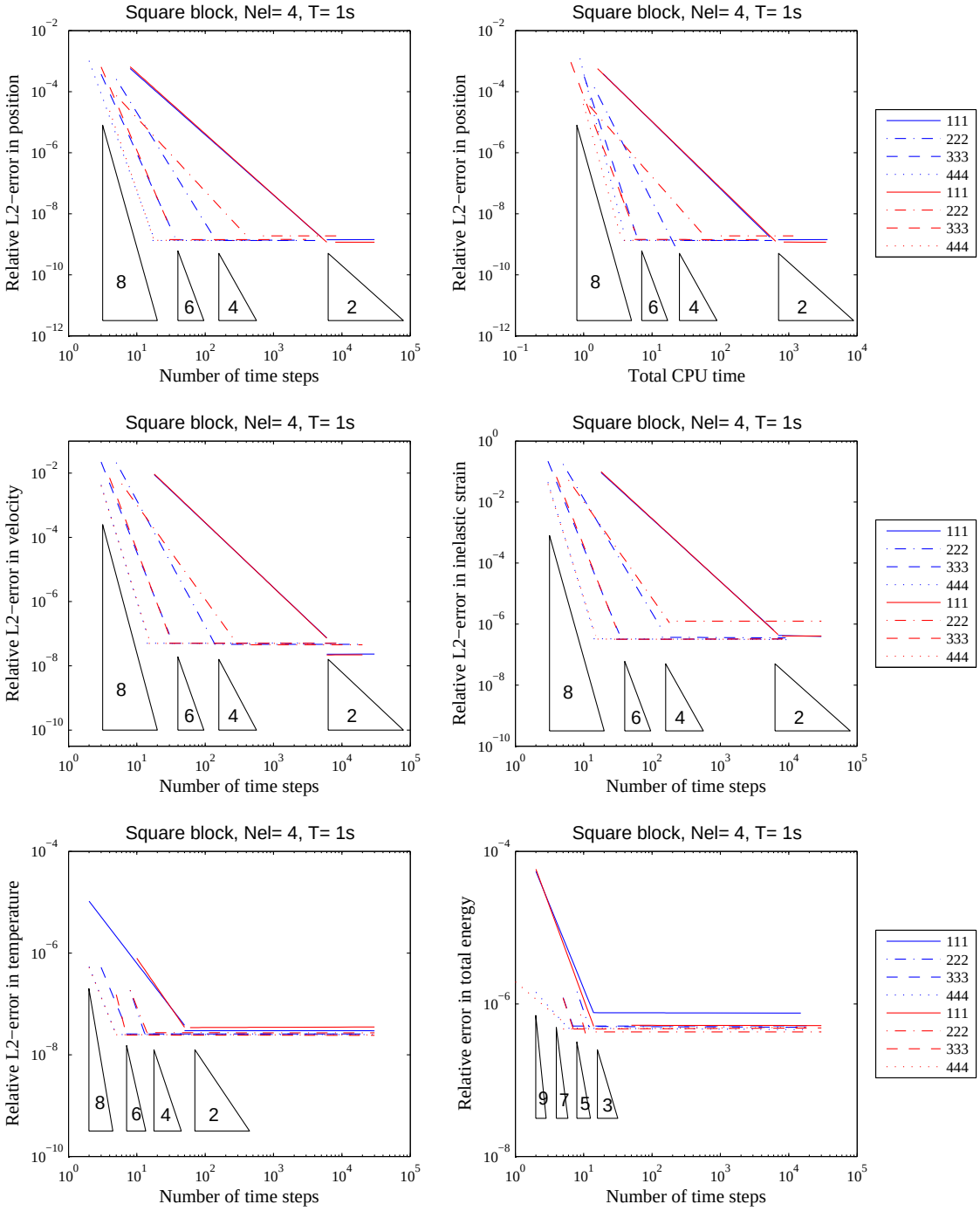


Figure 5: Accuracy orders of the energy consistent time stepping scheme presented above using the new assumed 'strain' approximation, determined with a flying stiff square discretized by four four-node quadrilateral elements. We investigated the thermo-mechanically problem in Reference [13]. The plots show the relative L_2 errors at the final simulation time $T = 1$. For shape functions of degree k in the mechanical and thermal fields (labels 'kkk'), we obtain the accuracy order $2k$. The order of the total energy has the order $2k + 1$, because the temperature is calculated with an energy consistent discontinuous GALERKIN method. Through the strong thermo-mechanical coupling, the temperature shows the same order as the displacements or current positions, respectively.

Table 1: Lagrangian shape functions in time of degree k and $k-1$ with respect to the parent domain $[0,1]$.

k	$M_j(\alpha)$	α_j	$\tilde{M}_i(\alpha)$	$\tilde{\alpha}_i$
1	$1-\alpha$ α	0 1	1	1
2	$(2\alpha-1)(\alpha-1)$ $-4\alpha\,(\alpha-1)$ $(2\alpha-1)\,\alpha$	0 $\frac{1}{2}$ 1	$1-\alpha$ α	0 1
3	$-\frac{9}{2}\,(\alpha-\frac{1}{3})(\alpha-\frac{2}{3})(\alpha-1)$ $\frac{27}{2}\,(\alpha-\frac{2}{3})(\alpha-1)\,\alpha$ $-\frac{27}{2}\,(\alpha-\frac{1}{3})(\alpha-1)\,\alpha$ $\frac{9}{2}\,(\alpha-\frac{1}{3})(\alpha-\frac{2}{3})\,\alpha$	0 $\frac{1}{3}$ $\frac{2}{3}$ 1	$(2\alpha-1)(\alpha-1)$ $-4\alpha\,(\alpha-1)$ $(2\alpha-1)\,\alpha$	0 $\frac{1}{2}$ 1
4	$\frac{32}{3}\,(\alpha-\frac{1}{4})(\alpha-\frac{1}{2})(\alpha-\frac{3}{4})(\alpha-1)$ $-\frac{128}{3}\,(\alpha-\frac{1}{2})(\alpha-\frac{3}{4})(\alpha-1)\alpha$ $64\,(\alpha-\frac{1}{4})(\alpha-\frac{3}{4})(\alpha-1)\alpha$ $-\frac{128}{3}\,(\alpha-\frac{1}{4})(\alpha-\frac{1}{2})(\alpha-1)\alpha$ $\frac{32}{3}\,(\alpha-\frac{1}{4})(\alpha-\frac{1}{2})(\alpha-\frac{3}{4})\alpha$	0 $\frac{1}{4}$ $\frac{1}{2}$ $\frac{3}{4}$ 1	$-\frac{9}{2}\,(\alpha-\frac{1}{3})(\alpha-\frac{2}{3})(\alpha-1)$ $\frac{27}{2}\,(\alpha-\frac{2}{3})(\alpha-1)\,\alpha$ $-\frac{27}{2}\,(\alpha-\frac{1}{3})(\alpha-1)\,\alpha$ $\frac{9}{2}\,(\alpha-\frac{1}{3})(\alpha-\frac{2}{3})\,\alpha$	0 $\frac{1}{3}$ $\frac{2}{3}$ 1

Table 2: Gaussian quadrature with N_{qp} Gauss points with respect to the temporal parent domain $[0,1]$.

N_{qp}	ξ_l	w_l
1	$1/2$	1
2	$(1-1/\sqrt{3})/2$ $(1+1/\sqrt{3})/2$	$1/2$ $1/2$
3	$(1-\sqrt{3/5})/2$ $1/2$ $(1+\sqrt{3/5})/2$	$5/18$ $4/9$ $5/18$
4	$(1-\sqrt{3/7+2\sqrt{6/5/7}})/2$ $(1-\sqrt{3/7-2\sqrt{6/5/7}})/2$ $(1+\sqrt{3/7-2\sqrt{6/5/7}})/2$ $(1+\sqrt{3/7+2\sqrt{6/5/7}})/2$	$(3-\sqrt{5/6})/12$ $1/2-(3-\sqrt{5/6})/12$ $1/2-(3-\sqrt{5/6})/12$ $(3-\sqrt{5/6})/12$

REFERENCES

[1] Kaliske M (2000) A formulation of elasticity and viscoelasticity for fibre reinforced material at small and finite strains. Comput Methods Appl Mech Engrg 185:225–243.

[2] Erler N and Groß M (2015) Energy-momentum conserving higher-order time integration of nonlinear dynamics of finite elastic fiber-reinforced continua. Computational Mechanics 55(5):921–942.

[3] Groß M and Betsch P and Steinmann P (2005) Conservation properties of a time FE method. Part IV: higher order energy and momentum conserving schemes. Int J Numer Methods Engng 63:1849–1897.

[4] Schröder J and Neff P and Balzani D (2005) A variational approach for materially stable anisotropic hyperelasticity. Int J Sol Struc 42:4352–4371.

[5] Schlögl T. and Leyendecker S. (2016) Electrostatic-viscoelastic finite element model of dielectric actuators. Comput Methods Appl Mech Engrg 299:421–439.

[6] Simo JC and Tarnow N (1992) The Discrete Energy-Momentum Method. Conserving Algorithms for Nonlinear Elastodynamics. Z angew Math Phys 43:757–792.

[7] Betsch P and Janz A (2016) An energymomentum consistent method for transient simulations with mixed finite elements developed in the framework of geometrically exact shells. Int J Numer Methods Engng, early view, DOI: 10.1002/nme.5217.

[8] Armero F and Zambrana-Rojas C (2007) Volume-preserving energy-momentum schemes for isochoric multiplicative plasticity. Comput Methods Appl Mech Engrg 196:4130–4159.

[9] Schröder B and Kuhl D (2015) Small strain plasticity: classical versus multifield formulation. Archive of Applied Mechanics 85(8):1127–1145.

[10] Schröder J, Wriggers P and Balzani D (2011) A new mixed finite element based on different approximations of the minors of deformation tensors. Comput Methods Appl Mech Engrg 200:3583–3600.

[11] Betsch P and Steinmann P (2000) Inherently Energy Conserving Time Finite Elements for Classical Mechanics. Journal of Computational Physics 160:88–116.

[12] Simo JC (1987) On a Fully Three-Dimensional Finite-Strain Viscoelastic Damage Model: Formulation and Computational Aspects. Comput Methods Appl Mech Engrg 60:153–173.

[13] Groß M and Betsch P (2011) Galerkin-based energy-momentum consistent time-stepping algorithms for classical nonlinear thermo-elastodynamics. Math Comp Sim 82:718–770.

[14] Schröder J, Neff P, Balzani D (2005) A variational approach for materially stable anisotropic hyperelasticity. Int J Solid Struct 42:4352–4371.

[15] Betsch P and Steinmann P (2001) Conservation properties of a time FE method. Part II: Time-Stepping Schemes for Nonlinear Elastodynamics. Int J Numer Methods Engng 50:1931–1955.

[16] Klinkel S and Wagner W (1997) A geometrical non-linear brick element based on the EAS-method. Int J Numer Methods Engng 40:4529–4545.

[17] Gonzalez O (2000) Exact Energy and Momentum Conserving Algorithms for General Models in Nonlinear Elasticity. Comput Methods Appl Mech Engrg 190:1763–1783.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

STRUCTURE-PRESERVING OPTIMAL CONTROL OF DISCRETE MECHANICAL SYSTEMS

Peter Betsch[1] **and Christian Becker**[1]

[1]Institute of Mechanics
Karlsruhe Institute of Technology (KIT)
Otto-Ammann-Platz 9
76131 Karlsruhe, Germany
{peter.betsch, christian.becker}@kit.edu

Keywords: Structure-Preserving Methods; Optimal Control; Noether's Theorem; Rotational symmetry; Generalized Momentum Maps

Abstract. *In this paper a structure-preserving direct method for the optimal control of mechanical systems is developed. The new method accommodates a large class of one-step integrators for the underlying state equations. The state equations under consideration govern the motion of affine Hamiltonian control systems. If the optimal control problem has symmetry, associated generalized momentum maps are conserved along an optimal path. This is in accordance with an extension of Noether's theorem to the realm of optimal control problems. In the present work we focus on optimal control problems with rotational symmetries. The newly proposed direct approach is capable of exactly conserving generalized momentum maps associated with rotational symmetries of the optimal control problem. This is true for a variety of one-step integrators used for the discretization of the state equations. Examples are the one-step theta method, a partitioned variant of the theta method, and energy-momentum consistent integrators. Numerical investigations confirm the theoretical findings.*

1 INTRODUCTION

The present work deals with the design of structure-preserving numerical methods for the solution of optimal control problems in mechanics. Although there exist many alternative numerical methods for the solution of optimal control problems (see, for example, the books by Agrawal & Fabien [1], Betts [4], Biegler [5] and Gerdts [14], and the survey articles by Binder et al. [6] und Diehl et al. [11]), little attention has been paid to structure-preservation on the level of the optimal control problem.

It is well-known that the Hamiltonian of autonomous optimal control problems is conserved along an optimal path (see, for example, Little & Pinch [23]). In addition to that, the optimal control problem may inherit symmetries from the underlying control system. Typically, in the context of mechanical control systems rotational symmetries are of paramount importance. Due to a generalization of Noether's theorem to optimal control (see Djukić [12] and Torres [28, 29]), symmetries of the optimal control problem lead to associated generalized momentum maps that are conserved along an optimal path. In the case of mechanical systems these symmetries have been explored to reduce the dimension of the optimal control problem (see Grizzle & Marcus [16] and van der Schaft [30]). However, to the best of our knowledge, symmetries and associated momentum maps have not been taken into account in the development of structure-preserving numerical methods for optimal control.

In contrast to that, structure-preserving numerical integrators originally developed in the framework of Hamiltonian systems with symmetry have been recently employed in the optimal control of mechanical systems. These integrators are capable of conserving important mechanical quantities in the forward dynamics of mechanical systems. In particular, energy-momentum integrators yield algorithmic conservation of energy and angular momentum in the case of Hamiltonian systems with rotational symmetry. Similarly, symplectic-momentum integrators are capable of conserving angular momentum and preserve the symplectic transformation property in phase space.

Structure-preserving numerical integrators can be directly used in optimal control problems by applying the direct transcription approach. For example, energy-momentum integrators have been used in the solution of optimal control problems by Bottasso & Croce [9], Betsch et al. [3] and Koch & Leyendecker [21]. Similarly, symplectic-momentum integrators have been applied in the framework of the direct approach to optimal control problems by Leyendecker et al. [22], Ober-Blöbaum et al. [26] and Bloch et al. [8].

In the present work we focus on mechanical optimal control problems with rotational symmetries. We propose a new direct approach to the design of structure-preserving schemes that are capable of conserving generalized momentum maps associated with rotational symmetries on the level of the optimal control problem. The newly proposed method accommodates a whole family of one-step integrators for the underlying state equations. Among those one-step integrators are the symplectic Euler methods, the implicit mid-point rule and 2nd-order accurate energy-momentum schemes. By design, the proposed method exactly conserves the generalized momentum maps associated with rotational symmetries of the optimal control problem, independent of the underlying integrator and the number of time steps used to resolve the time interval of interest.

An outline of the rest of the paper is as follows. In Section 2 we summarize the equations governing both the mechanical control systems and the optimal control problems under investigation. In particular, conditions for the presence of rotational symmetries are provided, the

form of the associated generalized momentum maps is addressed, and a proof of the Noether-type conservation law on the level of the optimal control problem is given. Section 3 contains the newly proposed structure-preserving method. The generation of specific optimal control schemes based on the choice of specific one-step integrators for the state equations is presented in Section 4. These schemes are further investigated in Section 5 with the numerical example of a nonlinear spring pendulum. Eventually, conclusions are drawn in Section 6.

2 GOVERNING EQUATIONS

In the present work we consider finite-dimensional mechanical control systems whose motion is governed by state equations of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{1}$$

Here, the vector of state coordinates $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p}) \in \mathcal{P} \equiv \mathbb{R}^{2n}$ contains the configuration coordinates $\boldsymbol{q} \in \mathcal{Q} \equiv \mathbb{R}^n$ and the momenta $\boldsymbol{p} \in \mathbb{R}^n$. Accordingly, the configuration space $\mathcal{Q}$ is assumed to be linear and the state space $\mathcal{P}$ coincides with the canonical phase space ($\mathcal{P} = T^*\mathcal{Q}$). The momentum vector can be written in the form $\boldsymbol{p} = \boldsymbol{M}\dot{\boldsymbol{q}}$, where $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is a non-singular mass matrix and $\dot{\boldsymbol{q}} = \frac{d}{dt}\boldsymbol{q} \in \mathbb{R}^n (= T_q\mathcal{Q})$ denotes the velocity vector. On the right-hand side of (1), $\boldsymbol{f} : \mathcal{P} \times \mathcal{U} \mapsto \mathbb{R}^{2n}$ is a continuously differentiable function and $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^m$ are the controls. In the present work we restrict ourselves to **affine Hamiltonian control systems** (see, for example, Bloch [7]), so that a more detailed form of the state equations (1) is given by

$$\begin{aligned} \dot{\boldsymbol{q}} &= \partial_{\boldsymbol{p}}\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) \\ \dot{\boldsymbol{p}} &= -\partial_{\boldsymbol{q}}\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) + \boldsymbol{B}(\boldsymbol{q})\boldsymbol{u} \end{aligned} \tag{2}$$

Here, $\mathcal{H} : \mathcal{P} \mapsto \mathbb{R}$ is the Hamiltonian of the mechanical system and $\boldsymbol{B}(\boldsymbol{q}) \in \mathbb{R}^{n\times m}$ is a control distribution matrix. In the uncontrolled case (i.e. if $\boldsymbol{u} = \boldsymbol{0}$), the Hamiltonian control system (2) reduces to a standard Hamiltonian system which we call the **underlying Hamiltonian system**. In the sequel we focus on natural mechanical systems with **separable** Hamiltonians of the form

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\boldsymbol{p} \cdot \boldsymbol{M}^{-1}\boldsymbol{p} + \mathcal{V}(\boldsymbol{q}) \tag{3}$$

where $\mathcal{V} : \mathcal{Q} \mapsto \mathbb{R}$ is the potential energy. In the following it often proofs convenient to use 'nodal' configuration vectors $\boldsymbol{q}_I \in \mathbb{R}^3$ and associated nodal momentum vectors $\boldsymbol{p}^I \in \mathbb{R}^3$, so that the state coordinates of the system are comprised of partitioned configuration and momentum vectors

$$\begin{aligned} \boldsymbol{q} &= \left(\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{n_n}\right) \in \mathbb{R}^{3n_n} \\ \boldsymbol{p} &= \left(\boldsymbol{p}^1, \boldsymbol{p}^2, \ldots, \boldsymbol{p}^{n_n}\right) \in \mathbb{R}^{3n_n} \end{aligned} \tag{4}$$

where n_n denotes the number of nodes. To summarize, in the present work we focus on mechanical control systems whose state equations are given by

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{f}^q(\boldsymbol{p}) = \boldsymbol{M}^{-1}\boldsymbol{p} \\ \dot{\boldsymbol{p}} &= \boldsymbol{f}_p(\boldsymbol{q}, \boldsymbol{u}) \end{aligned} \quad \text{or} \quad \begin{aligned} \dot{\boldsymbol{q}}_I &= \boldsymbol{f}^q_I(\boldsymbol{p}) = (M^{-1})_{IJ}\boldsymbol{p}^J \\ \dot{\boldsymbol{p}}^I &= \boldsymbol{f}^I_p(\boldsymbol{q}, \boldsymbol{u}) = -\partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) + \boldsymbol{f}^I_{\text{ext}}(\boldsymbol{q}, \boldsymbol{u}) \end{aligned} \tag{5}$$

Note that the control forces are contained in

$$\boldsymbol{f}^I_{\text{ext}}(\boldsymbol{q}, \boldsymbol{u}) = \boldsymbol{B}^I_j(\boldsymbol{q})u^j$$

In this connection $(M^{-1})_{IJ}$ denotes the components of $\boldsymbol{M}^{-1}$ corresponding to the nodes I and J. Throughout this work the summation convention applies to repeated lower case and upper case letters. The state equations (5) govern, for example, the controlled motion of (i) n_n particles in 3-space, (ii) semi-discrete systems arising from the discretization in space of nonlinear elastodynamics, and (iii) elastic Cosserat points.

2.1 Optimal control problem

Assume that on a time interval $\mathcal{I} = [0, T]$, the mechanical system under consideration is to be steered from an initial state $\boldsymbol{x}(0) = \overline{\boldsymbol{x}}_0 \in \mathcal{P}$ to a terminal state specified by the terminal constraint $\boldsymbol{g}_T(\boldsymbol{x}(T)) = \boldsymbol{0}$, $\boldsymbol{g}_T : \mathcal{P} \mapsto \mathbb{R}^{n_T}$, $n_T \leq 2n$, such that a specific cost functional is minimized. The corresponding **optimal control problem** is given by

$$J[\boldsymbol{x}(\cdot), \boldsymbol{u}(\cdot)] = K(\boldsymbol{x}(T)) + \int_0^T L(\boldsymbol{x}(t), \boldsymbol{u}(t))dt \longrightarrow \text{min}$$

subject to

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{x}(0) &= \overline{\boldsymbol{x}}_0 \\ \boldsymbol{0} &= \boldsymbol{g}_T(\boldsymbol{x}(T)) \\ \boldsymbol{u} &\in \mathcal{U} \end{aligned}$$

Here, $L : \mathcal{P} \times \mathcal{U} \mapsto \mathbb{R}$ denotes the density cost function and $K : \mathcal{P} \mapsto \mathbb{R}$ is the final cost function. Since we focus on structural properties of the optimal control problem, we make the simplifying assumption $\mathcal{U} \equiv \mathbb{R}^m$, that is, there are no constraints on the controls. In accordance with the Pontryagin maximum principle we introduce the Hamiltonian of the optimal control problem $H : \mathcal{P} \times \mathbb{R}^{2n} \times \mathcal{U} \mapsto \mathbb{R}$ given by

$$H(\boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{u}) = \boldsymbol{\psi} \cdot \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) - L(\boldsymbol{x}, \boldsymbol{u}) \tag{6}$$

Now necessary conditions for the optimality of a solution $(\boldsymbol{x}^*(t), \boldsymbol{u}^*(t))$ to the optimal control problem are given by

$$\dot{\boldsymbol{x}} = \partial_{\boldsymbol{\psi}} H(\boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{7a}$$

$$\dot{\boldsymbol{\psi}} = -\partial_{\boldsymbol{x}} H(\boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{u}) = \partial_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{u}) - \boldsymbol{\psi} \cdot \partial_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{7b}$$

$$\boldsymbol{0} = \partial_{\boldsymbol{u}} H(\boldsymbol{x}, \boldsymbol{\psi}, \boldsymbol{u}) = -\partial_{\boldsymbol{u}} L(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\psi} \cdot \partial_{\boldsymbol{u}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{7c}$$

together with the endpoint conditions

$$\boldsymbol{x}(0) = \overline{\boldsymbol{x}}_0 \tag{8a}$$

$$\boldsymbol{0} = \boldsymbol{g}_T(\boldsymbol{x}(T)) \tag{8b}$$

$$\boldsymbol{\psi}(T) = -DK(\boldsymbol{x}(T)) - D\boldsymbol{g}_T(\boldsymbol{x}(T))^T \boldsymbol{\eta} \tag{8c}$$

Accordingly, in addition to the state equations (7a), the necessary conditions consist of the adjoint equations (7b) along with the control equations (7c). The variables $\boldsymbol{\psi} \in \mathbb{R}^{2n}$ in (7b) are termed adjoint (or costate) variables. Moreover, $\boldsymbol{\eta} \in \mathbb{R}^{n_T}$ in (8c) are Lagrange multipliers associated with the terminal constraints (8b). Note that the endpoint conditions (8) consist of

a mixed set of initial and terminal conditions. Altogether the system of differential algebraic equations (7) together with the endpoint conditions (8) constitute a two-point boundary value problem.

In analogy to the partitioning of the state coordinates the adjoint variables can be arranged in partitioned form $\boldsymbol{\psi} = (\boldsymbol{\psi}_q, \boldsymbol{\psi}^p)$. Then the adjoint equations (7b) can be written as

$$\begin{aligned} \dot{\boldsymbol{\psi}}_q &= \partial_{\boldsymbol{q}} L\big((\boldsymbol{q},\boldsymbol{p}),\boldsymbol{u}\big) - \boldsymbol{\psi}^p \cdot \partial_{\boldsymbol{q}} \boldsymbol{f}_p(\boldsymbol{q},\boldsymbol{u}) \\ \dot{\boldsymbol{\psi}}^p &= \partial_{\boldsymbol{p}} L\big((\boldsymbol{q},\boldsymbol{p}),\boldsymbol{u}\big) - \boldsymbol{\psi}_q \cdot \partial_{\boldsymbol{p}} \boldsymbol{f}^q(\boldsymbol{p}) \end{aligned} \tag{9}$$

In addition to that, the control equations (7c) can be recast in the form

$$\boldsymbol{0} = \partial_{\boldsymbol{u}} L\big((\boldsymbol{q},\boldsymbol{p}),\boldsymbol{u}\big) - \boldsymbol{\psi}^p \cdot \partial_{\boldsymbol{u}} \boldsymbol{f}_p(\boldsymbol{q},\boldsymbol{u}) \tag{10}$$

Moreover, using the partitioned form of the state and adjoint variables, the optimal control Hamiltonian (6) may also be written in the form

$$H\left((\boldsymbol{q},\boldsymbol{p}),(\boldsymbol{\psi}_q,\boldsymbol{\psi}^p),\boldsymbol{u}\right) = \boldsymbol{\psi}_q \cdot \boldsymbol{f}^q(\boldsymbol{p}) + \boldsymbol{\psi}^p \cdot \boldsymbol{f}_p(\boldsymbol{q},\boldsymbol{u}) - L((\boldsymbol{q},\boldsymbol{p}),\boldsymbol{u}) \tag{11a}$$

$$= \boldsymbol{\psi}_q^I \cdot \boldsymbol{f}_I^q(\boldsymbol{p}) + \boldsymbol{\psi}_I^p \cdot \boldsymbol{f}_p^I(\boldsymbol{q},\boldsymbol{u}) - L((\boldsymbol{q},\boldsymbol{p}),\boldsymbol{u}) \tag{11b}$$

Note that this notation is consistent with the partitioned form of the state equations (5). In (11b) nodal quantities consistent with (4) have been used. In what follows these alternative partitioned forms will be employed occasionally.

2.2 Symmetries of the optimal control problem

We focus on rotational symmetries of a mechanical system related to a one-parameter family of rotation matrices $\boldsymbol{R}^s \in \mathrm{SO}(3)$, the special orthogonal group in 3-space. In particular, we consider superposed rotations of the form

$$\begin{aligned} \boldsymbol{q}_I^s &= \boldsymbol{R}^s \boldsymbol{q}_I \\ (\boldsymbol{p}^I)^s &= \boldsymbol{R}^s \boldsymbol{p}^I \end{aligned} \tag{12}$$

for $I = 1, \ldots, n_n$, and $\boldsymbol{R}^s = \exp(s\widehat{\boldsymbol{\xi}}) \in \mathrm{SO}(3)$, where the exponential map is given by the Rodrigues formula (see, for example, [25])

$$\exp(s\widehat{\boldsymbol{\xi}}) = \boldsymbol{I} + \frac{\sin(s\|\boldsymbol{\xi}\|)}{\|\boldsymbol{\xi}\|}\widehat{\boldsymbol{\xi}} + \frac{1}{2}\left[\frac{\sin(s\|\boldsymbol{\xi}\|/2)}{\|\boldsymbol{\xi}/2\|}\right]^2 \widehat{\boldsymbol{\xi}}^2 \tag{13}$$

Here, $\widehat{\boldsymbol{\xi}} \in \mathrm{so}(3)$ is a skew-symmetric tensor with associated axial vector $\boldsymbol{\xi} \in \mathbb{R}^3$. That is, $\widehat{\boldsymbol{\xi}}\boldsymbol{a} = \boldsymbol{\xi} \times \boldsymbol{a}$ for any $\boldsymbol{a} \in \mathbb{R}^3$. In the present context $\boldsymbol{\xi}$ is a fixed non-zero vector in $\mathbb{R}^3$. It can be easily verified that $\boldsymbol{R}^s = \exp(s\widehat{\boldsymbol{\xi}})$ satisfies

$$\boldsymbol{R}^{s=0} = \boldsymbol{I} \qquad \text{and} \qquad \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{R}^s = \widehat{\boldsymbol{\xi}}$$

Thus we have $\boldsymbol{q}_I^0 = \boldsymbol{q}_I$, $(\boldsymbol{p}^I)^0 = \boldsymbol{p}^I$, along with the infinitesimal generators

$$\begin{aligned} \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{q}_I^s &= \boldsymbol{\xi} \times \boldsymbol{q}_I \\ \left.\frac{d}{ds}\right|_{s=0} (\boldsymbol{p}^I)^s &= \boldsymbol{\xi} \times \boldsymbol{p}^I \end{aligned} \tag{14}$$

2.2.1 Symmetry of the mechanical control system

Let a one-parameter family of maps on $\mathcal{P}$ be given by (12). The mechanical control system (5) has rotational symmetry if there exist controls $\boldsymbol{u}^s(t) \in \mathcal{U}$, where $\boldsymbol{u}^0(t) = \boldsymbol{u}(t)$, such that

$$\begin{aligned} \boldsymbol{R}^s \boldsymbol{f}_I^q(\boldsymbol{p}) &= \boldsymbol{f}_I^q(\boldsymbol{R}^s \circ \boldsymbol{p}) \\ \boldsymbol{R}^s \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) &= \boldsymbol{f}_p^I(\boldsymbol{R}^s \circ \boldsymbol{q}, \boldsymbol{u}^s) \end{aligned} \tag{15}$$

where the shorthand notation

$$\begin{aligned} \boldsymbol{R}^s \circ \boldsymbol{q} &= \left(\boldsymbol{R}^s \boldsymbol{q}_1, \dots, \boldsymbol{R}^s \boldsymbol{q}_{n_n}\right) \\ \boldsymbol{R}^s \circ \boldsymbol{p} &= \left(\boldsymbol{R}^s \boldsymbol{p}^1, \dots, \boldsymbol{R}^s \boldsymbol{p}^{n_n}\right) \end{aligned}$$

has been introduced.

2.2.2 Symmetry of the optimal control problem

The optimal control problem has symmetry if, in addition to the symmetry of the mechanical control system, (15), the density cost function is invariant under superposed rotations. That is,

$$L\left(\boldsymbol{R}^s \circ \boldsymbol{q}, \boldsymbol{R}^s \circ \boldsymbol{p}, \boldsymbol{u}^s\right) = L\left(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{u}\right) \tag{16}$$

According to the generalization of Noether's theorem to optimal control, the symmetry of an optimal control problem implies the conservation of a generalized momentum map. In the present case, the rotational symmetry of the optimal control problem under consideration is associated with a generalized momentum map of the form

$$J_\xi(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{\psi}_q, \boldsymbol{\psi}^p) = \boldsymbol{\xi} \cdot \left(\boldsymbol{q}_I \times \boldsymbol{\psi}_q^I + \boldsymbol{p}^I \times \boldsymbol{\psi}_I^p\right) \tag{17}$$

The generalized momentum map (65) is conserved provided that conditions (15) and (16) are satisfied. This can be shown by a direct calculation along the lines of Torres [29]. We start with the infinitesimal version of the symmetry condition (16) given by

$$\begin{aligned} \left.\frac{d}{ds}\right|_{s=0} L\left(\boldsymbol{R}^s \circ \boldsymbol{q}, \boldsymbol{R}^s \circ \boldsymbol{p}, \boldsymbol{u}^s\right) &= 0 \\ \partial_{\boldsymbol{q}_I} L(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{u}) \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I + \partial_{\boldsymbol{p}^I} L(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{u}) \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}^I + \partial_{\boldsymbol{u}} L(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{u}) \cdot \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{u}^s &= 0 \end{aligned} \tag{18}$$

where use has been made of (14). Inserting from the adjoint equations (9) and the control equations (10) into the last equation yields

$$\left(\dot{\boldsymbol{\psi}}_q^I + \boldsymbol{\psi}^p \cdot \partial_{\boldsymbol{q}_I} \boldsymbol{f}_p(\boldsymbol{q}, \boldsymbol{u})\right) \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I + \left(\dot{\boldsymbol{\psi}}_I^p + \boldsymbol{\psi}_q \cdot \partial_{\boldsymbol{p}^I} \boldsymbol{f}^q(\boldsymbol{p})\right) \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}^I + \boldsymbol{\psi}^p \cdot \partial_{\boldsymbol{u}} \boldsymbol{f}_p(\boldsymbol{q}, \boldsymbol{u}) \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{u}^s = 0$$

The last equation can be recast in the form

$$\dot{\boldsymbol{\psi}}_q^I \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I + \dot{\boldsymbol{\psi}}_I^p \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}^I + \boldsymbol{\psi}_q^I \cdot \partial_{\boldsymbol{p}^J} \boldsymbol{f}_I^q(\boldsymbol{p}) \widehat{\boldsymbol{\xi}} \boldsymbol{p}^J + \boldsymbol{\psi}_I^p \cdot \left(\partial_{\boldsymbol{q}_J} \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) \widehat{\boldsymbol{\xi}} \boldsymbol{q}_J + \partial_{\boldsymbol{u}} \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{u}^s\right) = 0$$

Performing the derivative $d/ds|_{s=0}$ of (15) and taking into account the state equations (5), we obtain

$$\begin{aligned} \partial_{\boldsymbol{p}^J} \boldsymbol{f}_I^q(\boldsymbol{p}) \widehat{\boldsymbol{\xi}} \boldsymbol{p}^J &= \widehat{\boldsymbol{\xi}} \boldsymbol{f}_I^q(\boldsymbol{p}) &&= \widehat{\boldsymbol{\xi}} \dot{\boldsymbol{q}}_I \\ \partial_{\boldsymbol{q}_J} \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) \widehat{\boldsymbol{\xi}} \boldsymbol{q}_J + \partial_{\boldsymbol{u}} \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) \left.\tfrac{d}{ds}\right|_{s=0} \boldsymbol{u}^s &= \widehat{\boldsymbol{\xi}} \boldsymbol{f}_p^I(\boldsymbol{q}, \boldsymbol{u}) &&= \widehat{\boldsymbol{\xi}} \dot{\boldsymbol{p}}^I \end{aligned} \tag{19}$$

Taking into account the last two equations, we arrive at

$$\dot{\boldsymbol{\psi}}_q^I \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I + \dot{\boldsymbol{\psi}}_I^p \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}^I + \boldsymbol{\psi}_q^I \cdot \widehat{\boldsymbol{\xi}} \dot{\boldsymbol{q}}_I + \boldsymbol{\psi}_I^p \cdot \widehat{\boldsymbol{\xi}} \dot{\boldsymbol{p}}^I = 0$$

The last equation is equivalent to

$$\frac{d}{dt} \left(\boldsymbol{\psi}_q^I \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I + \boldsymbol{\psi}_I^p \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}^I \right) = 0$$

or

$$\boldsymbol{\xi} \cdot \frac{d}{dt} \left(\boldsymbol{q}_I \times \boldsymbol{\psi}_q^I + \boldsymbol{p}^I \times \boldsymbol{\psi}_I^p \right) = 0$$

Accordingly, the generalized momentum map (65) is a conserved quantity of the optimal control problem.

Remark 1. If condition (15) holds for the special case $\boldsymbol{u}^s = \boldsymbol{u}$ (i.e. for any $s \in \mathbb{R}$), the mechanical control system has state-space symmetry in the sense of Grizzle & Marcus [16, 17]. Then symmetry condition (15) can be recast in the form

$$D\Phi_g(\boldsymbol{x}) \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{f}(\Phi_g(\boldsymbol{x}), \boldsymbol{u}) \tag{20}$$

where the map $\Phi_g : \mathcal{P} \mapsto \mathcal{P}$ corresponds to the Lie group action of $g \in \mathrm{SO}(3)$ on $\boldsymbol{x} \in \mathcal{P}$. Using the nodal configuration and momentum vectors this map coincides with (12). If the system is uncontrolled, that is, if $\boldsymbol{u} = \boldsymbol{0}$, the state equations (2) can also be written in the form

$$\dot{\boldsymbol{x}} = X_{\mathcal{H}}(\boldsymbol{x})$$

If the underlying Hamiltonian system at hand has rotational symmetry, the Hamiltonian vector field $X_{\mathcal{H}}$ is $\mathrm{SO}(3)$-equivariant in the sense that

$$D\Phi_g(\boldsymbol{x}) X_{\mathcal{H}}(\boldsymbol{x}) = X_{\mathcal{H}}(\Phi_g(\boldsymbol{x})) \tag{21}$$

By comparing the last equation with (20), it is obvious that for mechanical control systems with state-space symmetry the symmetry of the uncontrolled system is a prerequisite for the symmetry of the mechanical control system. In general, the rotational symmetry of the underlying (uncontrolled) Hamiltonian system is passed on to the mechanical control system provided $\boldsymbol{u}^s$ exists such that symmetry condition (15) is satisfied.

3 STRUCTURE-PRESERVING DISCRETIZATION

We next devise a direct approach to the solution of the optimal control problem at hand that is capable to exactly conserve generalized momentum maps of the form (65) for mechanical optimal control systems with rotational symmetry. Our newly proposed approach accommodates a large class of one-step schemes for the integration of the state equations (1). To this end, the time interval of interest, $\mathcal{I} = [0, T]$, is devided into N non-overlapping sub-intervals whose length $h = t_{n+1} - t_n$ coincides with the time-step of the underlying one-step integrator.

Our approach relies on the introduction of a function $\mathbf{f_d} : \mathcal{P} \times \mathcal{P} \times \mathcal{U} \mapsto \mathbb{R}^{2n}$, which can be viewed as algorithmic approximation to the vector-valued function $h\boldsymbol{f} : \mathcal{P} \times \mathcal{U} \mapsto \mathbb{R}^{2n}$. Depending on the choice of $\mathbf{f_d}$ alternative one-step schemes of the form

$$\boldsymbol{x}_{n+1} - \boldsymbol{x}_n = \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) \tag{22}$$

are generated. Similarly, the density cost function $L : \mathcal{P} \times \mathcal{U} \mapsto \mathbb{R}$ is approximated on each time interval by the expression $\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n)/h$. We refer to Section 4 for specific examples.

The newly proposed direct method is based on the introduction of the following discrete optimal control Lagrangian

$$\begin{aligned} J_d = &\boldsymbol{\psi}_0 \cdot (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0) + \boldsymbol{\eta} \cdot \boldsymbol{g}_T(\boldsymbol{x}_N) + K(\boldsymbol{x}_N) \\ &+ \sum_{n=0}^{N-1} \left[\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) + \boldsymbol{\psi}_{n+1} \cdot (\boldsymbol{x}_{n+1} - \boldsymbol{x}_n - \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n)) \right] \\ &+ \sum_{n=0}^{N-1} \left[\boldsymbol{\alpha}_n \cdot (\boldsymbol{a}_n - \boldsymbol{x}_n - \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n)) \right] \end{aligned} \tag{23}$$

The necessary optimality conditions are derived by imposing the stationary of (23). Accordingly, the following algebraic system of equations need be satisfied for arbitrary variations of $\boldsymbol{x}_n, \boldsymbol{\psi}_n$ ($n = 0, \dots, N$), $\boldsymbol{a}_n, \boldsymbol{u}_n, \boldsymbol{\alpha}_n$ ($n = 0, \dots, N-1$) and $\boldsymbol{\eta}$:

$$\sum_{n=0}^{N-1} \delta\boldsymbol{\psi}_{n+1} \cdot (\boldsymbol{x}_{n+1} - \boldsymbol{x}_n - \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n)) = 0 \tag{24a}$$

$$\sum_{n=0}^{N-1} \delta\boldsymbol{\alpha}_n \cdot (\boldsymbol{a}_n - \boldsymbol{x}_n - \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n)) = 0 \tag{24b}$$

$$\sum_{n=0}^{N-1} \delta\boldsymbol{x}_n \cdot \left(\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n + \boldsymbol{\alpha}_n - D_1\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) + (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_1\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) \right) = 0 \tag{24c}$$

$$\sum_{n=0}^{N-1} \delta\boldsymbol{a}_n \cdot \left(\boldsymbol{\alpha}_n + D_2\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_2\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) \right) = 0 \tag{24d}$$

$$\sum_{n=0}^{N-1} \delta\boldsymbol{u}_n \cdot \left(D_3\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_3\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{a}_n, \boldsymbol{u}_n) \right) = 0 \tag{24e}$$

for $n = 0, \dots, N-1$ along with

$$\delta\boldsymbol{\psi}_0 \cdot (\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_0) = 0 \tag{25a}$$

$$\delta\boldsymbol{\eta} \cdot \boldsymbol{g}_T(\boldsymbol{x}_N) = 0 \tag{25b}$$

$$\delta\boldsymbol{x}_N \cdot \left(\boldsymbol{\psi}_N + D\boldsymbol{g}_T(\boldsymbol{x}_N)^T\boldsymbol{\eta} + DK(\boldsymbol{x}_N) \right) = 0 \tag{25c}$$

Due to the independence of $\delta\boldsymbol{\psi}_{n+1}, \delta\boldsymbol{\alpha}_n$ ($n = 0, \dots, N-1$), equations (24a) and (24b) yield $\boldsymbol{a}_n = \boldsymbol{x}_{n+1}$ ($n = 0, \dots, N-1$) along with the discrete state equations (22). Altogether, the

present approach leads to the following system of algebraic equations:

$$\boldsymbol{x}_{n+1} - \boldsymbol{x}_n = \mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) \tag{26a}$$
$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n + \boldsymbol{\alpha}_n = D_1\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_1\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) \tag{26b}$$
$$-\boldsymbol{\alpha}_n = D_2\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_2\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) \tag{26c}$$
$$\mathbf{0} = D_3\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot D_3\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) \tag{26d}$$

along with

$$\boldsymbol{x}_0 = \overline{\boldsymbol{x}}_0 \tag{27a}$$
$$\mathbf{0} = \boldsymbol{g}_T(\boldsymbol{x}_N) \tag{27b}$$
$$\boldsymbol{\psi}_N = -DK(\boldsymbol{x}_N) - D\boldsymbol{g}_T(\boldsymbol{x}_N)^T\boldsymbol{\eta} \tag{27c}$$

The above equations provide the general formulation of the necessary optimality conditions for the optimal control problem at hand. Specific schemes emanate from (26) by choosing appropriate expressions for $\mathbf{f_d}$ and $\mathrm{L_d}$. This will be further investigated in Section 4. First, however, we show that the general method (26) always generates schemes that are capable of conserving generalized momentum maps of the form (65).

3.1 Algorithmic conservation of the generalized momentum map

With regard to the partitioned form of the state equations, (5), the right-hand side of the discrete state equations introduced in (22) can be written in the form

$$\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = \begin{bmatrix} \mathbf{f_d}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) \\ \mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \end{bmatrix} \tag{28}$$

Similarly, the discrete version of the density cost function can be written as

$$\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = \mathrm{L_d}\left((\boldsymbol{q}_n, \boldsymbol{p}_n), (\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}), \boldsymbol{u}_n\right) \tag{29}$$

The partitioned form of the necessary optimality conditions (26) for the optimal control problem is given by

$$\boldsymbol{q}_{n+1} - \boldsymbol{q}_n = \mathbf{f_d}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) \tag{30a}$$
$$\boldsymbol{p}_{n+1} - \boldsymbol{p}_n = \mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \tag{30b}$$
$$\boldsymbol{\psi}_{q_{n+1}} - \boldsymbol{\psi}_{q_n} + \boldsymbol{\alpha}_{q_n} = \partial_{\boldsymbol{q}_n}\mathrm{L_d} - (\boldsymbol{\psi}^p_{n+1} + \boldsymbol{\alpha}^p_n) \cdot \partial_{\boldsymbol{q}_n}\mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \tag{30c}$$
$$\boldsymbol{\psi}^p_{n+1} - \boldsymbol{\psi}^p_n + \boldsymbol{\alpha}^p_n = \partial_{\boldsymbol{p}_n}\mathrm{L_d} - (\boldsymbol{\psi}_{q_{n+1}} + \boldsymbol{\alpha}_{q_n}) \cdot \partial_{\boldsymbol{p}_n}\mathbf{f_d}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) \tag{30d}$$
$$-\boldsymbol{\alpha}_{q_n} = \partial_{\boldsymbol{q}_{n+1}}\mathrm{L_d} - (\boldsymbol{\psi}^p_{n+1} + \boldsymbol{\alpha}^p_n) \cdot \partial_{\boldsymbol{q}_{n+1}}\mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \tag{30e}$$
$$-\boldsymbol{\alpha}^p_n = \partial_{\boldsymbol{p}_{n+1}}\mathrm{L_d} - (\boldsymbol{\psi}_{q_{n+1}} + \boldsymbol{\alpha}_{q_n}) \cdot \partial_{\boldsymbol{p}_{n+1}}\mathbf{f_d}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) \tag{30f}$$
$$\mathbf{0} = \partial_{\boldsymbol{u}_n}\mathrm{L_d} - (\boldsymbol{\psi}^p_{n+1} + \boldsymbol{\alpha}^p_n) \cdot \partial_{\boldsymbol{u}_n}\mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \tag{30g}$$

for $n = 0, \ldots, N-1$. In the above equations $\mathrm{L_d}$ stands for $\mathrm{L_d}\left((\boldsymbol{q}_n, \boldsymbol{p}_n), (\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}), \boldsymbol{u}_n\right)$. Now assume that the optimal control problem under consideration has rotational symmetry. That is, conditions (15) and (16) are satisfied. Then viable discrete versions of the state equations and the density cost function will inherit these symmetry properties (see Section 4 for

specific examples). In particular, the following two conditions will be fulfilled on the discrete level:

$$\begin{aligned} \boldsymbol{R}^s \mathrm{f}_{\mathrm{d}I}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) &= \mathrm{f}_{\mathrm{d}I}^q\left(\boldsymbol{R}^s \circ \boldsymbol{p}_n, \boldsymbol{R}^s \circ \boldsymbol{p}_{n+1}\right) \\ \boldsymbol{R}^s \mathrm{f}_{\mathrm{d}p}^I(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) &= \mathrm{f}_{\mathrm{d}p}^I\left(\boldsymbol{R}^s \circ \boldsymbol{q}_n, \boldsymbol{R}^s \circ \boldsymbol{q}_{n+1}, \boldsymbol{u}_n^s\right) \end{aligned} \tag{31}$$

and

$$\mathrm{L_d}\left((\boldsymbol{R}^s \circ \boldsymbol{q}_n, \boldsymbol{R}^s \circ \boldsymbol{p}_n), (\boldsymbol{R}^s \circ \boldsymbol{q}_{n+1}, \boldsymbol{R}^s \circ \boldsymbol{p}_{n+1}), \boldsymbol{u}_n^s\right) = \mathrm{L_d}\left((\boldsymbol{q}_n, \boldsymbol{p}_n), (\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}), \boldsymbol{u}_n\right) \tag{32}$$

If these conditions hold the present method does conserve the generalized momentum map (65) in the sense that

$$J_\xi(\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}, \boldsymbol{\psi}_{q_{n+1}}, \boldsymbol{\psi}_{n+1}^p) = J_\xi(\boldsymbol{q}_n, \boldsymbol{p}_n, \boldsymbol{\psi}_{q_n}, \boldsymbol{\psi}_n^p) \tag{33}$$

for $n = 0, \ldots, N-1$. To prove this we start with the infinitesimal version of the discrete symmetry condition (32). Similar to (18) we calculate

$$\left.\frac{d}{ds}\right|_{s=0} \mathrm{L_d}\left((\boldsymbol{R}^s \circ \boldsymbol{q}_n, \boldsymbol{R}^s \circ \boldsymbol{p}_n), (\boldsymbol{R}^s \circ \boldsymbol{q}_{n+1}, \boldsymbol{R}^s \circ \boldsymbol{p}_{n+1}), \boldsymbol{u}_n^s\right) = 0$$

$$\partial_{\boldsymbol{q}_{I_n}} \mathrm{L_d} \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_n} + \partial_{\boldsymbol{p}_n^I} \mathrm{L_d} \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_n^I + \partial_{\boldsymbol{q}_{I_{n+1}}} \mathrm{L_d} \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_{n+1}} + \partial_{\boldsymbol{p}_{n+1}^I} \mathrm{L_d} \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_{n+1}^I + \partial_{\boldsymbol{u}_n} \mathrm{L_d} \cdot \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{u}_n^s = 0$$

Inserting from (30c) through (30g) into the last equation yields

$$\begin{aligned} &\left[\boldsymbol{\psi}_{q_{n+1}}^I - \boldsymbol{\psi}_{q_n}^I + \boldsymbol{\alpha}_{q_n}^I + (\boldsymbol{\psi}_{n+1}^p + \boldsymbol{\alpha}_n^p) \cdot \partial_{\boldsymbol{q}_n} \mathrm{f}_{\mathrm{d}p}^I\right] \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_n} \\ &+ \left[\boldsymbol{\psi}_{I_{n+1}}^p - \boldsymbol{\psi}_{I_n}^p + \boldsymbol{\alpha}_{I_n}^p + (\boldsymbol{\psi}_{q_{n+1}} + \boldsymbol{\alpha}_{q_n}) \cdot \partial_{\boldsymbol{p}_n} \mathrm{f}_{\mathrm{d}I}^q\right] \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_n^I \\ &+ \left[-\boldsymbol{\alpha}_{q_n}^I + (\boldsymbol{\psi}_{n+1}^p + \boldsymbol{\alpha}_n^p) \cdot \partial_{\boldsymbol{q}_{n+1}} \mathrm{f}_{\mathrm{d}p}^I\right] \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_{n+1}} \\ &+ \left[-\boldsymbol{\alpha}_{I_n}^p + (\boldsymbol{\psi}_{q_{n+1}} + \boldsymbol{\alpha}_{q_n}) \cdot \partial_{\boldsymbol{p}_{n+1}} \mathrm{f}_{\mathrm{d}I}^q\right] \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_{n+1}^I \\ &+ \left(\boldsymbol{\psi}_{n+1}^p + \boldsymbol{\alpha}_n^p\right) \cdot \partial_{\boldsymbol{u}_n} \mathrm{f}_{\mathrm{d}p} \cdot \left.\frac{d}{ds}\right|_{s=0} \boldsymbol{u}_n^s = 0 \end{aligned} \tag{34}$$

On the other hand, the infinitesimal version of the discrete symmetry condition (31) can be derived in analogy to (19) and is given by

$$\begin{aligned} \partial_{\boldsymbol{p}_n^I} \mathrm{f}_{\mathrm{d}J}^q \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_n^I + \partial_{\boldsymbol{p}_{n+1}^I} \mathrm{f}_{\mathrm{d}J}^q \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_{n+1}^I &= \widehat{\boldsymbol{\xi}} \mathrm{f}_{\mathrm{d}J}^q = \widehat{\boldsymbol{\xi}}\left(\boldsymbol{q}_{J_{n+1}} - \boldsymbol{q}_{J_{n+1}}\right) \\ \partial_{\boldsymbol{q}_{I_n}} \mathrm{f}_{\mathrm{d}p}^J \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_n} + \partial_{\boldsymbol{q}_{I_{n+1}}} \mathrm{f}_{\mathrm{d}p}^J \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_{n+1}} + \partial_{\boldsymbol{u}_n} \mathrm{f}_{\mathrm{d}p}^J \cdot \left.\tfrac{d}{ds}\right|_{s=0} \boldsymbol{u}_n^s &= \widehat{\boldsymbol{\xi}} \mathrm{f}_{\mathrm{d}p}^J = \widehat{\boldsymbol{\xi}}\left(\boldsymbol{p}_{n+1}^J - \boldsymbol{p}_n^J\right) \end{aligned} \tag{35}$$

On the right-hand side of (35) the discrete state equations (30a) and (30b) have been taken into account. Substituting (35) into (34) yields

$$\boldsymbol{\psi}_{q_{n+1}}^I \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_{n+1}} - \boldsymbol{\psi}_{q_n}^I \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_{I_n} + \boldsymbol{\psi}_{I_{n+1}}^p \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_{n+1}^I - \boldsymbol{\psi}_{I_n}^p \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{p}_n^I = 0$$

or

$$\boldsymbol{\xi} \cdot \left(\boldsymbol{q}_{I_{n+1}} \times \boldsymbol{\psi}_{q_{n+1}}^I + \boldsymbol{p}_{n+1}^I \times \boldsymbol{\psi}_{I_{n+1}}^p - \boldsymbol{q}_{I_n} \times \boldsymbol{\psi}_{q_n}^I - \boldsymbol{p}_n^I \times \boldsymbol{\psi}_{I_n}^p\right) = 0$$

This result confirms algorithmic conservation of the generalized momentum map as stated in (33).

4 SPECIFIC SCHEMES

According to the direct approach to the solution of optimal control problems devised in the last section, the specific choice of one-step method for the integration of the state equations together with the specific discretization of the density cost function leads to a structure-preserving optimal control method that is capable of conserving the generalized momentum map. The purpose of this section is to further investigate and illustrate the design of specific schemes.

4.1 One-stage theta method

The discretization of the state equations by means of the one-stage theta method (see, for example, Stuart & Humphries [27]) relies on the specific choice

$$\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = h\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \tag{36}$$

in the discrete state equations (22). Here,

$$\boldsymbol{x}_{n+\theta} = (1-\theta)\boldsymbol{x}_n + \theta\boldsymbol{x}_{n+1} \tag{37}$$

for $\theta \in [0,1]$. In the special case of uncontrolled Hamiltonian systems, the one-stage theta method comprises, for example, both forward ($\theta = 0$) and backward ($\theta = 1$) Euler methods and the implicit mid-point rule ($\theta = 1/2$). Similar to (36), for the discretization of the density cost function we choose $\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = hL(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n)$. Before we substitute (36) into (26a) through (26d), we recast (26b) and (26c) in the form

$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n = D_1\mathrm{L_d} + D_2\mathrm{L_d} - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot (D_1\mathbf{f_d} + D_2\mathbf{f_d}) \tag{38a}$$

$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n + 2\boldsymbol{\alpha}_n = D_1\mathrm{L_d} - D_2\mathrm{L_d} - (\boldsymbol{\psi}_{n+1} + \boldsymbol{\alpha}_n) \cdot (D_1\mathbf{f_d} - D_2\mathbf{f_d}) \tag{38b}$$

With regard to (36) we have

$$\begin{aligned} D_1\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) &= h(1-\theta)\partial_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \\ D_2\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) &= h\theta\partial_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \end{aligned}$$

Similarly, we get

$$\begin{aligned} D_1\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) &= h(1-\theta)\partial_{\boldsymbol{x}}L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \\ D_2\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) &= h\theta\partial_{\boldsymbol{x}}L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \end{aligned}$$

Inserting these relationships into (38a) and (38b), a straightforward calculation yields

$$\boldsymbol{\alpha}_n = -\theta\left(\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n\right) \tag{39}$$

together with

$$\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n = h\partial_{\boldsymbol{x}}L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) - h\boldsymbol{\psi}_{n+(1-\theta)} \cdot \partial_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \tag{40}$$

Eventually, (26d) yields the control equations

$$\mathbf{0} = \partial_{\boldsymbol{u}}L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) - \boldsymbol{\psi}_{n+(1-\theta)} \cdot \partial_{\boldsymbol{u}}\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \tag{41}$$

Note that (40) and (41) can be viewed as discrete counterparts of the adjoint equations (7b) and the control equations (7c), respectively. Box 1 contains a summary of the theta family of structure-preserving optimal control schemes just derived.

Remark 2. For $\theta =\in \{0, \frac{1}{2}\}$ the method summarized in Box 1 can be directly linked to the discrete maximum principle proposed by Guibout & Bloch [18] (see, respectively, *Störmer's rule* and the *midpoint scheme* in [18]).

Box 1: The structure-preserving theta method for optimal control.

Using the optimal control Hamiltonian (6), the structure-preserving theta method for optimal control can be written in the form

$$\begin{aligned}
\boldsymbol{x}_{n+1} - \boldsymbol{x}_n &= h\partial_{\boldsymbol{\psi}} H(\boldsymbol{x}_{n+\theta}, \boldsymbol{\psi}_{n+(1-\theta)}, \boldsymbol{u}_n) \\
\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n &= -h\partial_{\boldsymbol{x}} H(\boldsymbol{x}_{n+\theta}, \boldsymbol{\psi}_{n+(1-\theta)}, \boldsymbol{u}_n) \\
\boldsymbol{0} &= \partial_{\boldsymbol{u}} H(\boldsymbol{x}_{n+\theta}, \boldsymbol{\psi}_{n+(1-\theta)}, \boldsymbol{u}_n)
\end{aligned}$$

or

$$\begin{aligned}
\boldsymbol{x}_{n+1} - \boldsymbol{x}_n &= h\boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{\psi}_{n+1} - \boldsymbol{\psi}_n &= h\partial_{\boldsymbol{x}} L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) - h\boldsymbol{\psi}_{n+(1-\theta)} \cdot \partial_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{0} &= \partial_{\boldsymbol{u}} L(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n) - \boldsymbol{\psi}_{n+(1-\theta)} \cdot \partial_{\boldsymbol{u}} \boldsymbol{f}(\boldsymbol{x}_{n+\theta}, \boldsymbol{u}_n)
\end{aligned}$$

together with

$$\begin{aligned}
\boldsymbol{x}_0 &= \overline{\boldsymbol{x}}_0 \\
\boldsymbol{0} &= \boldsymbol{g}_T(\boldsymbol{x}_N) \\
\boldsymbol{\psi}_N &= -DK(\boldsymbol{x}_N) - D\boldsymbol{g}_T(\boldsymbol{x}_N)^T \boldsymbol{\eta}
\end{aligned}$$

4.2 Partitioned theta method

We next consider a variant of the one-stage theta method which takes into account the partitioned form of the state equations (5). In particular, we choose

$$\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = \begin{bmatrix} \mathbf{f_d}^q(\boldsymbol{p}_n, \boldsymbol{p}_{n+1}) \\ \mathbf{f_d}_p(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) \end{bmatrix} = \begin{bmatrix} h\boldsymbol{f}^q(\boldsymbol{p}_{n+(1-\theta)}) \\ h\boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{u}_n) \end{bmatrix} \tag{42}$$

for $\theta \in [0, 1]$. In the case of uncontrolled Hamiltonian systems the present choice embraces the symplectic Euler methods (cf. Hairer et al. [20]) which are obtained for $\theta = 0$ and $\theta = 1$, respectively. As before, the implicit mid-point rule is obtained for $\theta = 1/2$. Concerning the discrete version of the density cost function we further choose

$$\begin{aligned}
\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) &= \mathrm{L_d}\left((\boldsymbol{q}_n, \boldsymbol{p}_n), (\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}), \boldsymbol{u}_n\right) \\
&= hL(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n)
\end{aligned} \tag{43}$$

In analogy to the last section we start with (38) to derive the resulting optimal control scheme. For completeness, we consider a more general version of (42) which is also valid for non-separable Hamiltonians. Accordingly, instead of (38) we consider

$$\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = \begin{bmatrix} h\boldsymbol{f}^q(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}) \\ h\boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n) \end{bmatrix} \tag{44}$$

Performing the derivatives needed in (38), we obtain

$$D_1\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = h \begin{bmatrix} (1-\theta)\partial_{\boldsymbol{q}} \boldsymbol{f}^q(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}) & \theta\partial_{\boldsymbol{p}} \boldsymbol{f}^q(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}) \\ (1-\theta)\partial_{\boldsymbol{q}} \boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n) & \theta\partial_{\boldsymbol{p}} \boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n) \end{bmatrix}$$

$$D_2\mathbf{f_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = h \begin{bmatrix} \theta\partial_{\boldsymbol{q}} \boldsymbol{f}^q(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}) & (1-\theta)\partial_{\boldsymbol{p}} \boldsymbol{f}^q(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}) \\ \theta\partial_{\boldsymbol{q}} \boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n) & (1-\theta)\partial_{\boldsymbol{p}} \boldsymbol{f}_p(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{u}_n) \end{bmatrix}$$

Furthermore,

$$D_1\mathrm{L_d}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1},\boldsymbol{u}_n) = h\begin{bmatrix}(1-\theta)\partial_{\boldsymbol{q}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\\ \theta\partial_{\boldsymbol{p}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\end{bmatrix}$$
$$D_2\mathrm{L_d}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1},\boldsymbol{u}_n) = h\begin{bmatrix}\theta\partial_{\boldsymbol{q}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\\ \theta(1-\theta)\partial_{\boldsymbol{p}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\end{bmatrix}$$

Now, inserting these relationships into (38a) and (38b), a straightforward calculation along the lines of the last section yields

$$\begin{aligned}\boldsymbol{\alpha}_{q_n} =& -\theta\left(\boldsymbol{\psi}_{q_{n+1}}-\boldsymbol{\psi}_{q_n}\right)\\ \boldsymbol{\alpha}_n^p =&(\theta-1)\left(\boldsymbol{\psi}_{n+1}^p-\boldsymbol{\psi}_n^p\right)\end{aligned} \tag{45}$$

along with

$$\begin{aligned}\begin{bmatrix}\boldsymbol{\psi}_{q_{n+1}}-\boldsymbol{\psi}_{q_n}\\ \boldsymbol{\psi}_{n+1}^p-\boldsymbol{\psi}_n^p\end{bmatrix} =& h\begin{bmatrix}\partial_{\boldsymbol{q}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\\ \partial_{\boldsymbol{p}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\end{bmatrix}\\ &-h\begin{bmatrix}\partial_{\boldsymbol{q}}\boldsymbol{f}^{q^T}(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)}) & \partial_{\boldsymbol{q}}f_p^T(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\\ \partial_{\boldsymbol{p}}\boldsymbol{f}^{q^T}(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)}) & \partial_{\boldsymbol{p}}f_p^T(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\end{bmatrix}\begin{bmatrix}\boldsymbol{\psi}_{q_{n+(1-\theta)}}\\ \boldsymbol{\psi}_{n+\theta}^p\end{bmatrix}\end{aligned} \tag{46}$$

Eventually, (26d) yields the control equations

$$\boldsymbol{0} = \partial_{\boldsymbol{u}}L(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n) - \partial_{\boldsymbol{u}}f_p^T(\boldsymbol{q}_{n+\theta},\boldsymbol{p}_{n+(1-\theta)},\boldsymbol{u}_n)\boldsymbol{\psi}_{n+\theta}^p \tag{47}$$

To summarize, the newly derived structure-preserving partitioned theta method for optimal control problems consists of the discrete state equations resulting from (44) (or (42) for separable Hamiltonians associated with the mechanical control system), the discrete adjoint equations (46) and the discrete control equations (47). A compact form of this scheme is given in terms of the optimal control Hamiltonian and can be found in Box 2. Note that the special case of a separable Hamiltonian associated with the underlying mechanical control system results if the optimal control Hamiltonian (11) is used in Box 2.

Remark 3. For $\theta = 1$ the method summarized in Box 2 coincides with the symplectic discrete-time state-adjoint system developed by Flaßkamp & Murphey [13] in the context of a linear optimal control problem (LQR).

4.3 Energy-momentum schemes

In the forward dynamics of Hamiltonian systems energy-momentum (EM) integrators preserve both the total mechanical energy and momentum maps associated with symmetries of the mechanical system. EM schemes (and energy-dissipative variants thereof, see, for example Armero & Romero [2]) can also be used for the discretization of the underlying Hamiltonian system related to the mechanical control systems under consideration. For this purpose we now choose in the discrete state equations (22)

$$\mathbf{f_d}(\boldsymbol{x}_n,\boldsymbol{x}_{n+1},\boldsymbol{u}_n) = \begin{bmatrix}\mathbf{f_d}^q(\boldsymbol{p}_n,\boldsymbol{p}_{n+1})\\ \mathbf{f_d}_p(\boldsymbol{q}_n,\boldsymbol{q}_{n+1},\boldsymbol{u}_n)\end{bmatrix} = \begin{bmatrix}h\boldsymbol{M}^{-1}\boldsymbol{p}_{n+\frac{1}{2}}\\ -h\overline{\nabla}\mathcal{V}(\boldsymbol{q}_n,\boldsymbol{q}_{n+1}) + h\boldsymbol{B}(\boldsymbol{q}_{n+\frac{1}{2}})\boldsymbol{u}_n\end{bmatrix} \tag{48}$$

Here, $\overline{\nabla}\mathcal{V}(\boldsymbol{q}_n,\boldsymbol{q}_{n+1})$ denotes a discrete derivative of the potential function $\mathcal{V}$ in the sense of Gonzalez [15]. The control extension of EM schemes yields energy-momentum consistent

Box 2: The structure-preserving partitioned theta method for optimal control.

$$
\begin{aligned}
\boldsymbol{q}_{n+1} - \boldsymbol{q}_n &= h\partial_{\boldsymbol{\psi}_q} H(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{\psi}_{q_{n+(1-\theta)}}, \boldsymbol{\psi}^p_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{p}_{n+1} - \boldsymbol{p}_n &= h\partial_{\boldsymbol{\psi}^p} H(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{\psi}_{q_{n+(1-\theta)}}, \boldsymbol{\psi}^p_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{\psi}_{q_{n+1}} - \boldsymbol{\psi}_{q_n} &= -h\partial_{\boldsymbol{q}} H(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{\psi}_{q_{n+(1-\theta)}}, \boldsymbol{\psi}^p_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{\psi}^p_{n+1} - \boldsymbol{\psi}^p_n &= -h\partial_{\boldsymbol{p}} H(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{\psi}_{q_{n+(1-\theta)}}, \boldsymbol{\psi}^p_{n+\theta}, \boldsymbol{u}_n) \\
\boldsymbol{0} &= \partial_{\boldsymbol{u}} H(\boldsymbol{q}_{n+\theta}, \boldsymbol{p}_{n+(1-\theta)}, \boldsymbol{\psi}_{q_{n+(1-\theta)}}, \boldsymbol{\psi}^p_{n+\theta}, \boldsymbol{u}_n)
\end{aligned}
$$

together with

$$
\begin{aligned}
\boldsymbol{q}_0 &= \overline{\boldsymbol{q}}_0 \\
\boldsymbol{p}_0 &= \overline{\boldsymbol{p}}_0 \\
\boldsymbol{0} &= \boldsymbol{g}_T(\boldsymbol{q}_N, \boldsymbol{p}_N) \\
\boldsymbol{\psi}_{q_N} &= -\partial_{\boldsymbol{q}} K(\boldsymbol{q}_N, \boldsymbol{p}_N) - \boldsymbol{\eta} \cdot \partial_{\boldsymbol{q}} \boldsymbol{g}_T(\boldsymbol{q}_N, \boldsymbol{p}_N) \\
\boldsymbol{\psi}^p_N &= -\partial_{\boldsymbol{p}} K(\boldsymbol{q}_N, \boldsymbol{p}_N) - \boldsymbol{\eta} \cdot \partial_{\boldsymbol{p}} \boldsymbol{g}_T(\boldsymbol{q}_N, \boldsymbol{p}_N)
\end{aligned}
$$

schemes which are capable of exactly reproducing the balance laws for energy and angular momentum in the discrete setting. We refer to Betsch & Sänger for further investigations in this direction. Concerning the discrete version of the density cost function we choose

$$
\mathrm{L_d}(\boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{u}_n) = hL(\boldsymbol{x}_{n+\frac{1}{2}}, \boldsymbol{u}_n) \tag{49}
$$

which corresponds to the choice $\theta = \frac{1}{2}$ in the last two sections. Again, inserting (48) and (49) into (26) yields a direct method for the optimal control problem which is capable of conserving the generalized momentum map (65). We refer to Section 5 for a sample application of an EM scheme.

4.4 Structure-preservation on two levels

This section provides a summary of the structure-preserving properties of the numerical methods dealt with in this paper. First of all we recall the specific form of the affine Hamiltonian control systems under consideration. With regard to (5) the state equations can be written in the form

$$
\begin{aligned}
\dot{\boldsymbol{q}}_I &= \boldsymbol{f}^q_I(\boldsymbol{p}) &&= (M^{-1})_{IJ}\boldsymbol{p}^J \\
\dot{\boldsymbol{p}}^I &= \boldsymbol{f}^I_p(\boldsymbol{q}, \boldsymbol{u}) &&= -\partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) + \boldsymbol{f}^I_{\mathrm{ext}}(\boldsymbol{q}, \boldsymbol{u})
\end{aligned} \tag{50}
$$

Note that the underlying Hamiltonian system is given by

$$
\begin{aligned}
\dot{\boldsymbol{q}}_I &= (M^{-1})_{IJ}\boldsymbol{p}^J \\
\dot{\boldsymbol{p}}^I &= -\partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q})
\end{aligned} \tag{51}
$$

Suppose that the underlying Hamiltonian system has rotational symmetry. This implies that the potential energy function is invariant under superposed rotations. That is, using the notation introduced in Section 2.2,

$$
\mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q}) = \mathcal{V}(\boldsymbol{q}) \tag{52}
$$

Performing the time derivative of (52) yields

$$\begin{aligned} \tfrac{d}{dt}\mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q}) &= \tfrac{d}{dt}\mathcal{V}(\boldsymbol{q}) \\ \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q}) \cdot \boldsymbol{R}^s \dot{\boldsymbol{q}}_I &= \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) \cdot \boldsymbol{R}^s \dot{\boldsymbol{q}}_I \end{aligned}$$

A straightforward calculation leads to the result

$$\big(\boldsymbol{R}^s \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) - \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q})\big) \cdot \boldsymbol{R}^s \dot{\boldsymbol{q}}_I = 0$$

which has to hold for arbitrary nodal velocities $\dot{\boldsymbol{q}}_I$. Due to the non-singularity of the rotation matrix $\boldsymbol{R}^s$, the last equation implies

$$\boldsymbol{R}^s \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) = \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q}) \tag{53}$$

Moreover, in view of $(50)_1$ and $(51)_1$, we get

$$\boldsymbol{f}_I^q(\boldsymbol{R}^s \circ \boldsymbol{p}) = (M^{-1})_{IJ} \boldsymbol{R}^s \boldsymbol{p}^J = \boldsymbol{R}^s \boldsymbol{f}_I^q(\boldsymbol{p}) \tag{54}$$

Note that (54) and (53) corroborate $\mathrm{SO}(3)$-equivariance of the Hamiltonian vector field as stated in (21) (see Remark 1). We further calculate the infinitesimal version of invariance property (52):

$$\left.\frac{d}{ds}\right|_{s=0} \mathcal{V}(\boldsymbol{R}^s \circ \boldsymbol{q}) = \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}) \cdot \widehat{\boldsymbol{\xi}} \boldsymbol{q}_I = 0$$

The last equation can be recast in the form

$$\boldsymbol{\xi} \cdot \big(\boldsymbol{q}_I \times \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q})\big) = 0 \tag{55}$$

We remark that a similar calculation based on (53) yields the same result.

4.4.1 Mechanical control system

It is well-known that a mechanical control system satisfies fundamental mechanical balance laws. In the present work we focus on the balance laws for angular momentum and energy, respectively. A time-stepping scheme for the numerical integration of the state equations shall be called structure-preserving mechanical integrator[1] if these balance laws are exactly reproduced. Correspondingly, the specific integrator applied to the mechanical control system can be associated with the first level of structure preservation. The angular momentum of a mechanical system relative to the origin of an inertial frame is defined by

$$\boldsymbol{j} = \boldsymbol{q}_I \times \boldsymbol{p}^I \tag{56}$$

It can be verified by a straightforward calculation that the following relationship holds:

$$\boldsymbol{j}_{n+1} - \boldsymbol{j}_n = \boldsymbol{q}_{I_{n+\theta}} \times \big(\boldsymbol{p}^I_{n+1} - \boldsymbol{p}^I_n\big) - \boldsymbol{p}^I_{n+(1-\theta)} \times \big(\boldsymbol{q}_{I_{n+1}} - \boldsymbol{q}_{I_n}\big)$$

for $\theta \in [0,1]$. If we now apply the partitioned theta method (see (42) in Section 4.2), the last equation yields

$$\boldsymbol{j}_{n+1} - \boldsymbol{j}_n = h\boldsymbol{q}_{I_{n+\theta}} \times \partial_{\boldsymbol{q}_I}\mathcal{V}(\boldsymbol{q}_{n+\theta}) - h(M^{-1})_{IJ}\boldsymbol{p}^I_{n+(1-\theta)} \times \boldsymbol{p}^J_{n+(1-\theta)} + h\boldsymbol{q}_{I_{n+\theta}} \times \boldsymbol{f}^I(\boldsymbol{q}_{n+\theta}, \boldsymbol{u}_n)$$

[1]The term mechanical integrator has originally been coined by Marsden [24] in the context of conservative mechanical systems with symmetry.

Taking into account the symmetry properties (55) and $(M^{-1})_{IJ} = (M^{-1})_{JI}$, the last equation leads to the result

$$\boldsymbol{\xi} \cdot \left(\boldsymbol{j}_{n+1} - \boldsymbol{j}_n \right) = h \boldsymbol{\xi} \cdot \left(\boldsymbol{q}_{I_{n+\theta}} \times \boldsymbol{f}^I(\boldsymbol{q}_{n+\theta}, \boldsymbol{u}_n) \right) \tag{57}$$

Accordingly, all members of the theta method are capable of conserving angular momentum in the case of vanishing controls (i.e. $\boldsymbol{f}^I = \boldsymbol{0}$). Moreover, for mechanical control systems the balance law for angular momentum will be consistently reproduced according to the discrete balance law (57). Analogous conclusions can be drawn in the case of the energy-momentum scheme outlined in Section 4.3. Moreover, concerning the theta method dealt with in Section 4.1, the only angular momentum consistent method is the implicit mid-point rule ($\theta = \frac{1}{2}$). Note that a time-stepping scheme that is not capable to conserve angular momentum in the purely Hamiltonian case will in general fail to correctly reproduce balance of angular momentum for finite time steps.

Of the one-step schemes considered herein, the energy-momentum integrator is the only energy consistent method. Taking into account the properties of the discrete derivative in (48), it can be verified that, similar to (57), the discrete balance law for the energy is satisfied for arbitrary time steps according to

$$\mathcal{H}(\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}) - \mathcal{H}(\boldsymbol{q}_n, \boldsymbol{p}_n) = \left(\boldsymbol{q}_{I_{n+1}} - \boldsymbol{q}_{I_n} \right) \cdot \boldsymbol{f}^I(\boldsymbol{q}_{n+\frac{1}{2}}, \boldsymbol{u}_n) \tag{58}$$

Here, the Hamiltonian $\mathcal{H}(\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1})$ has been defined in (3) and corresponds to the total mechanical energy of the system. Note that in the case of vanishing controls, i.e. if $\boldsymbol{f}^I = \boldsymbol{0}$, the energy-momentum integrator conserves the total energy.

4.4.2 Optimal control problem

The second level of structure preservation is associated with the optimal control problem. Indeed, as has been proven in Section 3.1, by design the present approach yields numerical methods capable of conserving generalized momentum maps related to rotational symmetries of the optimal control problem. This 2nd level conservation property holds, irrespective of the one-step scheme used for the integration of the state equations. That is, the two levels of structure preservation turn out to be independent of each other.

Since the optimal control problem itself has an intrinsic Hamiltonian structure, the optimal control Hamiltonian defined in (6) is also a conserved quantity. Note however, that the numerical methods developed in this work are not capable of conserving the Hamiltonian of the optimal control problem for a finite number of time steps. This issue will be further investigated in the context of the numerical example presented in Section 5.

5 NUMERICAL EXAMPLE

5.1 Nonlinear spring pendulum

The numerical example deals with the nonlinear spring pendulum depicted in Fig. 1. The motion of the controlled mechanical system at hand is governed by state equations of the form (5). In particular, we have

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1} \boldsymbol{p} \\ \dot{\boldsymbol{p}} &= -D\mathcal{V}(\boldsymbol{q}) + \boldsymbol{B}(\boldsymbol{q}) \boldsymbol{u} \end{aligned}$$

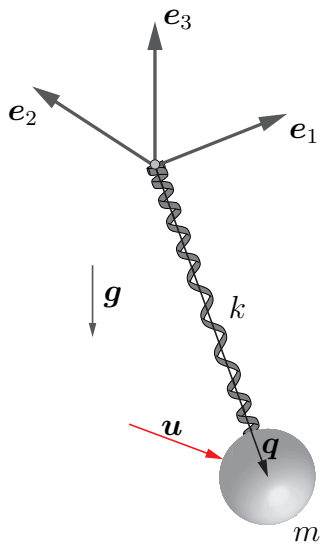


Figure 1: The nonlinear spring pendulum

Here, $\boldsymbol{q} \in \mathbb{R}^3$ is the position vector of the particle with mass m. Relative to the inertial Cartesian basis $\{\boldsymbol{e}_i\}$ (see Fig.1) we have $\boldsymbol{q}(t) = q_i(t)\boldsymbol{e}_i$. The mass matrix is diagonal and given by $\boldsymbol{M} = m\boldsymbol{I}$. The potential energy function assumes the form

$$\mathcal{V}(\boldsymbol{q}) = \underbrace{-\boldsymbol{g}\cdot\boldsymbol{M}\boldsymbol{q}}_{=\mathcal{V}_{\text{ext}}(\boldsymbol{q})} + \underbrace{\frac{1}{2}kl_0^2\varepsilon(\boldsymbol{q})^2}_{=\mathcal{V}_{\text{int}}(\boldsymbol{q})} \tag{59}$$

where

$$\boldsymbol{g} = -g\boldsymbol{e}_3 \tag{60}$$

is the gravitational acceleration and

$$\varepsilon(\boldsymbol{q}) = \frac{1}{2\,l_0^2}\left(\boldsymbol{q}\cdot\boldsymbol{q} - l_0^2\right) \tag{61}$$

is a scalar Green-Lagrangian-type strain measure. Note that the potential energy can be split into an external part $\mathcal{V}_{\text{ext}}$ due to gravity and an internal part $\mathcal{V}_{\text{int}}$ due to deformation of the spring. In (61), $l_0 = \sqrt{\boldsymbol{q}_0\cdot\boldsymbol{q}_0}$ defines the strain-free state of the spring. The motion of the spring pendulum is controlled by a force acting on particle m. That is, in the state equations $\boldsymbol{B}(\boldsymbol{q}) = \boldsymbol{I}$ such that the actuating force coincides with $\boldsymbol{u} \in \mathbb{R}^3$.

5.1.1 Symmetry properties

It can be easily verified that the potential energy is invariant under superposed rotations about the $\boldsymbol{e}_3$-axis. That is, setting $\boldsymbol{R}_3^s = \exp(s\widehat{\boldsymbol{e}}_3)$ and taking into account the identity $\boldsymbol{R}_3^s\boldsymbol{e}_3 = \boldsymbol{e}_3$, it can be deduced from (59) that

$$\mathcal{V}(\boldsymbol{R}_3^s\boldsymbol{q}) = \mathcal{V}(\boldsymbol{q}) \tag{62}$$

The last equation corresponds to symmetry property (52). Consequently, corresponding to (53), the identity

$$\boldsymbol{R}_3^s D\mathcal{V}(\boldsymbol{q}) = D\mathcal{V}(\boldsymbol{R}_3^s\boldsymbol{q})$$

holds. It can now be easily checked that symmetry condition (15) is satisfied provided that

$$\boldsymbol{u}^s = \boldsymbol{R}_3^s \boldsymbol{u} \tag{63}$$

Accordingly, the present mechanical control system has rotational symmetry about the $\boldsymbol{e}_3$-axis. Moreover, we choose the density cost function

$$\mathrm{L}(\boldsymbol{u}) = \frac{1}{2}\boldsymbol{u} \cdot \boldsymbol{u} \tag{64}$$

such that $\mathrm{L}(\boldsymbol{u}^s) = \mathrm{L}(\boldsymbol{u})$ and symmetry condition (16) is also satisfied. Accordingly, the optimal control problem under consideration has rotational symmetry about the $\boldsymbol{e}_3$-axis. Consequently, the associated generalized momentum map

$$J_3 = \boldsymbol{e}_3 \cdot \left(\boldsymbol{q} \times \boldsymbol{\psi}_q + \boldsymbol{p} \times \boldsymbol{\psi}^p \right) \tag{65}$$

is a conserved quantity.

5.1.2 Specific schemes under investigation

Of the newly developed structure-preserving schemes dealt with in Section 4 we apply the theta method (Section 4.1) with $\theta \in \{0, \frac{1}{2}, 1\}$, the partitioned theta method (Section 4.2) with $\theta \in \{0, \frac{1}{2}, 1\}$, and the energy-momentum scheme outlined in Section 4.3. Note that both the theta method and the partitioned theta method rely on

$$\begin{aligned} \mathbf{f}_{\mathrm{d}p}(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) &= -hD\mathcal{V}(\boldsymbol{q}_{n+\theta}) + h\boldsymbol{u}_n \\ &= -hmg\boldsymbol{e}_3 - hk\varepsilon(\boldsymbol{q}_{n+\theta})\boldsymbol{q}_{n+\theta} + h\boldsymbol{u}_n \end{aligned} \tag{66}$$

in the discrete state equations. In contrast to that, the energy-momentum integrator is based on

$$\begin{aligned} \mathbf{f}_{\mathrm{d}p}(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}, \boldsymbol{u}_n) &= -h\overline{\nabla}\mathcal{V}(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}) + h\boldsymbol{u}_n \\ &= -hmg\boldsymbol{e}_3 - \frac{hk}{2}\left(\varepsilon(\boldsymbol{q}_n) + \varepsilon(\boldsymbol{q}_{n+1})\right)\boldsymbol{q}_{n+\frac{1}{2}} + h\boldsymbol{u}_n \end{aligned} \tag{67}$$

where use has been made of the discrete derivative of the internal potential function given by

$$\overline{\nabla}\mathcal{V}_{\mathrm{int}}(\boldsymbol{q}_n, \boldsymbol{q}_{n+1}) = \frac{k}{2}\left(\varepsilon(\boldsymbol{q}_n) + \varepsilon(\boldsymbol{q}_{n+1})\right)\boldsymbol{q}_{n+\frac{1}{2}} \tag{68}$$

The methods under investigation are summarized in Table 1. Note that the schemes **theta-$\frac{1}{2}$** and **ptheta-$\frac{1}{2}$** coincide.

5.1.3 Numerical results

In the optimal control problem under investigation both the initial and the terminal state of the spring pendulum are prescribed. The goal is to determine the maneuver of the spring pendulum between the prescribed start and end points such that the control effort is minimized. The data used in the numerical example is summarized in Table 2.

First of all, the converged numerical solution obtained with a high resolution of $N = 10000$ time steps is depicted in Figs. 2, 3 and 4. In particular, Fig. 2 contains the time evolution of the state coordinates $\boldsymbol{q}(t) = q_i(t)\boldsymbol{e}^i$, $\boldsymbol{p}(t) = p^i(t)\boldsymbol{e}_i$ and the associated adjoint variables

denomination	method
theta-0	theta method
theta-$\frac{1}{2}$	(Section 4.1)
theta-1	$\theta \in \{0, \frac{1}{2}, 1\}$
ptheta-0	partitioned theta method
ptheta-$\frac{1}{2}$	(Section 4.2)
ptheta-1	$\theta \in \{0, \frac{1}{2}, 1\}$
EM	energy-momentum method
	(Section 4.3)

Table 1: Nonlinear spring pendulum: Methods under investigation

m (mass)	1
k (spring constant)	0.6
g (gravity)	9.8
l_0	5
$\overline{\boldsymbol{q}}_0$ (initial)	$(-2, -5, -5)$
$\overline{\boldsymbol{p}}_0$	$(-3, 0, 0)$
$\overline{\boldsymbol{q}}_N$ (terminal)	$(2, -10, -4)$
$\overline{\boldsymbol{p}}_N$	$(0, 0, 0)$
T	5
$\varepsilon_{\text{Newton}}$	10^{-9}

Table 2: Nonlinear spring pendulum: Data used in the numerical example

$\boldsymbol{\psi}_q(t) = \psi_q^i(t)\boldsymbol{e}_i$, $\boldsymbol{\psi}^p(t) = \psi_i^p(t)\boldsymbol{e}^i$. Fig. 3 depicts the controls $\boldsymbol{u} = u^i\boldsymbol{e}_i$ and Fig. 4 shows the evolution of mechanical energies and angular momentum.

In Fig. 5 the fulfillment of the balance law for the mechanical energy is checked for a comparatively rough resolution of $N = 50$ time steps. Specifically, formula (58) for the discrete balance of energy is applied to the present problem and yields

$$\mathcal{R}_{n+1}^{\text{En}} = \mathcal{H}(\boldsymbol{q}_{n+1}, \boldsymbol{p}_{n+1}) - \mathcal{H}(\boldsymbol{q}_n, \boldsymbol{p}_n) - \big(\boldsymbol{q}_{n+1} - \boldsymbol{q}_n\big) \cdot \boldsymbol{u}_n \tag{69}$$

As expected, the only energy consistent method which exactly satisfies the discrete balance of energy in the sense that $\mathcal{R}_{n+1} = 0$ (up to numerical round-off errors) for $n = 0, \ldots, N-1$ is the scheme **EM**. Of course, for $N \to \infty$ (or $h \to 0$), $\mathcal{R}_{n+1} \to 0$ for all schemes under consideration.

The discrete fulfillment of the balance law for angular momentum can be checked by adapting (57) to the present problem. Accordingly, in Fig. 6 we consider the residual

$$\mathcal{R}_{n+1}^{\text{AM}} = \boldsymbol{e}_3 \cdot \Big(\boldsymbol{j}_{n+1} - \boldsymbol{j}_n - h\boldsymbol{q}_{I_{n+\theta}} \times \boldsymbol{u}_n\Big) \tag{70}$$

for $n = 0, \ldots, N-1$. Note that for the scheme **EM**, $\theta = \frac{1}{2}$ has to be used in (70). It can be observed from Fig. 6 that the schemes **ptheta-0**, **theta-$\frac{1}{2}$**, **ptheta-1** and **EM** are indeed

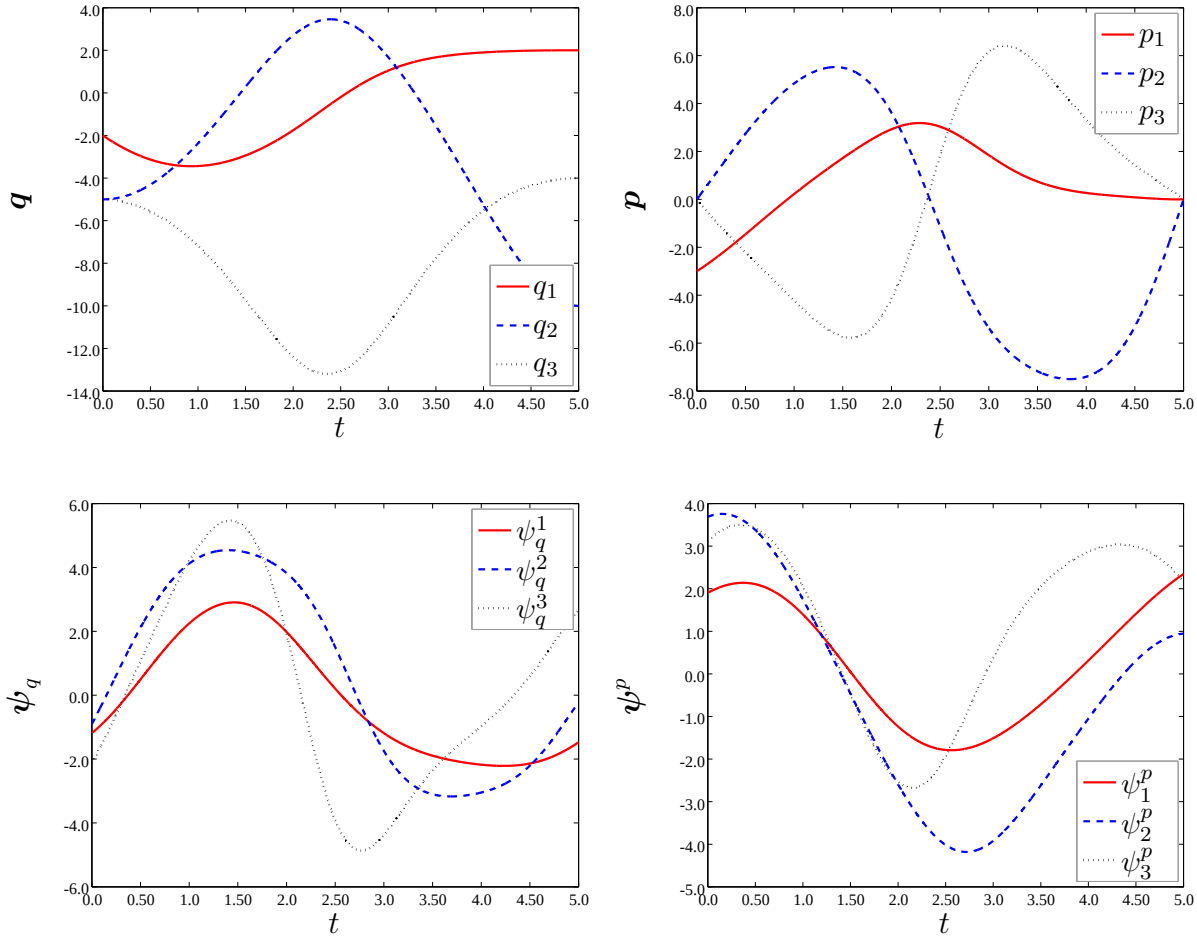


Figure 2: Nonlinear spring pendulum: State coordinates $(\boldsymbol{q}, \boldsymbol{p})$ and adjoint coordinates $(\boldsymbol{\psi}_q, \boldsymbol{\psi}^p)$ obtained with a high resolution of $N = 10000$ time steps

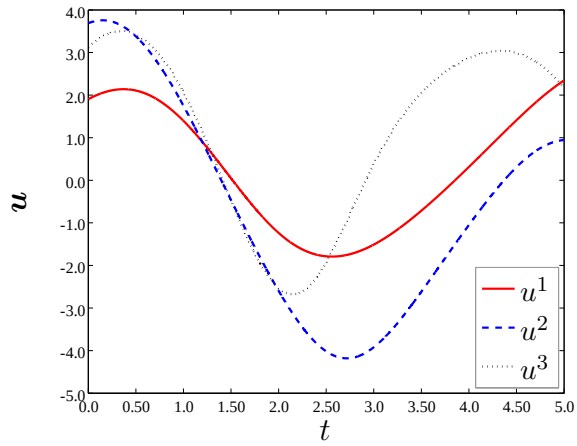


Figure 3: Nonlinear spring pendulum: Controls $\boldsymbol{u}$ obtained with a high resolution of $N = 10000$ time steps

consistently reproducing balance of angular momentum up to numerical round-off errors. This is in sharp contrast to the schemes **theta-0** and **theta-1**. Again this type of consistency error decreases when number N is increased.

Next the structure-preserving properties related to the second level (cf. Section 4.4.2) are

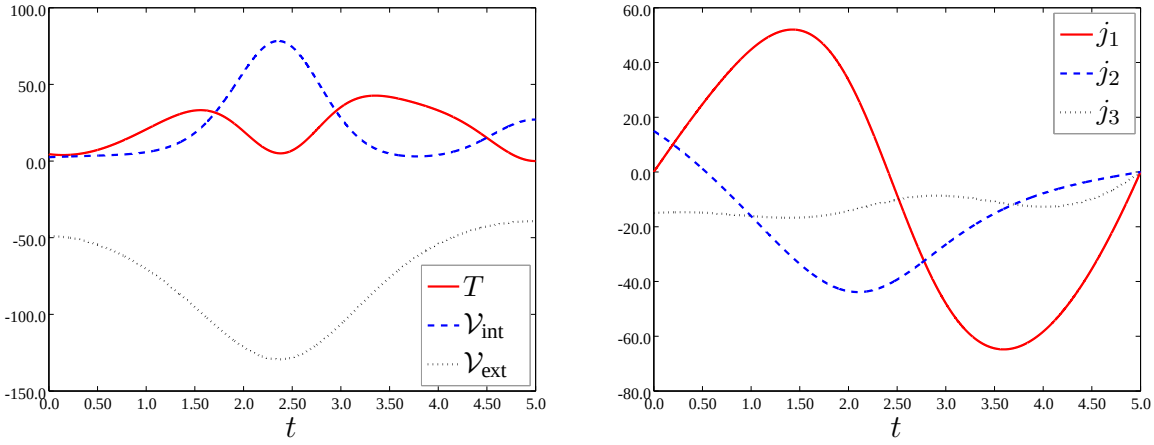


Figure 4: Nonlinear spring pendulum: Kinetic energy $T = \frac{1}{2}\boldsymbol{p} \cdot \boldsymbol{M}^{-1}\boldsymbol{p}$, gravitational potential energy $\mathcal{V}_{\text{ext}}$, strain energy $\mathcal{V}_{\text{int}}$ of the nonlinear spring and components $j_i = \boldsymbol{e}_i \cdot (\boldsymbol{q} \times \boldsymbol{p})$ of angular momentum obtained with a high resolution of $N = 10000$ time steps

considered. By design all schemes under investigation exactly preserve the generalized momentum map (65) associated with the rotational symmetry of the present optimal control problem. For completeness this conservation property is corroborated by Fig. 7 for the schemes **ptheta-0**, **theta-$\frac{1}{2}$**, **ptheta-1** and **EM**. Similarly, this conservation property holds (again up to numerical round-off errors) for the remaining schemes **theta-0** and **theta-1**.

As has been mentioned in Section 4.4.2, the optimal control Hamiltonian is also a 2nd level conserved quantity. However, as expected, none of the schemes under investigation exactly reproduces this conservation property. This can be observed from Figs. 8 through 10. The reference value of the constant optimal control Hamiltonian is equal to -0.37. All of the schemes under investigation converge to this reference value. However, for low resolutions of the optimal control problem significant deviations from the reference value can be observed (Figs. 8 through 10).

The errors in the state variables $\boldsymbol{x}$ and the adjoint variables $\boldsymbol{\psi}$ have been calculated via

$$\begin{aligned} \varepsilon_x &= \frac{1}{N+1}\sum_{n=0}^{N}\frac{||\boldsymbol{x}(t_n) - \boldsymbol{x}_{\text{ref}}(t_n)||_2}{||\boldsymbol{x}_{\text{ref}}||_2} \\ \varepsilon_\psi &= \frac{1}{N+1}\sum_{n=0}^{N}\frac{||\boldsymbol{\psi}(t_n) - \boldsymbol{\psi}_{\text{ref}}(t_n)||_2}{||\boldsymbol{\psi}_{\text{ref}}||_2} \end{aligned} \tag{71}$$

The corresponding plots are depicted in Figs. 11 and 12. It can be seen that the schemes **theta-$\frac{1}{2}$** and **EM** are second-order accurate whereas the schemes **theta-0**, **theta-1**, **ptheta-0** and **ptheta-1** are first-order accurate.

Eventually, to illustrate the calculated optimal solution, Fig. 13 shows a sequence of snapshots at successive points in time. In addition to the trajectory of particle m, the arrows indicate both magnitude (corresponding to the lengths of the arrows) and direction of the control force acting on particle m.

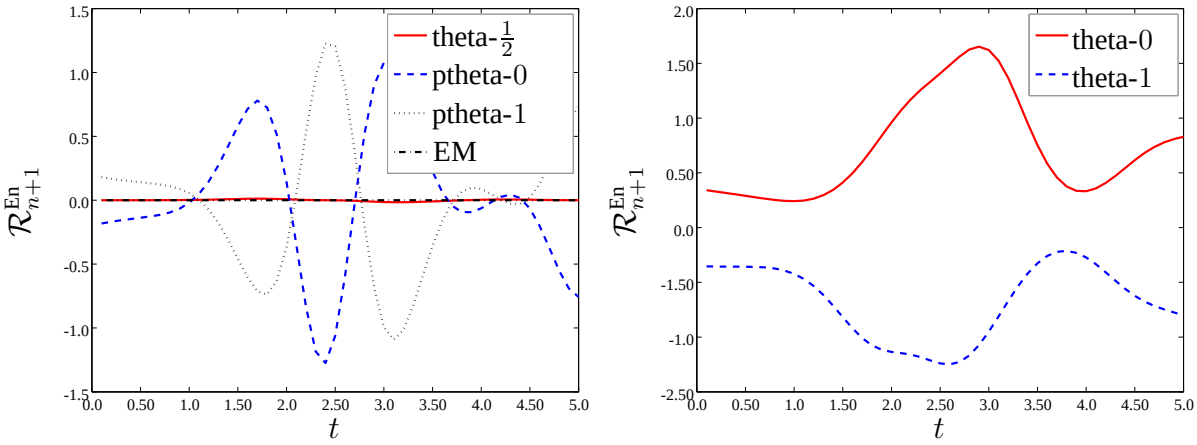


Figure 5: Nonlinear spring pendulum: Balance of energy for a low resolution of $N = 50$ time steps. The non-fulfillment of the balance of energy is measured by the residual $\mathcal{R}^{\mathrm{En}}_{n+1}$ defined in (69)

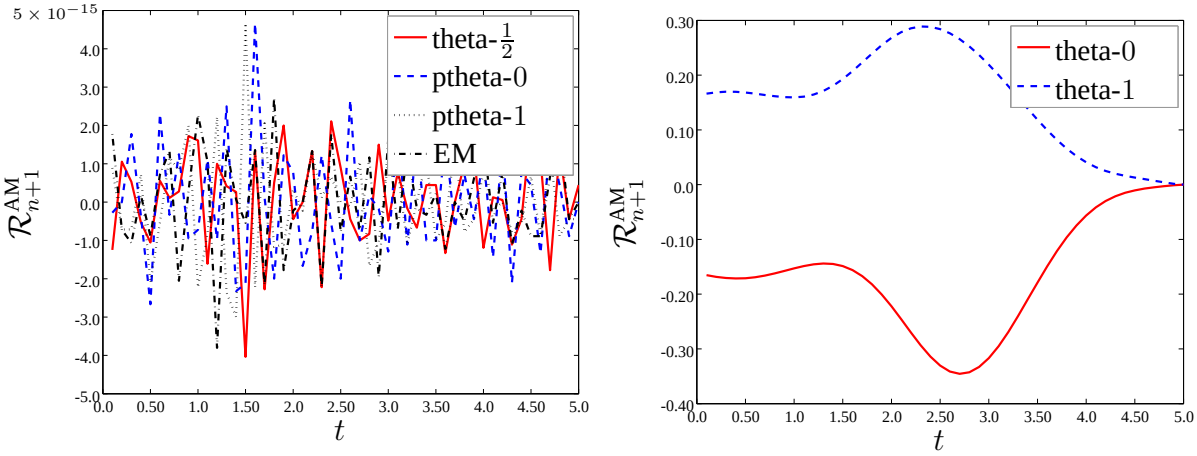


Figure 6: Nonlinear spring pendulum: Balance of angular momentum for a low resolution of $N = 50$ time steps. The non-fulfillment of the balance of angular momentum is measured by the residual $\mathcal{R}^{\mathrm{AM}}_{n+1}$ defined in (70)

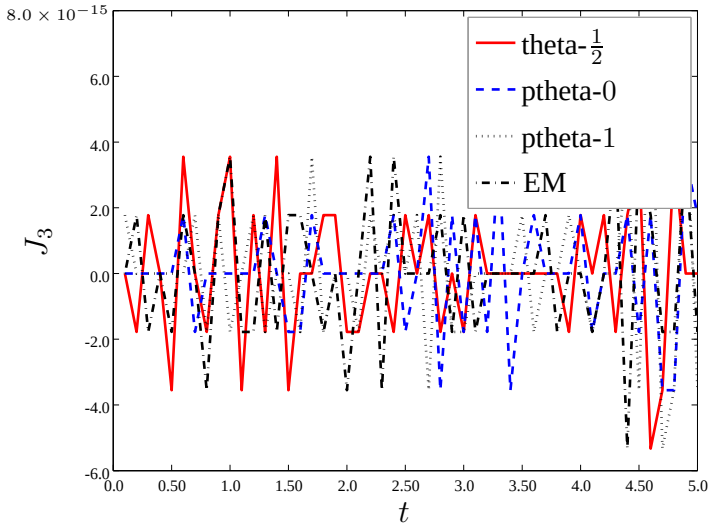


Figure 7: Nonlinear spring pendulum: Algorithmic conservation of the generalized momentum map J_3, cf. (65), up to numerical round-off errors. The optimal control problem has been solved with a low resolution of $N = 50$ time steps

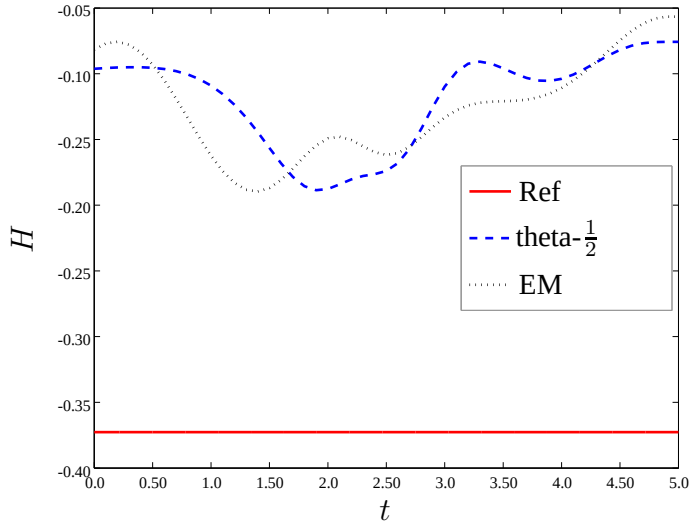


Figure 8: Nonlinear spring pendulum: Optimal control Hamiltonian obtained with the schemes **theta-$\frac{1}{2}$** and **EM** for $N = 50$ time steps. The (constant) reference value equals -0.37

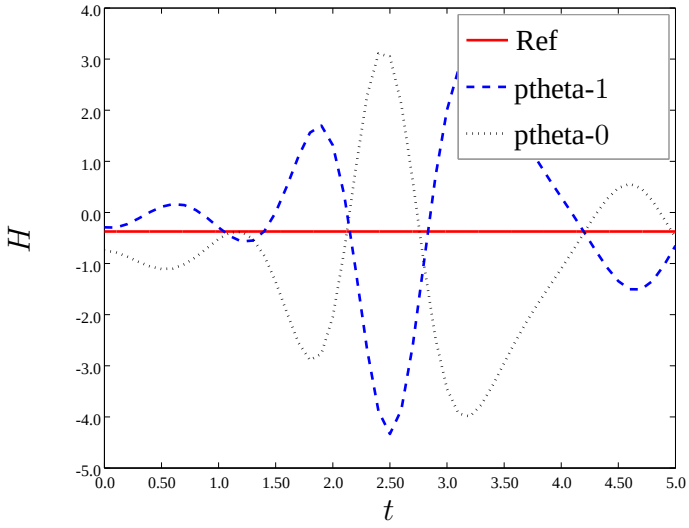


Figure 9: Nonlinear spring pendulum: Optimal control Hamiltonian obtained with the schemes **ptheta-0** and **ptheta-1** for $N = 50$ time steps. The (constant) reference value equals -0.37

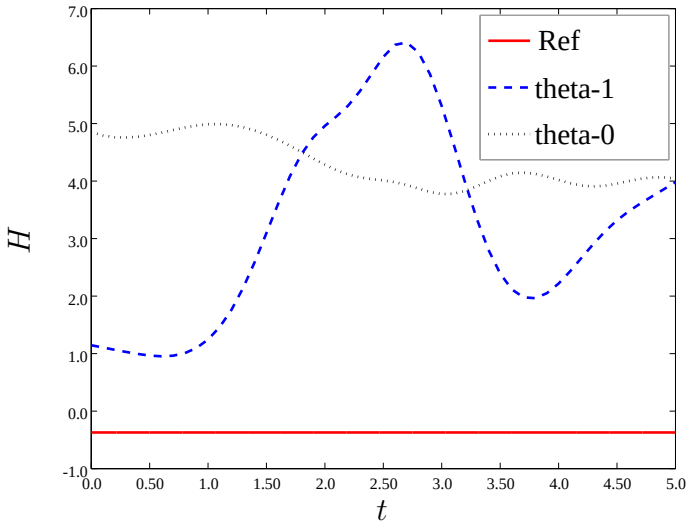


Figure 10: Nonlinear spring pendulum: Optimal control Hamiltonian obtained with the schemes **theta-0** and **theta-1** for $N = 50$ time steps. The (constant) reference value equals -0.37

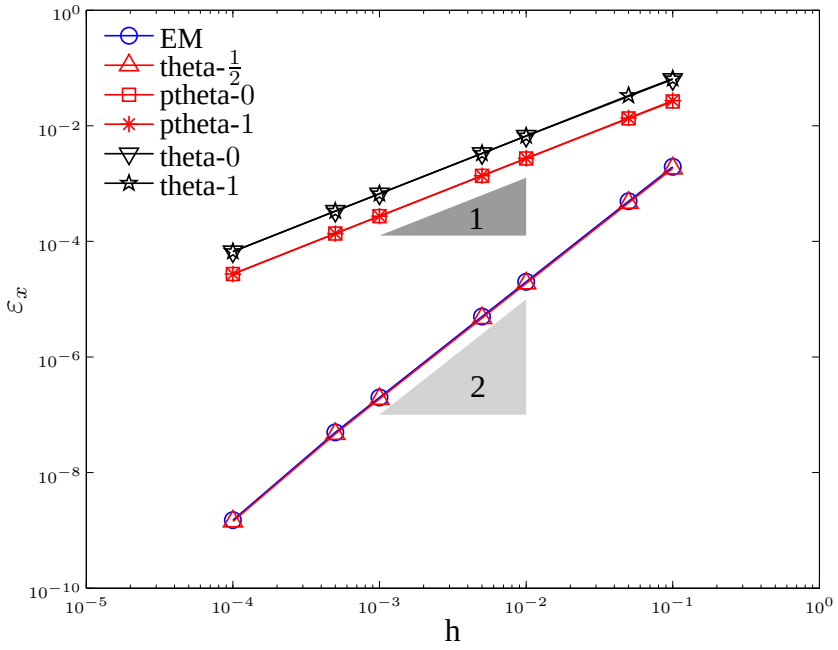


Figure 11: Nonlinear spring pendulum: Error ε_x in the state variables

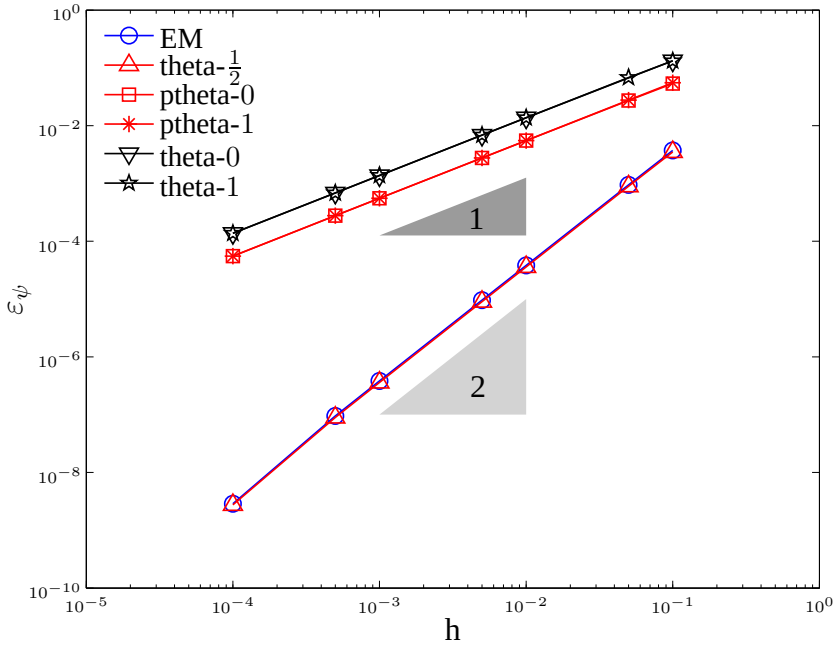


Figure 12: Nonlinear spring pendulum: Error ε_ψ in the adjoint variables

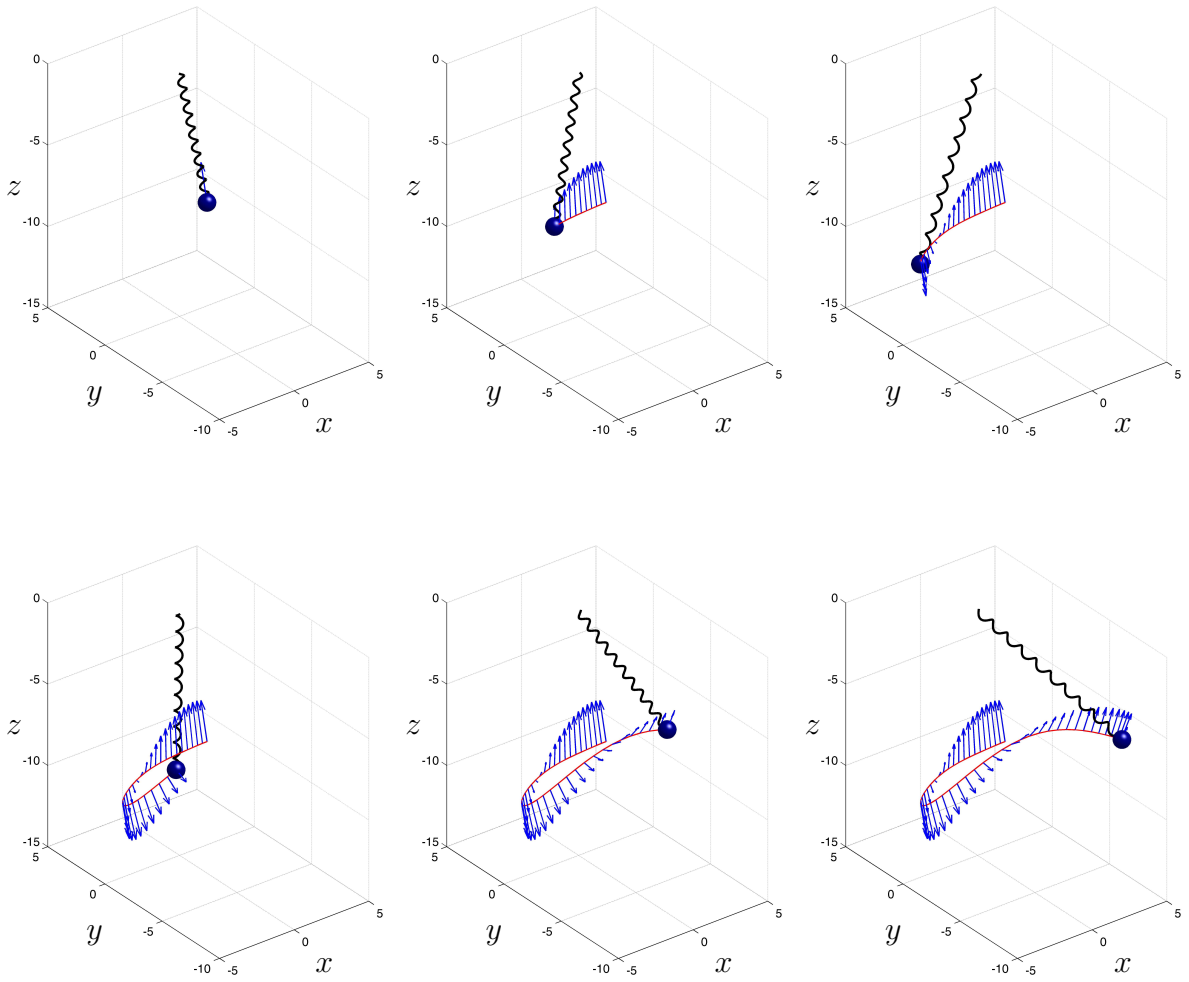


Figure 13: Nonlinear spring pendulum: Snapshots of the motion at times $t \in \{0, 1, 2, 3, 4, 5\}$

6 CONCLUSIONS

We devised a new structure-preserving direct approach to the numerical solution of optimal control problems in the framework of finite-dimensional mechanical systems. We may distinguish between two levels of structure-preservation. The first level of structure-preservation is linked to the forward dynamics of the underlying mechanical control system. That is, structure-preserving integrators such as momentum-symplectic or energy-momentum schemes can be applied to the numerical integration of the state equations. The second level of structure-preservation is connected to the optimal control problem itself. We saw that under specific symmetry conditions generalized momentum maps are conserved along an optimal path.

The newly devised direct method for the solution of the optimal control problem in Section 3 generates a family of schemes that by design conserve generalized momentum maps associated with rotational symmetries of the optimal control problem. To some extend our approach is related to previous work of Hager [19], Ober-Blöbaum et al. [26], Flaßkamp & Murphey [13] and Campos et al. [10]. Although those works do not consider symmetries and associated generalized momentum maps, the design of the discrete optimal control Lagrangian (23) hinges heavily on the *State Uniqueness Property* of the discrete control problem considered in Hager [19, Sec. 2].

Acknowledgement

Support for this research was provided by the Deutsche Forschungsgemeinschaft (DFG) under Grant BE 2285/10-1. This support is gratefully acknowledged.

REFERENCES

[1] S.K. Agrawal and B.C. Fabien. *Optimization of Dynamic Systems*. Kluwer Academic Publishers, 1999.

[2] F. Armero and I. Romero. On the formulation of high-frequency dissipative time-stepping algorithms for nonlinear dynamics. Part II: Second-order methods. *Comput. Methods Appl. Mech. Engrg.*, 190:6783–6824, 2001.

[3] P. Betsch, R. Siebert, and N. Sänger. Natural coordinates in the optimal control of multibody systems. *J. Comput. Nonlinear Dynam.*, 7(1):011009/1–8, 2012.

[4] J.T. Betts. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. SIAM, Philadelphia, PA, 2nd edition, 2010.

[5] L.T. Biegler. *Nonlinear Programming. Concepts, Algorithms, and Applications to Chemical Processes*. SIAM, Philadelphia, 2010.

[6] T. Binder, L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder, and O. von Stryk. Introduction to model based optimization of chemical processes on moving horizons. In M. Grötschel, S.O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 295–340. Springer-Verlag, 2001.

[7] A.M. Bloch. *Nonholonomic Mechanics and Control*. Springer-Verlag, 2003.

[8] A.M. Bloch, I.I. Hussein, M. Leok, and A.K. Sanyal. Geometric structure-preserving optimal control of a rigid body. *Journal of Dynamical and Control Systems*, 15(3):307–330, 2009.

[9] C.L. Bottasso and A. Croce. Optimal control of multibody systems using an energy preserving direct transcription method. *Multibody System Dynamics*, 12(1):17–45, 2004.

[10] C.M. Campos, S. Ober-Blöbaum, and E. Trlélat. High order variational integrators in the optimal control of mechanical systems. *Discrete and Continuous Dynamical Systems – Series A (DCDS-A)*, 35(9):4193–4223, 2015.

[11] M. Diehl, H.G. Bock, H. Diedam, and P.-B. Wieber. Fast direct multiple shooting algorithms for optimal robot control. In M. Diehl and K. Mombaur, editors, *Fast Motions in Biomechanics and Robotics*, volume 340 of *Lecture Notes in Control and Information Sciences*, pages 65–93. Springer-Verlag, 2006.

[12] D.S. Djukić. Noether's theorem for optimum control systems. *Int. J. Control*, 18(3):667–672, 1973.

[13] K. Flaßkamp and T.D. Murphey. Structure-preserving integration for linear optimal control and filtering of mechanical systems. Preprint March 2015.

[14] M. Gerdts. *Optimal Control of ODEs and DAEs*. De Gruyter, Berlin/Boston, 2012.

[15] O. Gonzalez. Time integration and discrete Hamiltonian systems. *J. Nonlinear Sci.*, 6:449–467, 1996.

[16] J. Grizzle and S. Marcus. Optimal control of systems possessing symmetries. *IEEE Transactions on Automatic Control*, 29(11):1037–1040, 1984.

[17] J. Grizzle and S. Marcus. The structure of nonlinear control systems possessing symmetries. *IEEE Transactions on Automatic Control*, 30(3):248–258, 1985.

[18] V. Guibout and A.M. Bloch. A discrete maximum principle for solving optimal control problems. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 1806–1811, Atlantis, Paradise Island, Bahamas, 2004.

[19] W.W. Hager. Runge-Kutta methods in optimal control and the transformed adjoint system. *Numer. Math.*, 87(2):247–282, 2000.

[20] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer-Verlag, 2nd edition, 2006.

[21] M.W. Koch and S. Leyendecker. Energy momentum consistent force formulation for the optimal control of multibody systems. *Multibody System Dynamics*, 29:381–401, 2013.

[22] S. Leyendecker, S. Ober-Blöbaum, J.E. Marsden, and M. Ortiz. Discrete mechanics and optimal control for constrained systems. *Optim. Control Appl. Meth.*, 31(6):505–528, 2010.

[23] G. Little and E.R. Pinch. The Pontryagin maximum principle: the constancy of the Hamiltonian. *IMA J Math Control Info*, 13(4):403–408, 1996.

[24] J.E. Marsden. *Lectures on Mechanics*. Cambridge University Press, 1992.

[25] J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, 2nd edition, 1999.

[26] S. Ober-Blöbaum, O. Junge, and J.E. Marsden. Discrete mechanics and optimal control: An analysis. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(2):322–352, 2011.

[27] A.M. Stuart and A.R. Humphries. *Dynamical Systems and Numerical Analysis*. Cambridge University Press, 1996.

[28] D.F.M. Torres. Conservation laws in optimal control. In F. Colonius and L. Grüne, editors, *Dynamics, Bifurcations, and Control*, volume 273 of *Lecture Notes in Control and Information Sciences*, pages 287–296. Springer-Verlag, 2002.

[29] D.F.M. Torres. On the Noether theorem for optimal control. *European Journal of Control*, 8(1):56–63, 2002.

[30] A.J. van der Schaft. Symmetries in optimal control. *SIAM J. Control and Optimization*, 25(2):245–259, 1987.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

VARIATIONAL INTEGRATORS OF MIXED ORDER FOR DYNAMICAL SYSTEMS WITH MULTIPLE TIME SCALES AND SPLIT POTENTIALS

Theresa Wenger[1], Sina Ober-Blöbaum[2], and Sigrid Leyendecker[1]

[1]Chair of Applied Dynamics
University of Erlangen-Nuremberg
Haberstrasse 1, 91058 Erlangen, Germany
e-mail: theresa.wenger@fau.de, sigrid.leyendecker@fau.de

[2] Department of Engineering Science
University of Oxford
Parks Road, Oxford, OX1 3PJ, United Kingdom
e-mail:sina.ober-blobaum@eng.ox.ac.uk

Keywords: multiple time scales, higher order variational integrators, Fermi-Pasta-Ulam problem

Abstract. *The simulation of mechanical systems that act on multiple time scales, caused e.g. by different types or stiffnesses in potentials, is challenging as a stable integration of the fast dynamics requires a highly accurate approximation whereas for the simulation of the slow part a coarser approximation is accurate enough. With regard to the general goals of any numerical method, high accuracy and low computational costs, the presented variational integrators of mixed order couple coarse and fine approximations. The idea builds up on the higher order Galerkin variational integrators in [9] that are derived via Hamilton's variational principle with a polynomial to approximate the configuration and an appropriate quadrature formula for the approximation of the integral of the Lagrangian. For the variational integration of systems with dynamics on multiple time scales, we use polynomials of different degrees to approximate the components that act on different time scales. Furthermore, quadrature formulas of different order approximate the integrals of the single energy contributions of the Lagrangian. This approach provides great flexibility in the design of the integrators. Their performance is investigated numerically by means of the Fermi-Pasta-Ulam problem and a numerical analysis regarding accuracy versus efficiency is carried out, where we focus on the integrators most promising to resolve the mentioned trade-off.*

1 INTRODUCTION

Many systems in Lagrangian mechanics exhibit dynamics on multiple time scales caused by different types and stiffnesses in potentials, e.g. highly oscillatory problems with a potential that can be split in a fast quadratic part and a slow nonlinear part. The fast dynamics require a highly accurate approximation whereas for the slow motion a coarser approximation is sufficient. This contrast leads to a conflict of the general goals, high accuracy and low computational costs, and thus is challenging for numerical methods. For the simulation of Lagrangian mechanical systems, variational integrators are convenient as they inherit specific properties of the underlying system by construction. The variational integrators base on a discrete version of Hamilton's principle inducing symplectic and momentum preserving integration schemes with excellent long time behaviour [8].

Multirate integrators resolve the mentioned trade-off by separating the degrees of freedom into fast and slow components and integrating them on different time grids, e.g. [5], [3], [1], [2]. A rather large time step is chosen for the integration of the slow components and a fine grid for the fast components. In [4], multirate schemes that base on partitioned Runge-Kutta methods are introduced. The multirate integrators in [7] base on variational mechanics allowing flexibility in design as the use of different quadrature rules for the approximation of the appearing integrals is enabled. Another approach is the splitting method that does not tend to actually resolve the fast motion, but to capture the multiscale phenomena of slow energy exchange and adiabatic invariants, e.g. implicit-explicit (IMEX) methods. The IMEX methods integrate the expensive slow forces explicitly and the cheap fast forces implicitly, see e.g. [5], [6]. A derivation of the IMEX numerical integration from a Lagrangian variational point of view is given in [10]. Another efficient framework for simulating systems with dynamics on multiple time scales is provided by averaging methods. The fast motion is not resolved, only an estimation is fed into the slow equations of motion. A powerful concept is known as HMM (heterogeneous multiscale methods [12]) with a wide range of applications in multiscale and multiphysics problems. Based on a top-down strategy, the motion of the microscale model is estimated through averaging and supplies the incomplete macroscale model with the missing information. Another example of averaging methods are FLAVORS (flow averaging integrators [11]). These structure preserving integrators base on variational methods and use a splitting and resynchronisation technique to average the flow.

However, the idea of the presented integrators here is different. The underlying approach originates from the higher order Galerkin variational integrator in [9], where a polynomial is used to approximate the configuration together with an appropriate quadrature formula to approximate the integral of the Lagrangian. Let the configuration be split in components acting on various time-scales. Instead of multiple time grids, only one time grid is used and the components that act on different time scales are approximated via polynomials of different degrees. For the approximation of the integral of the Lagrangian, different quadrature formulas for its contributions (the kinetic energy, the slow and the fast potential) are considered. This yields a high number of possibilities to construct the variational integrators. With regard to save computational runtime, low degree polynomials to approximate the slow configuration and quadrature formulas of low order for the approximation of the slow integral are of particular interest as the former has a direct influence on the number of unknowns and the latter on the number of costly function evaluations of the slow potential in one time step. Higher accuracy, however, can be achieved by increasing the degree of the polynomial approximating the fast configuration. The performance and the computational efficiency of the variational integrators of mixed order are investigated

numerically by means of a standard benchmark problem, the Fermi-Pasta-Ulam problem.

2 VARIATIONAL INTEGRATORS OF MIXED ORDER

2.1 Lagrangian dynamics

Consider an n-dimensional mechanical system on a configuration manifold $Q \subseteq \mathbb{R}^n$ with configuration vector $q(t) \in Q$ and velocity vector $\dot{q}(t) \in T_{q(t)}Q$ in the tangent space. The variable t denotes the time bounded on the interval $[0, T] \subset \mathbb{R}$. The Lagrangian $L : TQ \to \mathbb{R}$ of the mechanical system consists of the difference of the kinetic energy $T(\dot{q})$ and the potential $U(q)$. Let the potential $U = V + W$ be composed of different parts with strongly varying dynamics acting on different time scales such that it can be split in a slow potential V and a fast potential W. Furthermore, suppose that the n-dimensional variable q can be separated into n^s slow variables $q^s \in Q^s$ and n^f fast variables $q^f \in Q^f$ such that $Q^s \times Q^f = Q$ and $q = (q^s, q^f)$ with $n^s + n^f = n$. It is assumed that the fast potential $W = W(q^f)$ depends on the fast degrees of freedom only whereas the slow potential depends on the complete configuration variable q, i.e. $V = V(q)$. Under these assumptions the Lagrangian $L : TQ \to \mathbb{R}$ reads

$$L(q, \dot{q}) = T(\dot{q}) - V(q) - W(q^f)$$

The action $S : C^2([0, T], Q^s \times Q^f) \to \mathbb{R}$ is the integral in $[0, T]$ of the Lagrangian

$$S(q) = \int_0^T L(q, \dot{q})\, dt \tag{1}$$

depending on the smooth curves $q^s \in C^2([0, T], Q^s)$ and $q^f \in C^2([0, T], Q^f)$ with fixed initial value $q(0) = (q^s(0), q^f(0))$ and fixed final value $q(T) = (q^s(T), q^f(T))$. Hamilton's principle requires that the first variation of the action S vanishes and yields the Euler-Lagrange equations of motion

$$\begin{aligned} \frac{d}{dt}\frac{\partial T}{\partial \dot{q}^s} + \frac{\partial V}{\partial q^s} &= 0 \\ \frac{d}{dt}\frac{\partial T}{\partial \dot{q}^f} + \frac{\partial V}{\partial q^f} + \frac{\partial W}{\partial q^f} &= 0 \end{aligned}$$

The Legendre-transform $\mathbb{F}L : TQ \to T^*Q$ reading in coordinate form

$$\mathbb{F}L(q, \dot{q}) = (q, p) = \left(q, \frac{\partial L}{\partial \dot{q}}(q, \dot{q})\right)$$

relates $(q, \dot{q}) \in TQ$ to $(q, p) \in T^*Q$ in the cotangent space, where $p \in T_q^*Q$ denotes the conjugate momentum.

2.2 Discrete variational principle

The variational integrators of mixed order base on the higher order Galerkin variational integrators presented in [9], where a polynomial is used to approximate the continuous curve q on one time step and the integral of the Lagrangian is approximated via an appropriate quadrature formula. The approach here allows for the dynamics acting on different time scales. For the approximation of the slow variables q^s and the fast variables q^f, polynomials of different degrees are considered. With the splitting of the potential in a fast and a slow one different quadrature

formulas for the approximation of each can be used.

The space of trajectories $\mathcal{C}([0,h],Q^s) = \{q^s : [0,h] \to Q^s | q^s(0) = q_k^{0,s}, q^s(h) = q_{k+1}^{0,s}\}$ and $\mathcal{C}([0,h],Q^f) = \{q^f : [0,h] \to Q^f | q^f(0) = q_k^{0,f}, q^f(h) = q_{k+1}^{0,f}\}$ reduces to the finite-dimensional space of trajectories $\mathcal{C}^{ps}([0,h],Q^s) = \{q^s \in \mathcal{C}([0,h],Q^s) \mid q^s \in \Pi^{ps}\}$ and $\mathcal{C}^{pf}([0,h],Q^f) = \{q^f \in \mathcal{C}([0,h],Q^f) \mid q^f \in \Pi^{pf}\}$, whereby Π^{ps} and Π^{pf} denote the space of polynomials of degree ps respectively pf. With $ps+1$ configurations $q_k^s = (q_k^{0,s}, \dots, q_k^{ps,s}) \in (Q^s)^{ps+1}$ at $ps+1$ control points $0 = d_0^s < d_1^s < \dots < d_{ps-1}^s < d_{ps}^s = 1$, the ps degree polynomial q_d^s is uniquely defined such that it passes through each $q_k^{\nu,s}$ at the time $d_\nu^s h$, $\nu = 0, \dots, ps$. The polynomial q_d^f of degree pf is uniquely defined by the $pf+1$ configurations $q_k^f = (q_k^{0,f}, \dots, q_k^{pf,f}) \in (Q^f)^{pf+1}$ at $pf+1$ control points $0 = d_0^f < d_1^f < \dots < d_{pf-1}^f < d_{pf}^f = 1$ passing through each $q_k^{\nu,f}$ at the time $d_\nu^f h$, $\nu = 0, \dots, pf$. Using the Lagrange polynomials $l_{\nu,ps} : [0,1] \to \mathbb{R}$, $\nu = 0, \dots, ps$ respectively $l_{\nu,pf} : [0,1] \to \mathbb{R}$, $\nu = 0, \dots, pf$, the polynomial q_d^s approximating the slow variables and the polynomial q_d^f approximating the fast variables on $[0,h]$ can be written in the form

$$q_d^s(t; q_k^s, h) = \sum_{\nu=0}^{ps} q_k^{\nu,s}\, l_{\nu,ps}\left(\frac{t}{h}\right) \qquad q_d^f(t; q_k^f, h) = \sum_{\nu=0}^{pf} q_k^{\nu,f}\, l_{\nu,pf}\left(\frac{t}{h}\right)$$

Differentiating q_d^s and q_d^f w.r.t. t provides approximations of $\dot{q}^s$ and $\dot{q}^f$ on $[0,h]$

$$\dot{q}_d^s(t; q_k^s, h) = \frac{1}{h}\sum_{\nu=0}^{ps} q_k^{\nu,s}\, \dot{l}_{\nu,ps}\left(\frac{t}{h}\right) \qquad \dot{q}_d^f(t; q_k^f, h) = \frac{1}{h}\sum_{\nu=0}^{pf} q_k^{\nu,f}\, \dot{l}_{\nu,pf}\left(\frac{t}{h}\right)$$

Let the time interval $[0,T]$ be divided into $N = T/h$ subintervals of lengths h. On each subinterval, the continuous curves of the slow and fast configurations are approximated piecewisely via the polynomials $q_d^s : [kh, (k+1)h] \to Q^s$ and $q_d^f : [kh, (k+1)h] \to Q^f$, $k = 0, \dots, N-1$. To guarantee continuity on $[0,T]$ the conditions $q_k^{ps,s} = q_{k+1}^{0,s}$ and $q_k^{pf,f} = q_{k+1}^{0,f}$ for all $k = 0, \dots, N-2$ must hold.

Replacing the continuous curves piecewisely by the corresponding polynomials and approximating the integral of the Lagrangian via appropriate quadrature formulas yield an approximation of the action S on each subinterval $[kh, (k+1)h]$, $k = 0, \dots, N-1$. In particular, different quadrature formulas for the approximation of the integral of the contributions of the Lagrangian are considered. The quadrature rule $(b_i^t, c_i^t)_{i=1}^{r^t}$ is chosen for the integral of the kinetic energy, $(b_i^v, c_i^v)_{i=1}^{r^v}$ for the integral of the slow potential and $(b_i^w, c_i^w)_{i=1}^{r^w}$ for the integral of the fast potential. The quadrature rules are all w.r.t. the time interval $[0,1]$ with quadrature points $c_i^j \in [0,1]$ and weights $b_i^j \in \mathbb{R}$ $i = 1, \dots, r^j$, $j = v, w, t$. The discrete action then reads

$$S_d(q_0, \dots, q_{N-1}) = \sum_{k=0}^{N-1} L_d(q_k)$$

with $q_k = (q_k^s, q_k^f)$ and the discrete Lagrangian

$$\begin{aligned} L_d(q_k) = h \sum_{i=1}^{r^t} b_i^t T \left((\dot{q}_d^s(c_i^t h; q_k^s, h), \dot{q}_d^f(c_i^t h; q_k^f, h)) \right) \\ - h \sum_{i=1}^{r^v} b_i^v V \left((q_d^s(c_i^v h; q_k^s, h), q_d^f(c_i^v h; q_k^f, h)) \right) - h \sum_{i=1}^{r^w} b_i^w W \left((q_d^f(c_i^w h; q_k^f, h)) \right) \end{aligned}$$

providing an approximation of the action (1). The discrete version of Hamilton's principle implies the discrete action S_d to be stationary. Hence the derivatives of the discrete action S_d w.r.t. $q_k^{\nu,s}$, $\nu = 0, \dots, ps$ and $q_k^{\nu,f}$, $\nu = 0, \dots, pf$ have to vanish $\forall k = 0, \dots, N-1$, with $\delta q_0^{0,s} = \delta q_0^{0,f} = \delta q_N^{0,s} = \delta q_N^{0,f} = 0$. This leads to the discrete Euler-Lagrange equations

$$\begin{aligned} 0 =& D_{ps+1} L_d(q_k^{0,s}, \dots, q_k^{ps,s}, q_k^{0,f}, \dots, q_k^{pf,f}) + D_1 L_d(q_{k+1}^{0,s}, \dots, q_{k+1}^{ps,s}, q_{k+1}^{0,f}, \dots, q_{k+1}^{pf,f}) \\ & k = 0, \dots, N-2 \\ 0 =& D_i L_d(q_k^{0,s}, \dots, q_k^{ps,s}, q_k^{0,f}, \dots, q_k^{pf,f}) \quad i = 2, \dots, ps \qquad k = 0, \dots, N-1 \\ 0 =& D_{(ps+1)+pf+1} L_d(q_k^{0,s}, \dots, q_k^{ps,s}, q_k^{0,f}, \dots, q_k^{pf,f}) \\ & + D_{(ps+1)+1} L_d(q_{k+1}^{0,s}, \dots, q_{k+1}^{ps,s}, q_{k+1}^{0,f}, \dots, q_{k+1}^{pf,f}) \qquad k = 0, \dots, N-2 \\ 0 =& D_{(ps+1)+i} L_d(q_k^{0,s}, \dots, q_k^{ps,s}, q_k^{0,f}, \dots, q_k^{pf,f}) \quad i = 2, \dots, pf \qquad k = 0, \dots, N-1 \end{aligned} \tag{2}$$

with D_α being the partial derivative of the discrete Lagrangian w.r.t. the α-th argument.
The discrete Legendre transforms $\mathbb{F}_{L_d}^{\pm} : Q \times Q \to T^*Q$ read

$$\mathbb{F}_{L_d}^{+} : \begin{pmatrix} (q_{k-1}^{0,s}, q_{k-1}^{ps,s}) \mapsto (q_{k-1}^{ps,s}, p_{k-1}^{ps,s}) & = & (q_{k-1}^{ps,s}, D_{ps+1} L_d(q_{k-1})) \\ (q_{k-1}^{0,f}, q_{k-1}^{pf,f}) \mapsto (q_{k-1}^{pf,f}, p_{k-1}^{pf,f}) & = & (q_{k-1}^{pf,f}, D_{(ps+1)+pf+1} L_d(q_{k-1})) \end{pmatrix}$$

$$\mathbb{F}_{L_d}^{-} : \begin{pmatrix} (q_k^{0,s}, q_k^{ps,s}) \mapsto (q_k^{0,s}, p_k^{0,s}) & = & (q_k^{0,s}, -D_1 L_d(q_k)) \\ (q_k^{0,f}, q_k^{pf,f}) \mapsto (q_k^{0,f}, p_k^{0,f}) & = & (q_k^{0,f}, -D_{(ps+1)+1} L_d(q_k)) \end{pmatrix}$$

Note that along the solution of the discrete Euler-Lagrange equations (2) it holds that $(p_{k-1}^{ps,s}, p_{k-1}^{pf,f}) = (p_k^{0,s}, p_k^{0,f}) = p_k^0$ and therefore the conjugate momentum p_k^0 is uniquely defined.
For given configurations $q_k = (q_k^s, q_k^f) = (q_k^{0,s}, \dots, q_k^{ps,s}, q_k^{0,f}, \dots, q_k^{pf,f})$ the unknown configurations $q_{k+1}^{i,s}$, $i = 1, \dots, ps$ and $q_{k+1}^{i,f}$, $i = 1, \dots, pf$ in $q_{k+1} = (q_{k+1}^s, q_{k+1}^f) = (q_{k+1}^{0,s}, \dots, q_{k+1}^{ps,s}, q_{k+1}^{0,f}, \dots, q_{k+1}^{pf,f})$ are obtained by solving the discrete Euler-Lagrange equations (2). Replacing the first and the third equation in (2) by the discrete Legendre transforms

$$\begin{aligned} p_0^{0,s} &= -D_1 L_d(q_0) \\ p_0^{0,f} &= -D_{(ps+1)+1} L_d(q_0) \end{aligned}$$

provides a set of equations to determine the unknown configurations in q_0 in the first time step for given initial configuration $q(0) = q_0^0 = (q_0^{0,s}, q_0^{0,f})$ and momentum $p(0) = p_0^0 = (p_0^{0,s}, p_0^{0,f})$.

Varying the quadrature formulas (in this work the Gauss and the Lobatto quadrature are used) and its orders and varying the degrees ps and pf of the polynomials, provides a high number

of possibilities to construct the integrators. The abbreviations $ordT$, $ordV$, $ordW$ are introduced to denote the order of the quadrature formula that approximates the integral of the kinetic energy T, of the slow potential V and of the fast potential W, respectively. Furthermore, if $ordT = ordV = ordW$, the notation $ordTVW$ is used, respectively if $ordT = ordW$ the notation $ordTW$. Finally, an integrator is indicated as $[ps_pf_ordT_$(Gau / Lob)$ordV_$(Gau / Lob) $ordW_$(Gau / Lob)] with the abbreviations Lob for Lobatto and Gau for Gauss quadrature.

Examples

- $[ps2pf5ordTW10(\text{Gau})ordV4(\text{Lob})]$ denotes the integrator that is obtained when the polynomial approximating the slow variables is of degree 2, the polynomial approximating the fast variables is of degree 5. For the approximation of the integral of the kinetic energy T and the fast potential W, the same quadrature formula is used, i.e. the Gauss quadrature of order 10, whereas the integral of the slow potential is approximated via the Lobatto quadrature of order 4.

- $[ps1pf1ordTW2(\text{Gau})ordV2(\text{Lob})]$ yields an integrator treating the slow potential explicitly and the fast potential implicitly. This integrator is called the IMEX (implicit-explicit) method, introduced and analyzed in [10].

- Consider a linear polynomial q_d^s and a polynomial q_d^f of degree pf to approximate the slow variable q^s and the fast variable q^f, respectively. Suppose the kinetic energy T and the fast potential W are both approximated via the Gauss quadrature of order $2pf$. For the approximation of the slow potential, different quadrature formulas are considered. Using the Lobatto quadrature of order 2 yields the integrator $[ps1pf_ordTW12(\text{Gau})ordV2(\text{Lob})]$ with the discrete slow potential

$$V_d\left(q_k^s, q_k^f\right) = \frac{h}{2}\left[V\left(q_k^{0,s}, q_k^{0,f}\right) + V\left(q_k^{1,s}, q_k^{pf,f}\right)\right]$$

  The corresponding discrete conservative slow forces in the discrete Euler-Lagrange equations are explicit, i.e. they only depend on known configurations. When the Gauss quadrature of order 2 is applied, the integrator is $[ps1pf_ordTW12(\text{Gau})ordV2(\text{Gau})]$. The discrete slow potential then reads

$$V_d\left(q_k^s, q_k^f\right) = hV\left(\frac{q_k^{0,s} + q_k^{1,s}}{2}, q_d^f\left(\frac{h}{2}; q_k^f, h\right)\right)$$

  and the corresponding discrete conservative slow forces in the discrete Euler-Lagrange equations are implicit, meaning they depend on known and unknown configurations.

3 NUMERICAL EXAMPLE

3.1 Fermi-Pasta-Ulam problem

The performance of the variational integrators of mixed order is investigated numerically by means of the Fermi-Pasta-Ulam (FPU) problem (see e.g. [5]). The FPU model consists of $2l$ unit masses that are chained together by alternating soft nonlinear and stiff linear springs with fixed end points. Appropriate coordinates are chosen that enable the separation of the multirate system in slow and fast variables. The scaled displacement of the i-th stiff spring's

center is described by the slow variables q_i^s, $i = 1, \dots, l$, while the fast variables q_i^f, $i = 1, \dots, l$ represent the scaled length of the i-th stiff spring. The Lagrangian reads

$$\begin{aligned} L(q, \dot{q}) =& \frac{1}{2} \sum_{i=1}^{l} \left((\dot{q}_i^s)^2 + (\dot{q}_i^f)^2 \right) + \\ & - \frac{1}{4} \left[(q_1^s - q_1^f)^4 + \sum_{i=1}^{l-1} (q_{i+1}^s - q_{i+1}^f - q_i^s - q_i^f)^4 + (q_l^s + q_l^f)^4 \right] - \frac{\omega^2}{2} \sum_{i=1}^{l} (q_i^f)^2 \end{aligned}$$

The first two terms in the Lagrangian, the kinetic energy and the slow potential V, are both depending on the complete configuration variable, whereas the fast potential W, i.e. the third term, depends on the fast variable only. The stiffness $\omega \in \mathbb{R}$ in W is supposed to be large. The FPU problem is a multirate system because it shows different behaviour on different time scales [5]. The vibration of the stiff springs happens on the time scale ω^{-1}. On the time scale ω^0, the motion of the soft nonlinear springs takes place and over intervals exponentially long in ω, an energy exchange among the stiff springs occurs. For the simulations, the FPU model consists of 6 point masses, i.e. $l = 3$, with mass $m = 1$ and the stiffness of the stiff springs is $\omega = 50$. All initial values are zero despite the displacement and the extension of the first stiff spring and their velocities with $q_1^s(0) = 1$, $q_1^f(0) = \omega^{-1}$, $\dot{q}_1^s(0) = 1$, $\dot{q}_1^f(0) = 1$.

The performance of the variational integrators of mixed order is demonstrated by means of the motion of the first stiff spring and the energy exchange between the stiff springs over a long term simulation. The solution of the standard variational integrator $[ps1pf1ordTVW2(\text{Gau})]$, i.e. the midpoint-rule, with time step $h = 10^{-5}$ serves as a reference. Moreover, a numerical analysis regarding efficiency versus accuracy is carried out.

In Fig. 1, the motion of the first stiff spring's center q_1^s and the oscillation of the first stiff spring's extensions q_1^f are shown. The reference solutions are always plotted in solid turquoise. In the left hand side plots in Fig. 1, one can see in purple dashed the simulation results of the standard variational integrator $[ps1pf1ordTVW2(\text{Gau})]$ with time step $h = 0.1$. Increasing the degree of the polynomial for the fast part from one to six, while the approximation of the slow variable is still linear and using a quadrature formula of order 12 for T and W, yields the integrator $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$. Using the time step $h = 0.1$, the right hand side of Fig. 1 shows its discrete solution for the slow and fast configuration in purple dashed. With $h = 0.1$, both integrators resolve the motion of the slow variable q_1^s quite well, see top plots in Fig. 1 a) and b). The high degree of 6 of the polynomial approximating the fast configuration causes that the simulated fast motion nicely coincides with the reference solution, cf. lower plot in Fig. 1 b). Whereas, when using the linear polynomial for the fast part, the fast motion is not captured as can be seen in the lower plot in Fig. 1 a). Note that the integrator $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$ provides the fast configurations $q_k^f = (q_k^{0,f}, \dots, q_k^{6,f})$ for one time step $[kh, (k+1)h]$ such that with equidistant control points d_ν^f, $\nu = 0, \dots, 6$ the configurations $q_k^{i,f}$ at the time points $kh + ih/6$, $i = 0, \dots, 6$ are determined. Moreover, to resolve the high oscillations of the first stiff spring with a standard variational integrator, a smaller time step is needed. Compare Fig. 1 c) where the fast configuration q^f simulated via the standard variational integrator $[ps1pf1ordTVW2(\text{Gau})]$ with time step $h = 0.01$ is plotted. The high oscillation of q_1^f is captured, however the solution is shifted compared to the reference solution as typically seen for the standard variational integrators. This shifting-effect amplifies with increasing simulation time. No phase shifting occurs in the lower plot of Fig. 1 b), where the simulation results of the integrator $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$ are depicted, even after a long simulation time e.g. of 30 seconds (Fig. 1 d)).

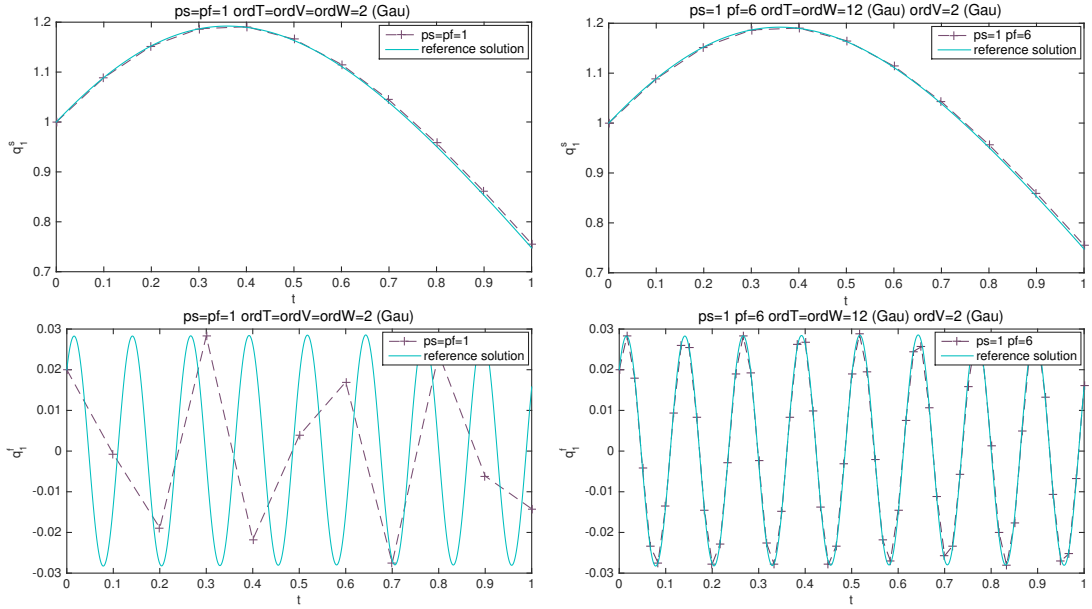
The energies of the stiff springs (red/purple/green) exchange over the time, while the sum of the stiff springs' energy (blue) remains close to a constant value, which is called an adiabatic invariant of the Hamiltonian system [5]. Fig. 2 is the reference solution. The plots in Fig. 3 show the simulation results of the integrators $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$, $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Lob})]$ and $[ps1pf6ordTVW12(\text{Gau})]$. In the left hand side plots of Fig. 3, the time step $h = 0.1$ is used for the simulations. The use of a quadrature formula of order 2 for the approximation of the slow potential, regardless of being the Gauss or the Lobatto quadrature, leads to high deviations in the oscillatory energies and their sum (Fig. 3 a) c)). However, both methods give qualitatively the correct rate of energy exchange and the total energy shows oscillations but no numerical drift. Increasing the order of $ordV$ to 12, the deviations of the oscillatory energies decrease and have the same magnitude as in the reference solution. The exchange of the stiff springs' energies is captured and the total oscillatory energy is nearly constant, see Fig. 3 e). Choosing $h = 0.2$ as time step, the performance of the integrators degrades, but the qualitative structure of the energy exchange behaviour is still maintained, see Fig. 3 b) d) f).

The computational efficiency of the variational integrators of mixed order is investigated numerically. The plots in Fig. 4 show the global errors of the slow configuration q^s (left) and of the fast configuration q^f (right) for the time steps $[0.2, 0.1, 0.05, 0.01, 0.005]$ against the run-times. The global errors of the configurations are determined by

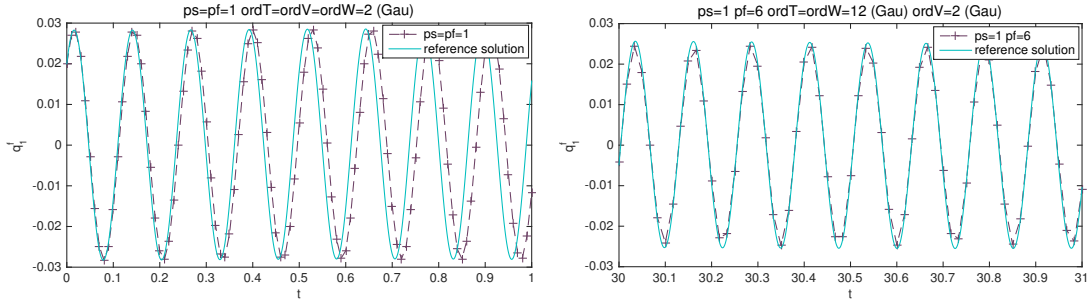
$$\max_{k\in\{0,\dots,N\}} \left\| q_k^{0,s} - q_{ref}^s(kh) \right\| \qquad \max_{k\in\{0,\dots,N\}} \left\| q_k^{0,f} - q_{ref}^f(kh) \right\|$$

where q_{ref}^s and q_{ref}^f are the configurations of the reference solution. Due to the limited accuracy of the reference solution itself, the global errors of the slow and the fast configuration can only be determined to a magnitude of approximately 10^{-8} and 10^{-7}, respectively. The simulation time is $T = 6$ seconds. Starting with the integrator $[ps6pf6ordTVW12(\text{Gau})]$ (blue dashed) in Fig. 4 a), the degree of the polynomial q_q^s approximating the slow variable is decreased successively from 6 to 1, while the degree of q_d^f is still 6 and the orders of the quadrature formulas remain 12. Reducing the degree ps of q_d^s decreases the number of unknowns in the Euler-Lagrange equations and thus the run-time decreases. Though, the accuracy of the solutions of the integrators with a lower degree ps remains nearly the same, which is nicely visible in both plots in Fig. 4 a) for the polynomial degrees $ps = 3, 4$ and 5. This result demonstrates the improved efficiency of the proposed mixed order integrators. However, decreasing ps even more causes a loss of accuracy of the simulation results, which becomes obvious for the degrees $ps = 1$ and $ps = 2$ in Fig. 4 a). To assess the computational efficiency of the integrators, the desired accuracy is crucial. For instance, when the focus lies on the fast configuration only with a lower accuracy being sufficient, the integrator $[ps1pf6ordTVW12(\text{Gau})]$ (red dashed) is the fastest in our implementation, see right plot in Fig. 4 a). More accurate results of the fast and the slow configuration in conjunction with shortest run-times are achieved by the integrator $[ps3pf6ordTVW12(\text{Gau})]$ (green dashed) in Fig. 4 a). Further savings in run-time can be reached, when the order $ordV$ decreases, as a low order $ordV$ means less quadrature nodes and therefore less costly function evaluations of the non-linear slow potential. Thus, successively reducing the order $ordV$ of the integrator $[ps1pf6ordTVW12(\text{Gau})]$ (Fig. 4 b)), respectively $[ps3pf6ordTVW12(\text{Gau})]$ (Fig. 4 c)), shortens the run-time even more with negligible influence on the accuracy, see Fig. 4 b) and c), where the errors for $ordV = 6, 8, 10, 12$ for $ps = 1$ ($ordV = 8, 10, 12$ for $ps = 3$, respectively) are of the same magnitude. However, decreasing the order $ordV$ too far leads to loss of accuracy in the configuration. First investigations indi-

cate that the loss is a result of convergence order reductions.
Fig. 5 shows the savings in run-time compared to $[ps6pf6ordTVW12(\text{Gau})]$ at a glance. If a lower accuracy of the fast configuration is sufficient, a decreased degree ps of 1 (red dashed) and in addition a low quadrature order $ordV = 4$ (yellow dash-dotted) bring shortest run-times in our implementation. A more accurate result can be gained cheaply with a decreased degree $ps = 3$ (green dashed) and most cheaply together with order $ordV = 8$ (purple dash-dotted). Furthermore, the run-time compared with the global error of q^f for the integrators $[ps1pf6ordTW12$ $(\text{Gau})ordV2$ $(\text{Gau})]$ (blue dash-dotted) and $[ps1pf6ordTW12$ $(\text{Gau})ordV2$ $(\text{Lob})]$ (blue dotted) and for the standard variational integrator $[ps1pf1ordTWV2(\text{Gau})]$ (turquoise dashed) are plotted in Fig. 5. It is nicely visible that treating the slow force explicitly (blue dotted) saves run-time compared to the implicit treatment of the slow force (blue dash-dotted). In comparison to the standard variational integrator, great improvements regarding efficiency versus accuracy are achieved with the presented integrators. The global error of the fast configuration simulated with the standard variational integrator $[ps1pf1ordTWV2(\text{Gau})]$ does not get smaller for the tested time steps between $h = 0.2$ and $h = 0.01$. The reason is the phase shifting that proceeds so far that the reference solution and the solution of the standard variational integrator are phase shifted by half a period causing a global error of two times the amplitude of the first spring's oscillation. Hence, the global error of the standard variational integrator $[ps1pf1ordTWV2(\text{Gau})]$ must be regarded in relation to the simulation time, as the shifting-effect amplifies with it.



a) $h = 0.1$ b) $h = 0.1$



c) $h = 0.01$ d) $h = 0.1$

Figure 1: FPU problem. Simulation results of the first slow q_1^s and first fast q_1^f configuration using a standard variational integrator (left) with time step $h = 0.1$ in a) and $h = 0.01$ in c) and using the integrator $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$ (right) with time step $h = 0.1$ in b) and d) in dashed purple, reference solution in solid turquoise.

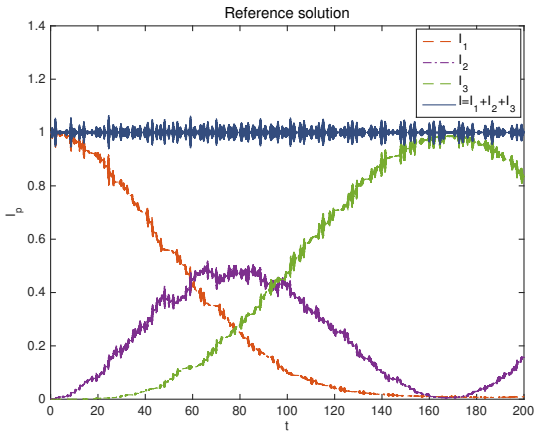


Figure 2: FPU problem. Energy of the three stiff springs (red dashed, purple dash-dotted, green dashed) and the total oscillatory energy (blue solid) of the reference solution.

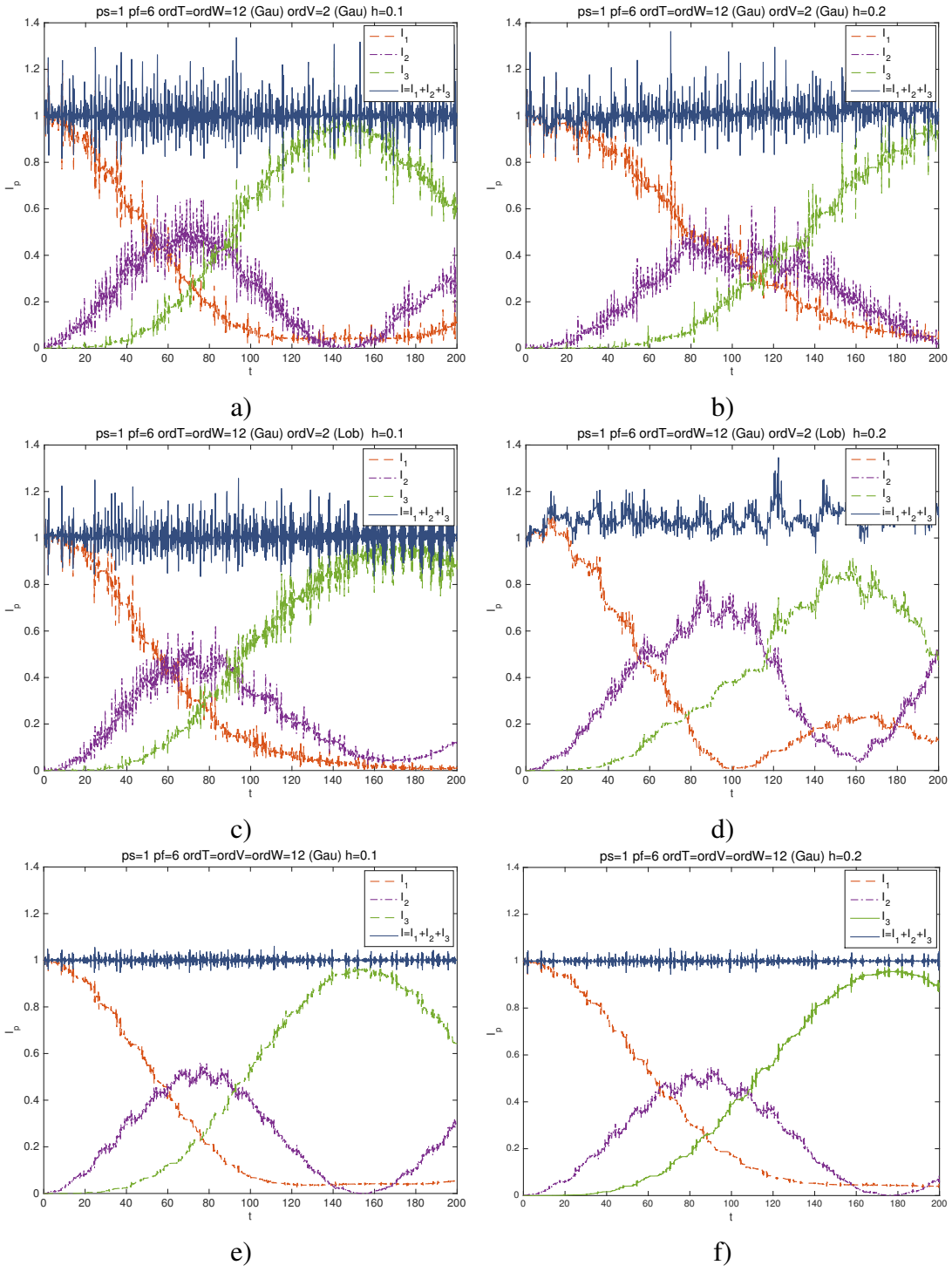
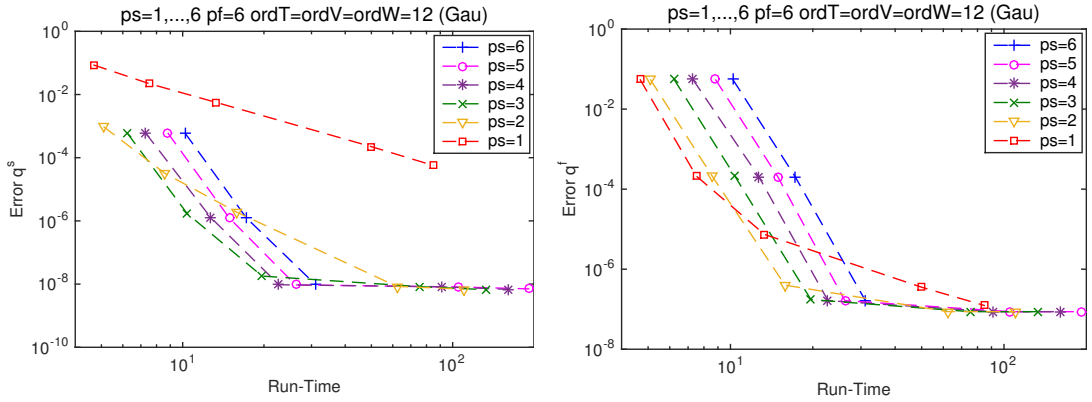
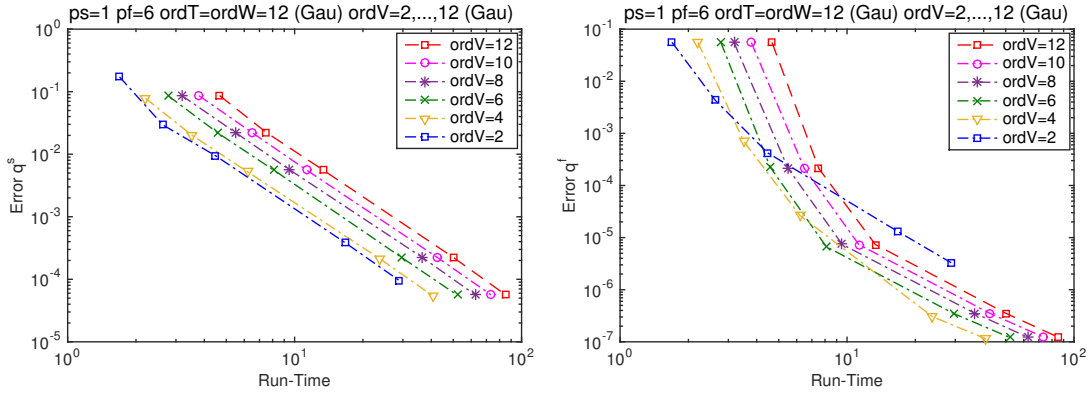


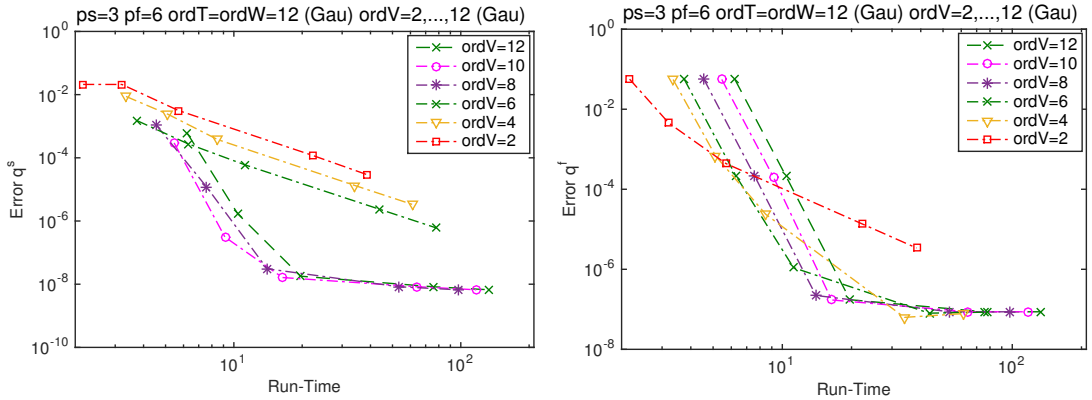
Figure 3: FPU problem. Energy of the three stiff springs (red dashed, purple dash-dotted, green dashed) and the total oscillatory energy (blue solid). Simulation results of the integrators $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Gau})]$, $[ps1pf6ordTW12(\text{Gau})ordV2(\text{Lob})]$ and $[ps1pf6ordTVW12(\text{Gau})]$ with time step $h = 0.1$ (left) and $h = 0.2$ (right).



a)



b)



c)

Figure 4: FPU problem. Global error of the slow configuration q^s (left) and the fast configuration q^f (right) against run-time for the simulation time of $T = 6$ seconds for the integrators $[ps1 \dots 6pf6ordTVW12(\text{Gau})]$ in a), $[ps1pf6ordTW12(\text{Gau})ordV2 \dots 12(\text{Gau})]$ in b) and $[ps3pf6ordTW12(\text{Gau})ordV2 \dots 12(\text{Gau})]$ in c) with time steps sizes $h \in \{0.2, 0.1, 0.05, 0.01, 0.005\}$.

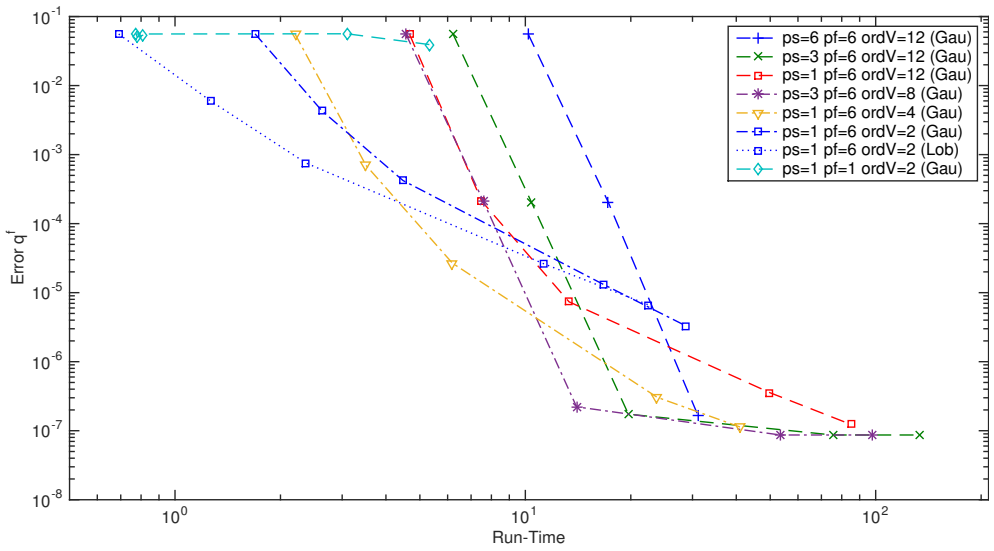


Figure 5: FPU problem. Global error of the fast configuration against run-time for the simulation time of $T = 6$ seconds for different integrators $[ps_pf_ordTW12(\text{Gau})ordV_(\text{Gau / Lob})]$ with time steps sizes $h \in \{0.2, 0.1, 0.05, 0.01, 0.005\}$.

4 CONCLUSION

The integrators of mixed order feature a great flexibility in design and thus in application. In this work, particular attention is paid to mechanical systems that act on different time scales caused by slow nonlinear and fast quadratic potentials. The trade-off between high accuracy and low computational costs, when simulating such mechanical systems, is resolved via a low degree polynomial for approximating the slow and a high degree polynomial for approximating the fast motion. The degrees of the polynomials have a direct impact on the number of unknowns in one time step and thus on run-time. Furthermore, due to its approximation via a high degree polynomial, the fast motion is captured and the typical phase shifting observed for the standard variational integrator is prevented. Considering the approximation of the integrals of the contributions of the Lagrangian separately enables the use of different quadrature formulas for each part and especially a low order quadrature formula for the integral of the slow potential reduces the number of costly function evaluations. This approach offers a possibility to decrease run-time with negligible loss in accuracy at once. However, higher savings in run-time come at a cost, as the accuracy suffers then. Therefore, the computational efficiency of the integrators has to be assessed with regard to the desired accuracy. The numerical investigation reveals that the multiscale phenomena of slow energy exchange and the conservation of adiabatic invariants are captured. However, the order of the quadrature formula approximating the integral of the slow energy influences the magnitude of the deviations of the oscillatory energies such that with increasing the order, the deviations decrease. The presented integrators base on a discrete version of Hamilton's principle. A detailed analysis of the inherited properties such as symplecticity and momentum preservation has to be done in the future, as well as a stability analysis. The convergence order of the integrators of mixed order are of interest, inter alia with regard to the observed losses in accuracy. Furthermore, a next step is to extend the approach to mechanical systems acting on multiple time scales with holonomic constraints. A comparison of the multirate variational integrators based on multiple time grids to the approach presented here is also subject of future works.

REFERENCES

[1] M. Arnold, Multi-rate time integration for large scale multibody system models. Proceedings of the IUTAM Symposium on Multiscale Problems in Multibody System Contacts, Stuttgart, Germany, February 20–23, 2006.

[2] C. Engstler and C. Lubich, Multirate extrapolation methods for differential equations with different time scales. *Computing*, **58**, 173–185, 1997.

[3] C. W. Gear and D. R. Wells, Multirate linear multistep methods. *BIT Numerical Mathematics*, **24**, 484–502, 1984.

[4] M. Günther, A. Kvaernø, and P. Rentrop, Multirate partitioned Runge-Kutta methods. *BIT Numerical Mathematics*, **41**, 504–514, 2001.

[5] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Vol. 31 of Springer Series in Computational Mathematics*. Springer, Berlin Heidelberg, 2006.

[6] B. Leimkuhler and S. Reich, *Simulating Hamiltonian dynamics, Vol. 14*. Cambridge University Press, 2004.

[7] S. Leyendecker and S. Ober-Blöbaum, A variational approach to multirate integration for constrained systems. In *Multibody dynamics*, 97–121, Springer, 2013.

[8] J. E. Marsden and M. West, Discrete mechanics and variational integrators. *Acta Numerica*, **10**, 357–514, 2001.

[9] S. Ober-Blöbaum and N. Saake, Construction and analysis of higher order Galerkin variational integrators. *Advances in Computational Mathematics*, **41**, 955–986, 2015.

[10] A. Stern and E. Grinspun, Implicit-explicit variational integration of highly oscillatory problems. *SIAM Multiscale Modeling & Simulation*, **7**, 1779–1794, 2009.

[11] M. Tao, H. Owhadi, and J. E. Marsden, Nonintrusive and structure preserving multiscale integration of stiff ODEs, SDEs, and Hamiltonian systems with hidden slow dynamics via flow averaging. *SIAM Multiscale Modeling & Simulation*, **8**, 1269–1324, 2010.

[12] E. Weinan, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden, Heterogeneous multiscale methods: a review. *Communications in Computational Physics*, **2**, 367–450, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

CO-SIMULATION OF LARGE MBD, FEA AND DEM SYSTEM

José L. Ortiz[1], **and Neil MacDonald**[2]

[1] MSC Software Corporation
201 Depot St. Suite 100, Ann Arbor, MI 48104, USA
jose.ortiz@mscsoftware.com

[2] Neil MacDonald
49 Queen St. Edinburgh EH2 3NH, UK
neil.macdonald@dem-solutions.com

Keywords: Co-simulation, multibody, finite elements, discrete elements.

Abstract. *An extension to an existing co-simulation algorithm [1] is presented. The original algorithm allows co-simulation of a large multibody dynamics (MBD) system with a large finite elements analysis (FEA) model; the present extension consists of a force-displacement co-simulation with a discrete element method (DEM) system.*

The proposed algorithm implements a force-displacement approach where the MBD system steps first using approximations for the forces provided by the DEM model. The DEM model later simulates its model using prescribed displacements provided by the MBD system.

As implemented in [1], there is no fixed communication interval. Each code (MBS, FEA and DEM) are free to take a simulation steps that fits well the current status of the numerical solution. The glue (master) code holds data at every step the co-simulationg codes take.

Implementation details and discussion follow.

1 INTRODUCTION

Co-simulation of large Multibody Systems (MBD) with both Finite Element models (FEA) and Discrete Element models (DEM) is nowadays a requirement from diverse industrial sectors which relay on simulation to speed up the design and development of vehicle, machinery and other devices. The present work is an extension of a recent development [1] where one large MBD system co-simulates with one or more FEA models. The extension consists of allowing the co-simulation with a DEM model.

The present work implements a basic force-displacement algorithm with the additional feature of computing a tangent stiffness matrix from the FEA models. The MBD model runs first with instantaneous forces provided by the FEA model and corresponding stiffness matrix. After the MBD system has advanced in time, the FEA model takes turn and advances in time with prescribed displacements at the points of interaction between the codes. Similarly, the co-simulation with the DEM model is implemented by requesting current forces at the interaction points. The MBD system simulates first using extrapolated values for the forces. Later, the DEM model simulates using prescribed motion at the interaction points. Currently, the DEM model does not support the computation of a tangent stiffness matrix at the interaction points.

2 ALGORITHM SPECIFICATIONS

As described in reference [1], the algorithm is based on a force-displacement interaction between the codes. Technical details of the case of co-simulation between an MBD model and multiple FEA models can be found in [1]. Here only main findings are presented as well as some details of the implementation to support a DEM model.

2.1 Interaction forces acting on the MBD system

The case of co-simulating with a FEA model is handled by applying a force on the MBD system given by the following equation:

$$\begin{Bmatrix} F_i \\ F_j \end{Bmatrix} = \begin{Bmatrix} F_{i0} \\ F_{j0} \end{Bmatrix} - \begin{bmatrix} K_{ij} \end{bmatrix} \begin{Bmatrix} U_i - U_{i0} \\ U_j - U_{j0} \end{Bmatrix}. \tag{1}$$

Equation (1) assumes there are two interaction points between the FEA model and the MBD model. The case of arbitrary number of interaction is similar. Forces *Fi* and *Fj* given by Equation (1) are applied on the MBD model to simulate the MBD system first. Here *Fio* and *Fjo* stand for the current value of the interaction forces (provided by the FEA model), *Uio* and *Ujo* stand for the displacements at the starting time step, *Ui* and *Uj* stand for the displacements at the interaction points *i*, and *j* respectively. Finally, *Kij* stands for the tangent stiffness matrix provided by the FEA code.

For the case of co-simulating with a DEM model, the interaction forces have the following expression (assuming two interaction points with the DEM model):

$$\begin{Bmatrix} F_i \\ F_j \end{Bmatrix} = \begin{Bmatrix} F_{ie}(t) \\ F_{je}(t) \end{Bmatrix}. \quad (2)$$

Functions *Fie(t)* and *Fje(t)* are continuous functions built with extrapolated forces obtained using the history of applied forces computed by the DEM model. The DEM model does not compute a stiffness matrix. The implementation allows selecting different extrapolation algorithms.

2.2 Variable communication algorithm

The original algorithm implements a variable communication algorithm (see reference [1]) between the co-simulating codes. The same implementation applies for the case of co-simulating with a DEM model. The basic idea is to allow all codes simulate using their best simulation setting (error settings, particular simulation options, etc.) without imposing a fixed communication interval. The master controller code keeps a history of the forces and displacements of all codes and communicates interpolated/extrapolated values to the co-simulating codes as needed.

2.3 Other features

The implemented algorithm permits all co-simulating codes run in parallel and execute in hosts across the network. Setup is relatively easy through the use of a configuration script. The configuration script is created by the user to define the topology of the co-simulating models and permits setting diverse options to tune the co-simulation. Options allow users to set different Newtonian reference frames for each model and set conversion factors when units are not consistent.

2.4 Limitations

The main limitation of the implemented algorithm is that it does not implement corrective measures when the relative error between computed forces by the MBD code (using either Equation (1) or Equation(2)) at the end of a simulation step, is greater than the actual forces computed by the FEA and DEM co-simulating models.

A second limitation is related to the topology of the co-simulating codes. Only co-simulation between the MBD model and the other models is supported. No FEA-FEA, nor FEA-DEM co-simulation is supported.

A third limitation is that users need to setup the simulation algorithms for each model independently. For instance, a simulation script is need for the MBD code, simulation data needs to be prepared in the FEA models, and simulation data needs to be included in the DEM model independently. Simulation data includes end time of simulation, error settings, etc. The end simulation must be consistent otherwise the co-simulation may be interrupted when a code finishes at an earlier time.

3 NUMERICAL EXAMPLES

The proposed algorithm has been successfully implemented in MSC Adams to co-simulate with MSC Marc and EDEM (flagship software from DEM Solutions). The following numerical examples show both prototype and industrial strength results obtained using the tool.

3.1 Oil pan damage

Figure 1 shows a co-simulation of full MBD model and a nonlinear FEA code. The box on the base of the car models a simple oil pan while the light-blue sphere simulates a stiff spherical rock hit during driving. The co-simulation is limited to low speed impact. The MBD model has no modelling limitations.

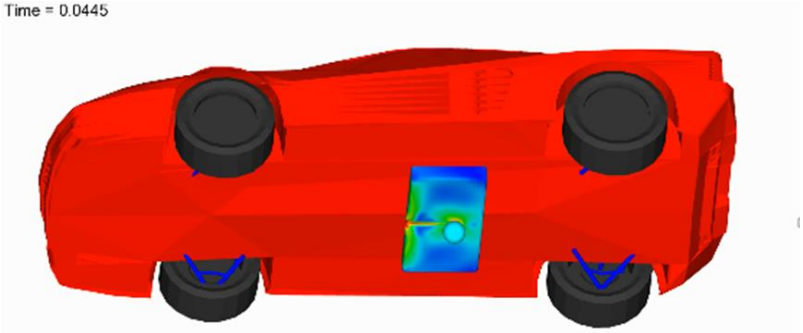


Figure 1: Oil pan damage.

3.2 Moderate misuse scenario

Figure 2 shows a co-simulation of a front suspension and tire MBD model hitting a curb. The control arm in dark blue is modelled in a FEA code and undergoes large nonlinear deformation and plasticity. Forces and displacements have been validated using a different approach showing excellent correlation.

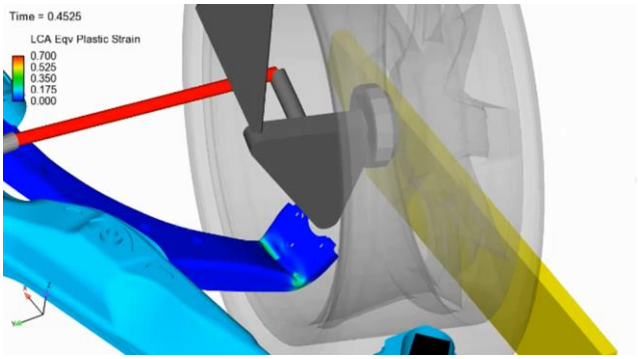


Figure 2: Moderate misuse scenario.

3.3 Torque modulator

Figure 3 shows the results of the simulations using the presented co-simulation algorithm (in red) versus the results using a full FEA model (in light blue). The results shown correspond to a torque modulator forces at a given control point. The full FEA model consists of a set of highly nonlinear coils inside a circular casing. The co-simulation model consists of the just the coils modelled in the FEA code while the casing and gadgets are modelled in the MBD code.

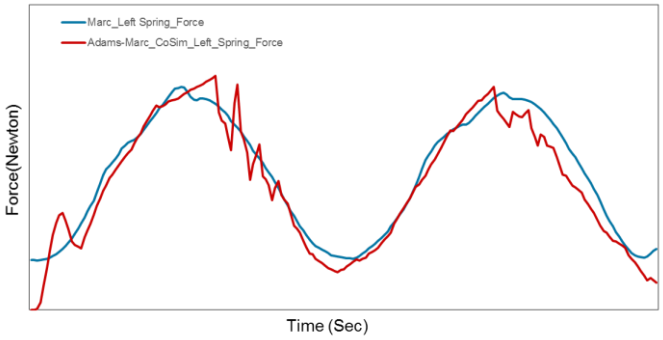


Figure 3: Torque modulator results.

3.4 Vehicle on granular terrain

Figure 4 shows a full ATV MBD model running on a bed of granular material. The MBD model co-simulation consisted of a static analysis to set the model on the ground followed by a dynamic simulation. Figure 5 shows a detail of one tire at time t=1 s.; notice some particles remained attached to the tire threads. The simulation lasted for 10 s and took approximately 20 hours to finish in a standard desktop computer.

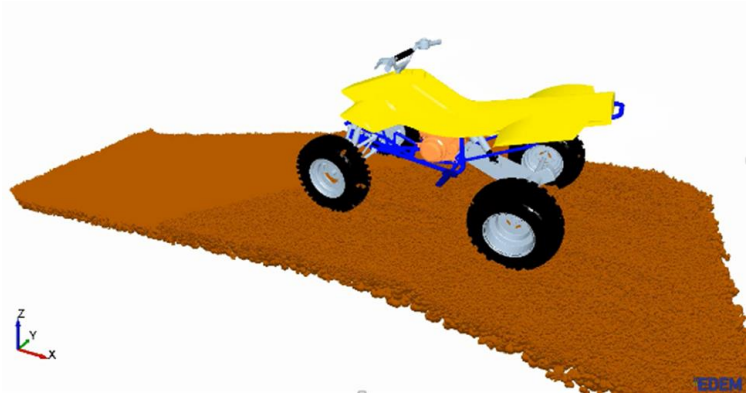


Figure 4: Vehicle on granular terrain.

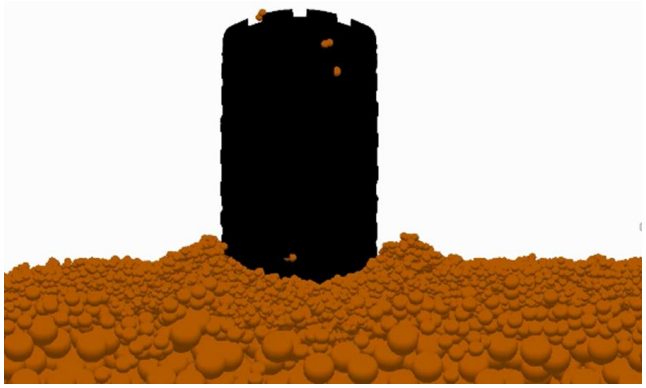


Figure 5: Detail of tire.

4 CONCLUSIONS

- Force-displacement co-simulation algorithms provide useful co-simulation results.
- Variable communication steps among co-simulating codes permits all codes run using their best simulation settings. Co-simulating codes constantly send their interaction data to the controller code.

REFERENCES

[1] I. Ishikawa, M. Tateishi, J. L. Ortiz. Variable communication and stiffness computation in co-simulation of large MBD and FEA systems. Proceedings of ECCOMAS 2015, Barcelona, Spain, July 2015

[2] E. Fahlgren, A. Elliot, M. Carlson. Knee Replacement Simulation Including Large Displacements and Viscoelastic Effects. Proceedings of BioMed 2008. June 18-20, Irvine, CA, USA.

[3] A. Elliot. A Highly Efficient General-Purpose Approach for Co-Simulation with ADAMS. Presented at the 15th European ADAMS Users Conference, November 2000, Rome, Italy.

[4] A. Elliot. Status Update on Advanced General-Purpose Co-Simulation with ADAMS. Presented at the ADAMS 2002 North American Users Conference, May 2002, Scottsdale, Arizona, USA.

[5] The FMI for Model Exchange and Co-simulation Specification v2.0.

[6] S. Sicklinger, V. Belsky, B. Engelmann, H. Elmqvist, H. Olsson, R. Wüchner[1]and, K.-U. Bletzinger. Interface Jacobian-based Co-Simulation. Int. J. Numer. Meth. Engng 2014; 98:418-444.

[7] C. Gear, D. Wells. Multirate linear multistep methods. BIT, 24, 484-502 (1984)

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

STOCHASTIC 3D MODELING OF AMORPHOUS MICROSTRUCTURES - A POWERFUL TOOL FOR VIRTUAL MATERIALS TESTING

Matthias Neumann[1], **Volker Schmidt**[1]

[1]Institute of Stochastics, Ulm University,
D-89069 Ulm, Germany
e-mail: matthias.neumann@uni-ulm.de, volker.schmidt@uni-ulm.de

Keywords: Virtual materials testing, Stochastic 3D modeling, Microstructure, Li-ion batteries, Solid oxide fuel cells, Effective conductivity.

Abstract. *We are going to introduce the concept of stochastic 3D modeling of geometrically complex (disordered) microstructures as a tool for virtual materials testing, including applications to battery electrodes as well as to electrodes of fuel cells, and solar cells. Using stochastic 3D models, one can generate a large variety of stochastically simulated micrsotructures with little computational effort. These virtual microstructures can be used as data basis to elucidate microstructure-property relationships. In this way, for example, effective conductivity can be expressed by microstructural characteristics such as volume fraction, tortuosity (windedness of transport paths) and costrictivity (bottleneck criterion) of the considered material phase. In another recent simulation study, we analysed more than 8000 virtual microstructures for various microstructural scenarios. Using data mining techniques like artificial neural networks and random forests, we were able to accurately predict effective conductivities given microstructure properties like volume fraction, tortuosity and constrictivity.*

1 INTRODUCTION

In many applications the functionality of materials strongly depends on their - mostly amorphous - microstructure. This is the case for, e.g. anodes in lithium-ion (Li-ion) batteries, which are a key technology for electric vehicles. The microstructure of these electrodes, consisting mainly of graphite particles and pores, strongly influences both, the transport of ions and electrons as well as degradation processes within the electrode, which are in turn important for the performance and lifetime of the battery. For microstructure effects in Li-ion batteries, see e.g. [1] and the references therein. Porous microstructures appear also in electrodes of solid oxide fuel cells (SOFC), a technology for electricity generation, which is more efficient, more reliable and has less environmental impact compared to conventional energy generation. Like in Li-ion batteries the porous microstructures in SOFC are of high relevance for the performance of these fuel cells as well as for their degradation processes. See e.g. [2] for the microstructure influence in SOFC electrodes where the solid phase consists of nickel (Ni) and yttria-stabilized zirconia (YSZ).

Due to this microstructural influence on the functionality of Li-ion batteries and SOFC, two issues have to be clarified in order to improve the functionality of these materials. At first, it is important to understand how the microstructure of the electrode is quantitatively related to the corresponding effective properties. The second issue is how the production parameters of the electrodes influence the microstructure.

The progress in tomographic 3D imaging during the last decades has enabled detailed microstructure analysis, see [3] for a methodological overview. For microstructure analysis in the context of Li-ion batteries and SOFC we refer, e.g., to [4] and [5]. By the aid of numerical modeling it is possible to simulate the corresponding effective properties, like effective conductivity. Although this combination of image analysis and numerical simulation allows a direct investigation of the relationship between well-defined microstructure characteristics and effective transport properties ([6] to [9]), it is strongly limited by the high costs of tomographic 3D imaging.

An alternative approach which combines image analysis and numerical simulations with stochastic 3D modeling of amorphous microstructures solves this problem. Tools from spatial statistics and stochastic geometry, see e.g. [10], are used to develop parametric stochastic 3D models for the generation of virtual but realistic microstructures on the basis of image data from real microstructures. Once developed, a stochastic 3D model can be used to approach both of the above mentioned issues. On the one hand, the parameters of the stochastic 3D model can be varied to obtain virtual microstructures within a wide range of microstructure characteristics. Then, numerical simulations can be applied to these virtual microstructures to simulate the corresponding effective properties, such that the quantitative relationship between microstructure characteristics and effective properties can be efficiently investigated, see e.g. [11] and [12]. On the other hand, in case that there is image data of microstructures available, where the microstructures are produced under different conditions (e.g. under different sinter temperatures like in [13]), the parameters of the stochastic 3D model might be fit to microstructures generated with different production parameters. An interpolation between model parameters and production parameters allows to predict virtual microstructures for production parameters where no 3D images are available. This was succesfully done for solar cells produced with different spin coating velocities in [14]. Once the relationship between production parameters and microstructure is understood one can find optimal production parameters by numerical simulations of effective properties.

Summarizing the above, stochastic 3D modeling can be used for investigating the relationship between microstructure characteristics and effective properties as well as the relationship between microstructure characteristics and production parameters. Where we test the functionality of virtual microstructures by simulating their effective properties, the combination of mathematical image analysis, stochastic 3D modeling of amorphous microstructures, and numerical modeling of (spatially resolved) effective properties on complex geometries is, what we refer to as virtual materials testing.

In the present paper we give an overview of stochastic 3D modeling of amorphous microstructures, which is essential for virtual materials testing. The paper is organized as follows. Models for electrode materials in Li-ion batteries ([15], [16]), and for SOFC ([13], [17]) are reviewed in Sections 2 and 3, respectively. In Section 4 we present a method to derive empirical formulas that describe the quantitative relationship between microstructure characteristics and effective conductivity in porous microstructures ([11], [12]). Section 5 concludes the paper.

2 LITHIUM-ION BATTERIES

In [16] a stochastic 3D model for the anode microstructure of so-called energy cells in Li-ion batteries has been developed on the basis of experimental image data obtained by X-ray tomography, see Figure 1. The anode mainly consists of a connected network of graphite particles and pores between the particles.

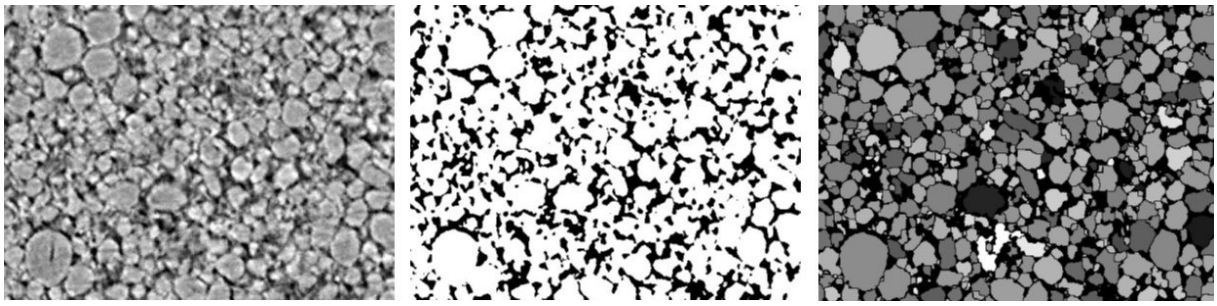


Figure 1: 2D slice of experimental image data: Grayscale image obtained by X-ray tomography (left) has been binarized (center). The application of the watershed transform has led to an extraction of single particles (right).[2]

2.1 Stochastic modeling of particle shapes

Since the shape of graphite particles observed in data is neither spherical nor ellipsoidal, the single particles in the considered anode material have been modeled by the so-called spherical harmonics expansion, see [15] and [18]. This approach is based on the fact that any (star-shaped) particle in 3D is completely determined by its barycenter and a function $r : [0,\pi] \times [0,2\pi) \longrightarrow [0,\infty)$, where $r(\theta,\phi)$ denotes the elongation of the particle in direction $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. By the aid of the so-called spherical harmonics functions $\{Y_l^m : [0,\pi] \times [0,2\pi) \longrightarrow [0,\infty) : l,m \geq 0\}$, the radius function r can be represented as

$$r(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta,\phi) \approx \sum_{l=0}^{L} \sum_{m=-l}^{l} c_l^m Y_l^m(\theta,\phi) \tag{1}$$

[2]Reprinted from [16], Figure 1, with permission from Elsevier.

for $L \geq 1$ large enough and spherical harmonics coefficients $c_l^m \in \mathbb{R}$ with $0 \leq l$ and $-l \leq m \leq l$. For details and the definition of the spherical harmonics functions Y_l^m the reader is referred to [15].

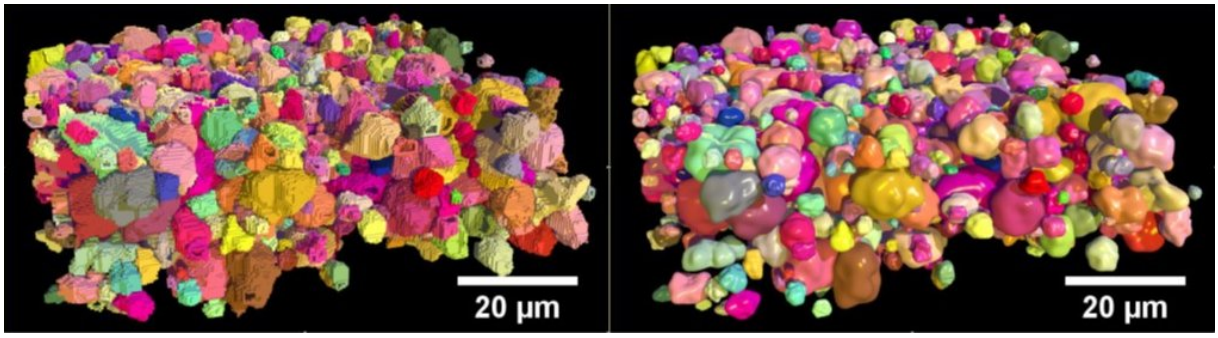


Figure 2: 3D voxel representation of graphite particles (left) and their spherical harmonics representation (right).[3]

Equation (1) implies that each (star-shaped) particle is completely determined by its spherical harmonics coefficients and, therefore, the particle can be approximated by a finite number of coefficients. In [15] a method for estimating L and the spherical harmonics coefficients $c_l^m \in \mathbb{R}$ with $0 \leq l \leq L$ and $-l \leq m \leq l$ from image data is described. This method has been successfully applied to experimental image data showing particles in battery anodes, see Figure 2. The single particles have been extracted from grayscale images with well-known methods from image analysis, to be more precise the watershed transform, see e.g. [19], has been used after binarization of the grayscale images, see Figure 1.

Since the shape of each graphite particle is appropriately described by a finite number of real-valued coefficients, the spherical harmonics representation leads to an enormous compression of information contained in the original (grayscale) image data.

2.2 Stochastic modeling of anode microstructure

In total, a parametric representation of image data is obtained by the method proposed in [15]. The anode material is completely described by the barycenters of particles and their spherical harmonics coefficients. In [16] statistical analysis of this parametric representation is the basis for stochastic 3D modeling of anode microstructures in Li-ion batteries.

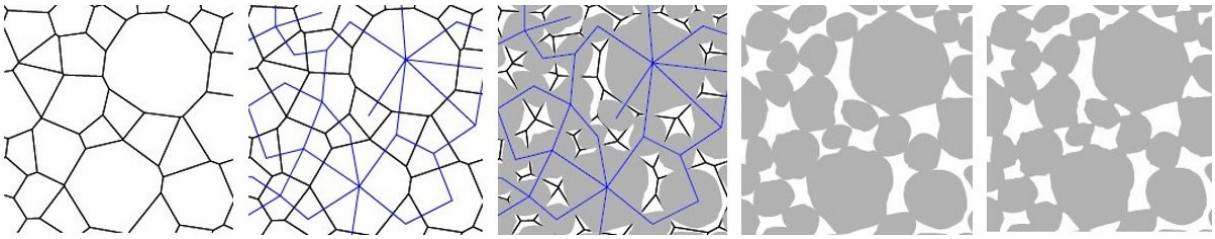


Figure 3: 2D sketch of the modeling idea: Random tesselation model determines the locations of particles (1). Connections between particles have been modeled by a random graph (2). Particle shapes have been modeled using spherical harmonics (3). Connected particle system before (4) and after morphological closing (5).[4]

[3]Reprinted from [15], Figure 11, with permission from Elsevier.

The model idea, sketched in Figure 3, is the following. At first the locations, at which the single particles are placed, have been modeled by a random Laguerre tessellation, see [20]. That is to say that the 3D space is partitioned into polytopes, the size distribution of which can be controlled by model parameters. In a second step, connections between single particles have been modeled by a completely connected random graph, i.e., edges are put between neighboring polytopes according to a certain probabilistic rule. Within each polytope a particle is placed using spherical harmonics, such that particles overlap if they are connected by an edge in the random graph. This graph model ensures that the modeled particle system is completely connected. Finally, morphological closing is applied to the particle system in order to smooth the sharp edges at the contacts between particles. Note that the volume fraction of the particle system can be adjusted by model parameters, which have been fitted to experimental image data.

Besides a visual comparison between virtual microstructures generated by the fitted model and experimental image data, see Figure 4, an extensive quantitative validation regarding certain microstructure characteristics has been performed and discussed in [16]. Exemplarily, the spherical contact distribution functions [10] of the particle phase, which are nearly identical for virtual and experimental microstructures, are shown in Figure 4.

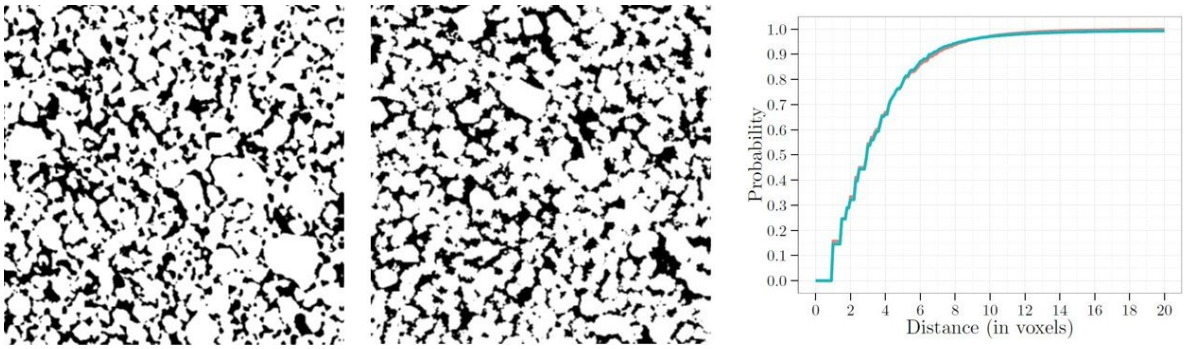


Figure 4: 2D slice of experimental image data from anode material (left) compared to a 2D slice of a virtual 3D microstructure generated by the stochastic model (center). Spherical contact distribution functions computed for experimental (red) and simulated (blue) microstructures are nearly identical (right).[5]

Thus, in [15] and [16] flexible modeling techniques have been developed for 3D microstructures in Li-ion batteries. In particular, a stochastic 3D model has been established which is able to generate virtual microstructures that are statistically similar to experimental microstructures. In a forthcoming paper, the model is further validated by numerical 3D electrochemical simulations, which are applied to experimental and virtual anodes [21]. Additionally, extensions of the stochastic microstructure model, which overcome small differences in electrochemical behaviour, are proposed. The model will be used for investigating the relationship between microstructure characteristics and the functionality of Li-ion batteries, i.e. for virtual materials testing.

3 SOLID OXIDE FUEL CELLS

In this section we summarize the modeling approaches proposed in [13] and [17]. While in [13] a stochastic microstructure model has been fitted to three cathode materials produced

[4]Reprinted from [16], Figure 2, with permission from Elsevier.
[5]Reprinted from [16], Figures 10 and 11, with permission from Elsevier.

with different sinter temperatures, the model in [17] is able to generate virtual, but realistic three-phase microstructures, where all three phases are completely connected.

3.1 Stochastic 3D modeling of $La_{0.6}Sr_{0.4}CoO_{3-\delta}$ cathodes

The considered $La_{0.6}Sr_{0.4}CoO_{3-\delta}$(LSC) cathodes consist of sintered spherical LSC particles, which form a connected particle system, and pores, see Figure 6. Thus the microstructure has been modeled by a completely connected union of slightly overlapping spheres. To begin with, a structural segmentation of the binarized image data, obtained by FIB-SEM tomography, has been performed to represent the LSC phase by a union of overlapping spheres. For this purpose, tools from image analysis like the watershed algorithm [19] and the Hough transform [22] are used.

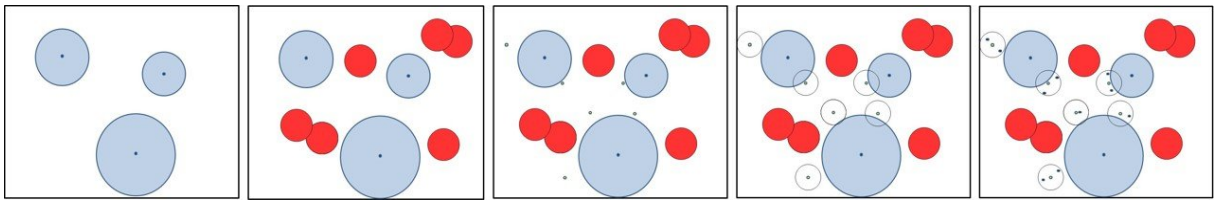


Figure 5: Modeling idea for the midpoints of the sphere system sketched in 2D: Non-overlapping large spheres are randomly distributed (1), where pore former (red circles) is modeled in the remaining space (2). Cluster centers of small spheres (3) determine regions (4) in which the midpoints of small spheres are distributed (5).

Like in the case of battery anodes considered in Section 2, this kind of structural segmentation can be understood as a compression of image data, since the microstructure has been appropriately represented by the midpoints and radii of the spheres. In a second step a model for random sphere systems, a so-called germ-grain model [10], is developed which generates sphere systems that are similar to those extracted from image data.

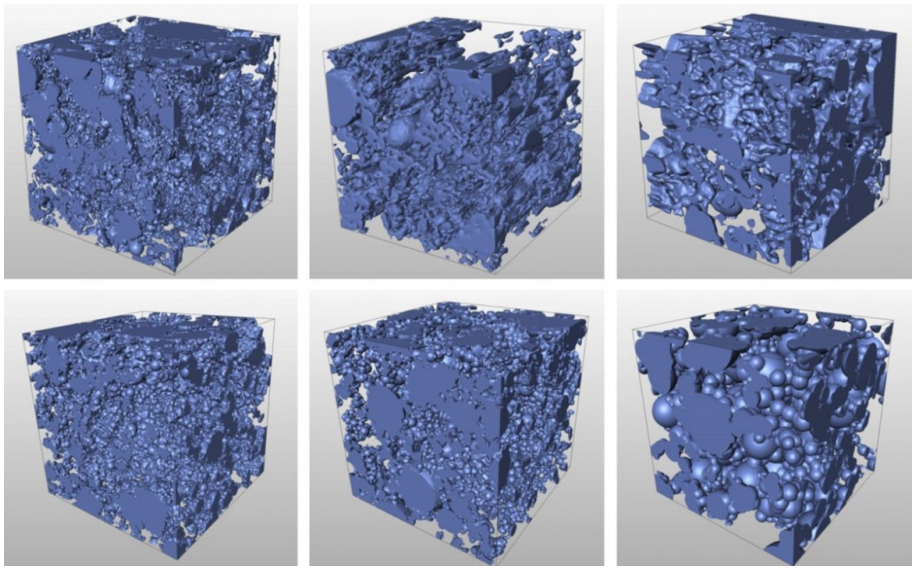


Figure 6: 3D images of real (top) and virtual microstructures (bottom). The structures correspond to different sinter temperatures (from left to right: $750°C, 850°C, 950°C$).[6]

The modeling of sphere midpoints takes into account that small particles are accumulated around large spheres and that large void regions occur due to pore former. The sphere midpoints have been modeled in three steps using random point processes, see e.g. [10]. At first large spheres are distributed completely at random under the condition that they do not overlap each other. Then, the void regions are determined according to a Boolean model [23] with spherical grains in the remaining space. Finally the midpoints of small spheres are clustered around the large spheres, see Figure 5. The radii of the small spheres are determined such that volume fraction and connectivity properties of the sphere systems extracted from experimental image data are statistically matched. Similar as in [16] a graph model is used for this purpose. For further details we refer to [13].

The model has been fitted to three cathodes, produced with three different sinter temperatures of $750°\mathrm{C}$, $850°\mathrm{C}$ and $950°\mathrm{C}$. Figure 6 shows a good visual accordance between image data and virtual structures generated by the model. Furthermore, in [13] a quantitative comparison of microstructure characteristics is performed and, in combination with numerical modeling, the influence of different sinter temperatures on the electrode performance can be investigated. The achieved goodness-of-fit shows a high flexibility of the model since all three microstructures produced with different sinter temperatures can be described by one and the same model type, although their microstructure characteristics differ strongly from each other.

3.2 Stochastic 3D modeling of three-phase Ni-YSZ anodes

In Sections 2.2 and 3.1 the material to be modeled consists of pores and mainly one single (homogeneous) solid phase. In this section a model for three-phase microstructures with two different solid (sub-) phases is described, which has been introduced in [17]. The model exhibits the desirable property that it reproduces microstructures where all three phases are completely connected. Three-phase microstructure modeling is important since not only the geometry of single phases influences the functionality of Ni-YSZ anodes, but also the length and the geometry of the triple phase boundary (TPB), at which chemical reactions take place.

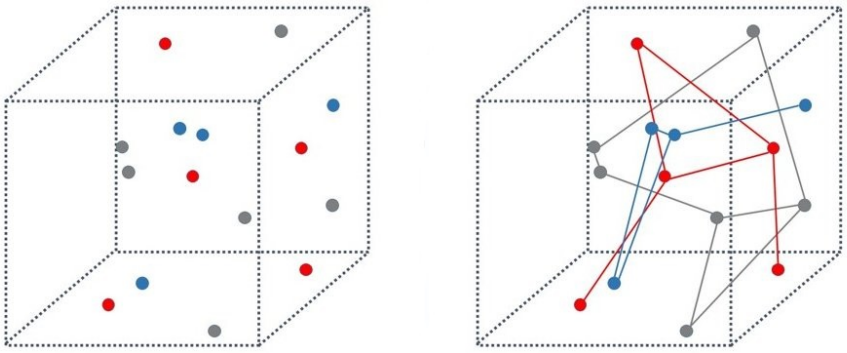


Figure 7: Modeling idea: Three random point patterns (consisting of blue, gray and red points, respectively) are distributed completely at random in the three-dimensional space (left). On each of these point patterns a geometric graph is modeled (right).

A parametric random graph in the three-dimensional space has been modeled for each phase, the vertices of which are distributed completely at random, see Figure 7. These graphs build the backbones of the three phases, which are Voronoi tessellations with respect to the edge set

[6]Reprinted from [13], Figure 15, with permission from Elsevier.

of the three graphs, i.e. each point $x \in \mathbb{R}^3$ is allocated to the phase the corresponding graph of which is closer to x than the other two graphs.

By this approach the model is able to mimic connectivity properties of experimental image data gained by FIB-tomography. It is even possible to simulate virtual microstructures with arbitrary volume fractions and completely connected phases. A slight generalization of the model has enabled a good fit to the experimental image data recently discussed in [24], see Figure 8.

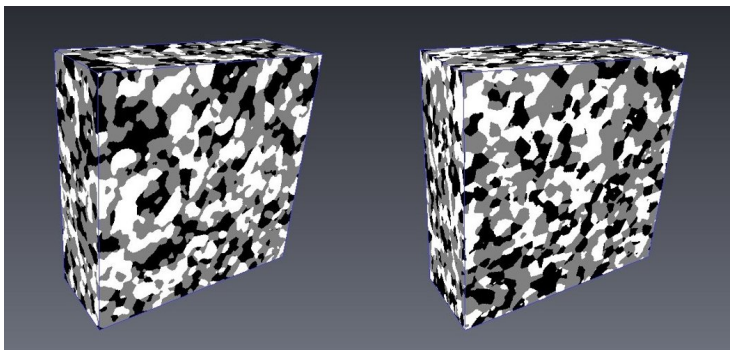


Figure 8: Image data of real (left) and virtual (right) Ni-YSZ anodes.

Simulations of the effective conductivity σ_{eff} for image data of real and virtual Ni-YSZ anodes by means of the finite element method (FEM) has shown that the model is close to the real microstructures with respect to effective properties. Besides the development of new techniques for stochastic 3D modeling of three-phase microstructures, the model proposed in [17] is the basis of virtual materials testing for Ni-YSZ anodes. In a further work the area specific resistance (ASR) of virtual Ni-YSZ microstructures with systematically varied microstructure characteristics is simulated by a FE model for the anode reaction mechanism that takes into account the specific microstructure characteristics, in particular the length of TPB. Performing this kind of virtual materials testing we expect new insights about the microstructure influence on the ASR.

4 VIRTUAL MATERIALS TESTING

In the previous sections stochastic 3D models have been presented, which are fitted to image data of real microstructures in order to study the microstructure of various energy materials. Moreover, it is of general interest to investigate quantitative relationships between microstructure characteristics and effective properties. In this section we focus on the so-called M-factor, a normed measure for the microstructure influence on effective conductivity. It is defined by

$$M = \frac{\sigma_{\text{eff}}}{\sigma_0}, \tag{2}$$

where σ_0 denotes the intrinsic conductivity of the material. Virtual materials testing contributes to find an expression of the M-factor in terms of microstructure characteristics ([11], [12]).

Leaned on [25] the investigations in [11] and [12] assumed that volume fraction ε, tortuosity τ measuring the length of (shortest) transport paths and a constriction factor β measuring the strength of bottleneck effects are the most important characteristics for effective conductivity in porous microstructures. For tortuosity, two different approaches, i.e., mean geometric and

mean geodesic tortuosity denoted by τ_{geom} and τ_{geod} have been taken into account [12]. The definition of β is based on the concept of continuous phase size distribution and a geometric simulation of mercury intrusion porosimetry [26].

The idea was to develop a stochastic 3D model which is flexible regarding the microstructure characteristics $\varepsilon, \tau_{\mathrm{geom}}, \tau_{\mathrm{geod}}$ and β. Then, σ_{eff}, and thus the M-factor, of virtual microstructures with different constellations for $\varepsilon, \tau_{\mathrm{geom}}, \tau_{\mathrm{geod}}$ and β can be simulated by FEM. The generated database has enabled an efficient investigation of the quantitative microstructure influence on the M-factor.

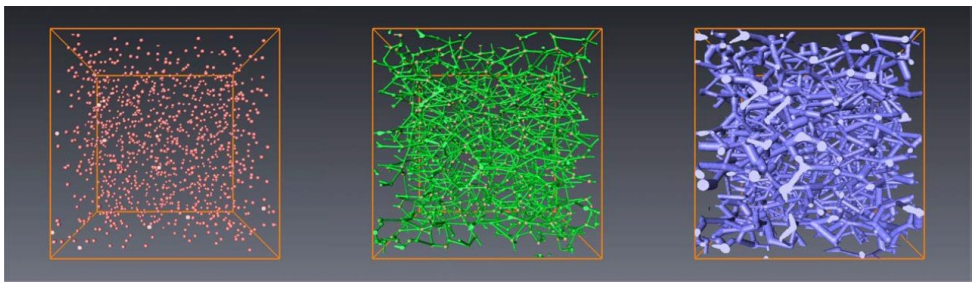


Figure 9: Modeling idea: Vertices are distributed completely at random (left), before edges are put (center). Random dilation of edges leads to virtual microstructures (right).[7]

The stochastic 3D model is based on an anisotropic random graph model, where the anisotropy can be controlled by model parameters. The vertices of the graph are distributed completely at random in the three-dimensional space. The edges of the graph are randomly dilated, where the variance of the dilation radii influences the constrictivity of the microstructure strongly. Note that the model graph is completely connected which ensures complete connectivity of the conducting phase. The modeling idea is visualized in Figure 9.

By means of the stochastic 3D model 49 virtual microstructures for a wide range of ε, τ_{geom}, τ_{geod} and β have been generated and the corresponding effective conductivites are simulated by FEM. Using error-minimization the relationship

$$M = \min\left\{1, 2.03\frac{\varepsilon^{1.57}\beta^{0.72}}{\tau_{\mathrm{geom}}^2}\right\} \tag{3}$$

has been established in [11] and further refined in [12] using the concept of geodesic tortuosity instead of geometric tortuosity. This has led to

$$M = \frac{\varepsilon^{1.15}\beta^{0.37}}{\tau_{\mathrm{geod}}^{4.39}}. \tag{4}$$

Both relationships fit better to the virtual microstructures than the existing ones from literature. Furthermore, validation with experimental image data has been performed in [12], which shows that Equation (4) derived by virtual materials testing is able to predict the M-factor of real microstructures quite well by the microstructure characteristics $\varepsilon, \tau_{\mathrm{geod}}$ and β.

A forthcoming paper [27] is under preparation in which an extensive simulation study with around 8000 virtual microstructures is performed. In this work methods from statistical learning [28] are combined with virtual materials testing for further improvement of quantitative relationships between microstructure characteristics and the M-factor.

[7] Reprinted from [11], Figure 3, with permission from J. Wiley & Sons.

5 CONCLUSION

In the present survey paper the concept of virtual materials testing is presented. It is a method, which combines mathematical image analysis, stochastic 3D modeling of amorphous microstructures, and numerical modeling of (spatially resolved) effective properties on complex geometries for an efficient investigation of the relationship between microstructure characteristics of functional materials and their effective properties.

The focus of the present paper is on stochastic 3D modeling of microstructures in Li-ion batteries and SOFC. In Section 4 it is described how quantitative relationships between the microstructure characteristics volume fraction, tortuosity, constrictivity on the one hand and effective conductivity on the other hand can be established. Virtual materials testing is neither restricted to Li-ion batteries and SOFC nor to effective conductivity. Virtual materials testing opens the possibility to investigate amorphous structures of various further materials, e.g., in polymers solar cells [12] or aluminium alloys [29] as well as various further effective properties like e.g. exciton quenching efficiency [14] or mechanical stress-strain curves [30].

REFERENCES

[1] J. R. Wilson, J. S. Cronin, S. A. Barnett, S. J. Harris. Measurement of three-dimensional microstructure in a $\mathrm{LiCoO_2}$ positive electrode. *Journal of Power Sources*, **196**, 3443–3447, 2011.

[2] B. S. Prakash, S. S. Kumar, S. T. Aruna. Properties and development of Ni/YSZ as an anode material in solid oxide fuel cell: A review. *Renewable and Sustainable Energy Reviews*, **36**, 149–179, 2014.

[3] S. Torquato. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, New York, 2013.

[4] P. R. Shearing, L. E. Howard, P. S. Jørgensen, N. P. Brandon, S. J. Harris. Characterization of the 3-dimensional microstructure of a graphite negative electrode from a Li-ion battery. *Electrochemistry Communications*, **12**, 374– 377, 2010.

[5] J. R. Wilson, W. Kobsiriphat, R. Mendoza, H.-Y. Chen, J. M. Hiller, D. J. Miller, K. Thornton, P. W. Voorhees, S. B. Adler, S. A. Barnett. Three-dimensional reconstruction of a solid-oxide fuel-cell anode. *Nature Materials*, **5**, 541–544, 2006.

[6] M. Doyle, T. F. Fuller, J. Newman. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical Society*, **140**, 1526–1533, 1993.

[7] L. Holzer, B. Iwanschitz, T. Hocker, L. Keller, O. M. Pecho, G. Sartoris, P. Gasser, B. Muench. Redox cycling of Ni–YSZ anodes for solid oxide fuel cells: Influence of tortuosity, constriction and percolation factors on the effective transport properties. *Journal of Power Sources*, **242**, 179–194, 2013.

[8] N. Shikazono, D. Kanno, K. Matsuzaki, H. Teshima, S. Sumino, N. Kasagi. Numerical assessment of SOFC anode polarization based on three-dimensional model microstructure reconstructed from FIB-SEM images. *Journal of the Electrochemical Society*, **157**, B665–B672, 2010.

[9] S. Tippmann, D. Walper, L. Balboa, B. Spier, W. G. Bessler. Low-temperature charging of lithium-ion cells part I: Electrochemical modeling and experimental investigation of degradation behavior. *Journal of Power Sources*, **252**, 305–316, 2014.

[10] S. N. Chiu, D. Stoyan, W. S. Kendall, J. Mecke. *Stochastic Geometry and its Applications*. J. Wiley & Sons, Chichester, 3rd edition, 2013.

[11] G. Gaiselmann, M. Neumann, O. M. Pecho, T. Hocker, V. Schmidt, L. Holzer. Quantitative relationships between microstructure and effective transport properties based on virtual materials testing. *AIChE Journal*, **60**, 1983–1999, 2014.

[12] O. Stenzel, O. M. Pecho, M. Neumann, V. Schmidt, L. Holzer. Predicting effective conductivities based on geometric microstructure characteristics. *AIChE Journal, in print*, 2016.

[13] G. Gaiselmann, M. Neumann, L. Holzer, T. Hocker, M. Prestat, V. Schmidt. Stochastic 3D modeling of $\mathrm{La_{0.6}Sr_{0.4}CoO_{3-\delta}}$ cathodes based on structural segmentation of FIB-SEM images. *Computational Materials Science*, **67**, 48–62, 2013.

[14] O. Stenzel, L. Koster, R. Thiedmann, S. D. Oosterhout, R. A. J. Janssen, V. Schmidt. A new approach to model-based simulation of disordered polymer blend solar cells. *Advanced Functional Materials*, **22**, 1236–1244, 2012.

[15] J. Feinauer, A. Spettl, I. Manke, S. Strege, A. Kwade, A. Pott, V. Schmidt. Structural characterization of particle systems using spherical harmonics. *Materials Characterization*, **106**, 123–133, 2015.

[16] J. Feinauer, T. Brereton, A. Spettl, M. Weber, I. Manke, V. Schmidt. Stochastic 3D modeling of the microstructure of lithium-ion battery anodes via gaussian random fields on the sphere. *Computational Materials Science*, **109**, 137–146, 2015.

[17] M. Neumann, J. Staněk, O. M. Pecho, L. Holzer, V. Beneš, V. Schmidt. Stochastic 3D modeling of complex three-phase microstructures in SOFC-electrodes with completely connected phases. *Computational Materials Science, under revision*, 2016.

[18] E. J. Garboczi. Three-dimensional mathematical analysis of particle shape using X-ray tomography and spherical harmonics: Application to aggregates used in concrete. *Cement and Concrete Research*, **32**, 1621–1638, 2002.

[19] J. B. T. M. Roerdink, A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, **41**, 187–228, 2000.

[20] C. Lautensack, S. Zuyev. Random Laguerre tessellations. *Advances in applied probability*, **40**, 630–650, 2008.

[21] S. Hein, J. Feinauer, D. Westhoff, I. Manke, V. Schmidt, A. Latz. 3D electrochemical simulations of experimental and virtual anodes in lithium-ion batteries. *Working paper*, 2016.

[22] B. Jähne. *Digital Image Processing*. Springer, Berlin, 2005.

[23] I. Molchanov. *Statistics of the Boolean Model for Practitioners and Mathematicians*. J. Wiley & Sons, Chichester, 1997.

[24] O. M. Pecho, O. Stenzel, P. Gasser, M. Neumann, V. Schmidt, T. Hocker, R. J. Flatt, L. Holzer. 3D microstructure effects in Ni-YSZ anodes: Prediction of effective transport properties and optimization of redox-stability. *Materials*, **8**, 5554–5585, 2015.

[25] J. Van Brakel, P. M. Heertjes. Analysis of diffusion in macroporous media in terms of a porosity, a tortuosity and a constrictivity factor. *International Journal of Heat and Mass Transfer*, **17**, 1093–1103, 1974.

[26] L. Holzer, D. Wiedenmann, B. Münch, L. Keller, M. Prestat, P. Gasser, I. Robertson, B. Grobéty. The influence of constrictivity on the effective transport properties of porous layers in electrolysis and fuel cells. *Journal of Materials Science*, **48**, 2934–2952, 2013.

[27] O. Stenzel, M. Neumann, O. M. Pecho, L. Holzer, V. Schmidt Big data for microstructure-property relationships: A case study of predicting effective conductivities. *Working paper*, 2016.

[28] J. Friedman, T. Hastie, R. Tibshirani. *The Elements of Statistical Learning*. Springer, Berlin, 2001.

[29] G. Gaiselmann, O. Stenzel, A. Kruglova, F. Muecklich, V. Schmidt. Competitive stochastic growth model for the 3D morphology of eutectic Si in Al–Si alloys. *Computational Materials Science*, **69**, 289–298, 2013.

[30] M. Roland, A. Kruglova, G. Gaiselmann, T. Brereton, V. Schmidt, F. Mücklich, S. Diebels. Numerical simulation and comparison of a real Al–Si alloy with virtually generated alloys. *Archive of Applied Mechanics*, **85**, 1161–1171, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PERCOLATION PROPERTIES OF THE FREE VOLUME GENERATED BY TWO ROUGH SURFACES IN CONTACT

P. Cinat[1], **M. Paggi**[2], **C. Borri**[3]

[1]IMT School for Advanced Studies Lucca
Piazza San Francesco 19, 55100 Lucca, Italy
e-mail: paolo.cinat@imtlucca.it

[2]marco.paggi@imtlucca.it, [3] claudia.borri@imtlucca.it

Keywords: Contact mechanics, free volume, sealing, surface statistics.

Abstract. *The mechanism of fluid leakage trough the free volume between rough surfaces in contact is relevant in physics and in many engineering applications. In the present study, the normal contact problem between randomly generated fractal rough surfaces is solved using the boundary element method. Then, an algorithm for the evaluation of the network involved in the percolation of fluid is proposed. Numerical results are synthetically collected in diagrams relating the free volume involved in the percolation to the dimensionless statistical parameters of the rough surface.*

1 INTRODUCTION

The mechanism of percolation in seals or simply between two rough surfaces in contact is an interesting topic for physics and engineering [1, 2, 3]. This topic is relevant for many applications such as hydraulic fracturing and sealing of mechanical components in contact. Also, the percolation of fluid in photovoltaic modules can lead to electrical power losses due to the oxidation of the grid line deposited on the solar cell caused by moisture. The free volume and the related network of channels is due to the fact that surfaces are never ideally flat and the actual area in contact is determined by the elastic interactions between asperities. Since the full contact solution could be attained only in case of very high contact pressures, in most of the cases only a small percentage of the nominal contact area is in contact and a significant amount of free volume is present. In this context, a line of research regards the discovery of the relation between the contact mechanics results to the statistical properties of the undeformed parent rough surface with a power-law power spectral density (PSD) function [4]. Yastrebov investigated the role of the lower and upper cut-offs to the PSD function in different contact regimes [5], while Paggi and Ciavarella [6] highlighted the effect of the bandwidth parameter α. Similarly, the features of the leakage domain by considering the statistical proprieties of the surface height distribution were investigated in the literature. As highlighted by Dapp [7], the role of the finite size length and the lower cut-off frequency of the rough surface mainly affect

the topology of the free volume and, consequently, leakage characteristics [1]. All of those studies pinpoint the fact that, below the full contact limit, morphological proprieties of the surface affect percolation of a fluid between two rough surfaces. However, during contact, the initial statistics of the deformed surface change as well as consequence of the elastic interactions.

The present study aims at investigating the dependency of the leakage characteristics of free volume on the variation of statistical parameters describing the evolution of the contact domain between rough surfaces. To this purpose, a single rough surface with fractal dimension $D = 2.1$ has been generated and the contact problem has been solved for a series of mean plane separations using the boundary element method (BEM). Afterwards, the free volume has been evaluated and its portion involved in the percolation has been computed, under the assumption of a laminar fluid flow. In Section 2, the numerical framework is briefly summarized. Results are presented in Section 3, in form of diagrams relating the variation of the statistical parameters of the contact domain for decreasing mean plane separations and the relation with the variation of the percolating volume.

2 COMPUTATIONAL METHOD

A single rough surfaces with fractal properties is numerically generated according to the random midpoint displacement (RMD) algorithm detailed in [8]. With this method, it is possible to obtain a detailed square surface by refining an initial square mesh, adding recursively in the mid point an intermediate height. The value of this height is equal to the mean value of the neighboring heights, plus a random number extracted from a Gaussian distribution with zero mean and variance $\sigma_1^2 = \sigma_0^2/2^{(3-D)/2}$, where $\sigma_0^2 = 1/\sqrt{0.09}$ and D is the surface fractal dimension ranging from 2 to 3. The refinement algorithm depends upon the parameter m, which is related to the number of heights per side of the squared generated grid, viz. $2^m + 1$.

Once the surface is generated, the contact problem between an elastically deformable rough surface and a rigid half space is solved using BEM [9] for different values of the far-field closing displacement Δ. For each imposed Δ, the contact problem is solved using the Non-Negative Least Squares (NNLS) algorithm, without warm starting, as detailed in [9].

After solving the frictionless contact problem, the grid points in contact are determined and the contact mechanics results and the statistical parameters of the deformed surface are computed for each imposed displacement. The total free volume V is evaluated by summing up the corresponding contributions of the boundary elements not in contact. The complete procedure is described in [10]. Next, the percolating volume V_{pg} is computed as the sum of all the boundary element volumes involved in the leakage phenomena. Those elements are evaluated with an algorithm examining all the available paths for the fluid and retaining in the computation only the free channels connecting one side of the surface to the opposite. Statistical parameters of the deformed surface are evaluated according to Nayak's theory [4]. Nayak investigated the effect of some characteristic moments which are related do the r.m.s. values of roughness ($\sigma = \sqrt{m_0}$), the r.m.s slope ($\sigma_m = \sqrt{m_2}$) and the r.m.s. curvature ($\sigma_k = \sqrt{m_4}$), and the bandwidth parameter $\alpha = \frac{m_0 m_4^2}{m_2}$. Here, this computation is repeated for each imposed contact displacement, using the algorithm published in [11].

3 RESULTS

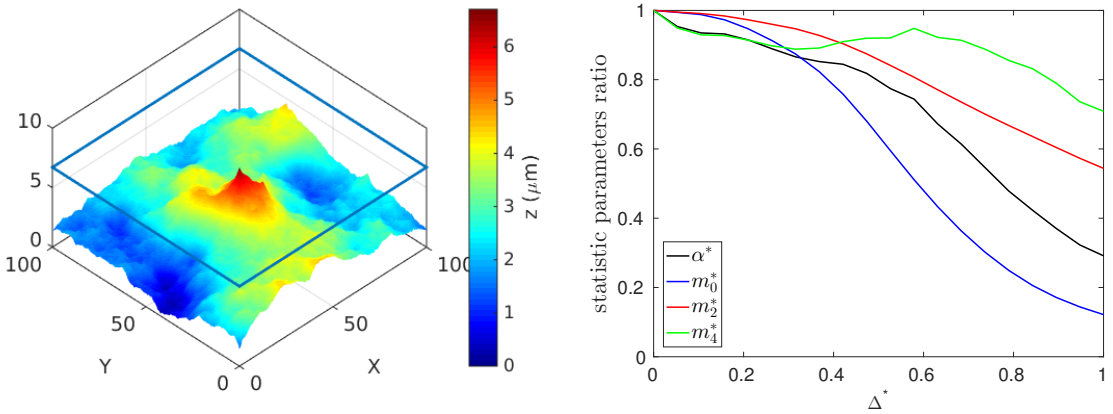
A rough surface with fractal dimension $D = 2.1$ has been generated using the RMD algorithm with a refinement corresponding to $m = 7$, see also [10] for more details. The surface lateral size is $L = 0.1$ mm, and the undeformed surface presents a maximum height equal to

$h_s = 6.72$ μm. Moreover, the mean surface elevation is equal to $\bar{h} = 2.78$ μm and the r.m.s. of heights is $\sigma = 0.95$ μ m^2, the r.m.s. of the profile slopes is $\sigma_m \cong 0.004$, and the r.m.s. of curvatures is $\sigma_k = 0.16$ μm^{-2}. The bandwidth parameter is $\alpha_0 = 921$.

The explored values of the far-field displacement Δ are in the range between $\Delta = 0$ (infinitesimal contact) and the maximum asperity height $\Delta = h_s$ computed in relation to the undeformed surface. This range is subdivided in twenty equal levels. The value of Δ is made dimensionless by dividing it for h_s, i.e, $\Delta^* = \frac{\Delta}{h_s}$. In this way, it is possible to the define the contact level from $\Delta^* = 0$ (infinitesimal contact regime) to $\Delta^* = 1$ (full contact regime). As detailed in [6], the dimensionless free volume $V^* = \frac{V}{L^2\sigma}$ is introduced, where L is the surface lateral size and σ is the r.m.s. value of height distribution. Therefore, the percolating volume V_{pg} is made dimensionless using its value when only one spot is in contact for $\Delta^* = 0$. The spectral moments are also made dimensionless by dividing them for their initial values, i.e., $m_n^* = \frac{m_n}{m_{n0}}$, where m_{n0} is the value of the $n-$th moment corresponding the undeformed surface.

The undeformed surface topography is shown in Fig.1(a), while Fig.1(b) shows the variation of the statistical parameters during contact. By increasing the contact interference Δ^*, statistical parameters diminish in value since the surface becomes flatter and flatter and the related r.m.s. becomes smaller and smaller. More specifically, m_0 presents small variations for infinitesimal contact, before significantly reducing its value for medium and full contact regimes. The same trend is noted for the variance of slopes, m_2. This transitional regime takes place for $\Delta^* \gtrsim 0.4$. The parameter m_4, for low values of Δ^*, tends to be almost constant with variations of about 10% up to $\Delta^* \cong 0.6$, while it significantly reduces afterwards. The variability of α vs. Δ^* is in line with the trend of m_0 and m_2, mitigated by the trend of m_4.

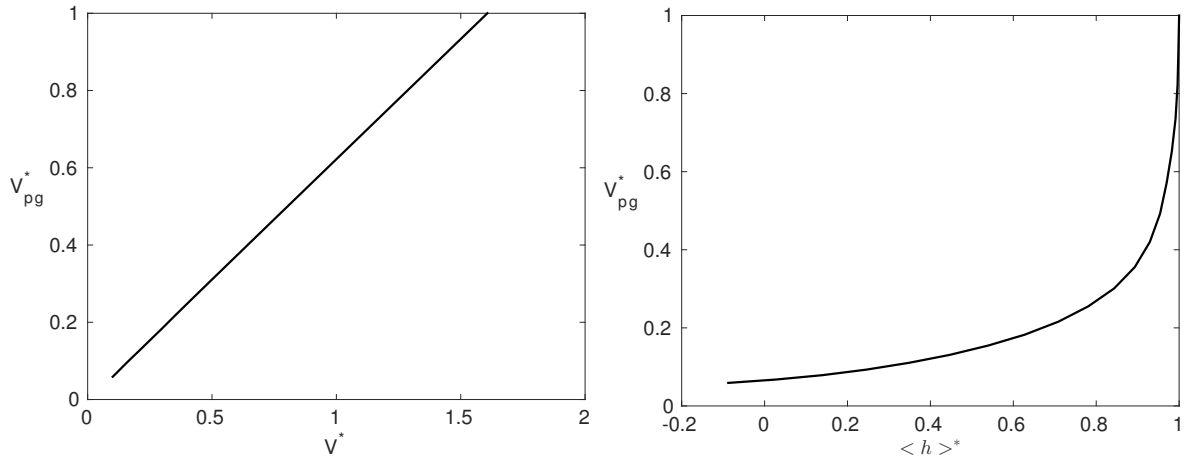
The dependencies of V_{pg} on the free volume V^* and on the dimensionless mean height $\bar{h}^* = \frac{h_s}{\bar{h}}$ are shown in Fig.2. It is important to notice that, for this surface, V_{pg} varies almost linearly with V_{pt}. This might be due by the morphology of the surface, which has the tallest asperity in the center. Moreover, the initial percolating volume V_{pg} is equal to the total free volume V^*. Examining Fig.2(b), the dependency of V_{pg} on the surface mean height presents a highly nonlinear trend, especially in the low contact regime ($\bar{h}^* = 1$). On the other hand, near the full contact limit ($\bar{h}^* = 0$), V_{pg} tends asymptotically to a constant value. Finally, V_{pg} is correlated to m_0^* and m_2^* during contact in Fig.3. In both cases, for an infinitesimal far-field displacement,



(a) Undeformed topography

(b) m_n and α vs Δ^*

Figure 1: Topography of the surface and variation of its Nayak's moments for different contact levels.



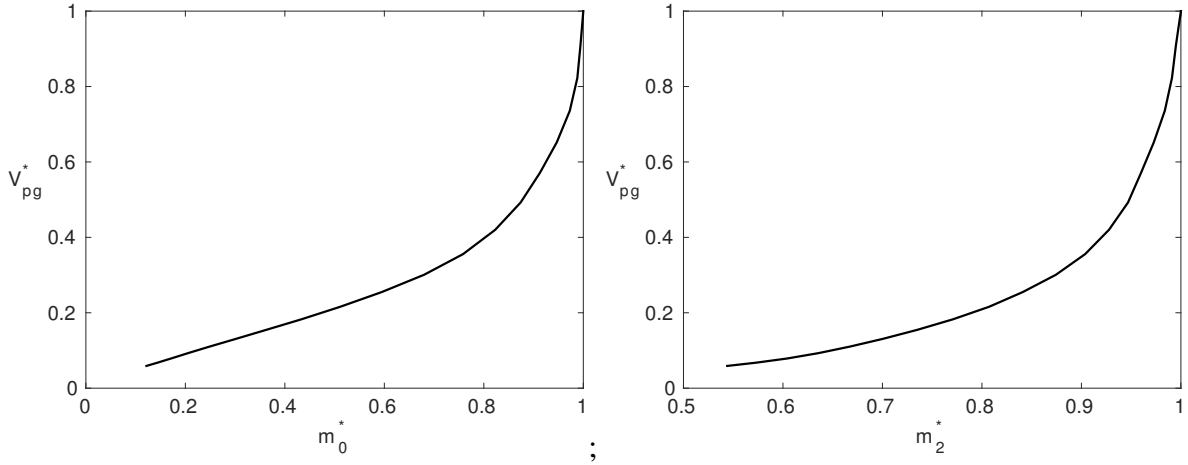
(a) V_p^* vs V^* (b) V_p^* vs $\bar{h}^*$

Figure 2: Variation of the leakage volume during contact. Fig.2(a) shows the dependency of the dimensionless leakage volume V_{pg}^* with respect to the overall free volume V^*. Fig.2(b) shows the dependency of V_{pg}^* on the dimensionless mean height of the deformed rough surface.

a strong dependency of V_{pg} on the moments of the PSD is reported. On the opposite, the percolating volume tends to be almost independent of the statistical proprieties of the surface near full contact. This could be connected directly with Persson theory of leakage [2, 3], which asserts that leakage phenomena do not depend on surface statistic in the full contact regime.

4 CONCLUSIONS

In this work, the free volume generated by two fractal rough surfaces in contact has been computed, determining also its percolating part, which is its portion through which a laminar fluid flow can cross the entire surface from on side to the opposite. A single surface has been generated with the RMD algorithm, and the contact problem has been solved with a BEM method for different far-field displacements, from the infinitesimal to the full contact regimes.



(a) V_p^* vs m_0 (b) V_p^* vs m_2

Figure 3: Variation of the leakage volume during contact. Fig.3(a) shows the dependency of the dimensionless leakage volume V_{pg}^* on m_0^*. Fig.3(b) shows the dependency on m_2^*.

Then, the percolating volume has been evaluated, considering its dependency on the statistical parameters of the surface computed according to Nayak's theory of roughness. Results shows that there is a linear dependency between the free volume V_* and the percolating one V_{pg}. Moreover, examining the height and slope r.m.s. values, it is possible to notice that the percolating volume strongly depends on their variation for small imposed far-field displacements. On the other hand, the percolating volume nearby the full contact limit appears to be almost independent of the statistical parameters variation, in agreement with the Persson's theory of leakage in seals. Future developments of the present study will regard a statistical analysis of the obtained results based on a wider population of numerically generated contact surfaces. Moreover, the computation of the portion of the free volume which is involved in leakage is also an important result, and its dependency on the statistics of the rough surfaces is also a crucial issue to deepen. This knowledge could be helpful to avoid surface configurations implying trapped fluid.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP/2007–2013)/ERC Grant Agreement No. 306622 (ERC Starting Grant Multi-field and Multi-scale Computational Approach to Design and Durability of PhotoVoltaic Modules – CA2PVM).

REFERENCES

[1] F. Bottiglione, G. Carbone, L. Mangialardi, G. Mantriota. Leakage mechanism in flat seals. *Journal of Applied Physics* **106**: 104902, 2009.

[2] B.N.J. Persson. Theory of rubber friction and contact mechanics *J. Chem. Phys.* **115 (8)**: 38840-3861, 2001.

[3] B. Lorenz, B.N.J. Persson. Leak rate of seals: Comparison of theory with experiment. *EPL (Europhysics Letters)* **86**: 44006, 2009.

[4] P.R. Nayak. Random process model of rough surfaces. *J Lubr Technol (ASME)* **93**:398-407, 1971.

[5] V.A. Yastrebov, G. Anciaux, J.-F. Molinari, From infinitesimal to full contact between rough surfaces: evolution of the contact area, *International Journal of Solids and Structures*, **52**;83-102, 2015.

[6] M. Paggi, M. Ciavarella, The coefficient of proportionality k between real contact area and load, with new asperity models. *Wear* **56**: 10201029, 2010.

[7] W.B. Dapp, A. Lucke, B.N.J. Persson, M.H. Muser, Self-affine elastic contacts: Percolation and leakage, *Phys Rev Lett* **108**: 244301, 2012.

[8] H.O. Peitgen, D. Saupe, The Science of Fractal Images, Springer-Verlag, New York, 1988.

[9] A. Bemporad, M. Paggi, Optimization algorithms for the solution of the frictionless normal contact between rough surfaces, *International Journal of Solids and Structures*, 2015, **06**; 69-70, 2015.

[10] M. Paggi, Q.-C. He. Evaluation of the free volume between rough surfaces in contact. *Wear*, **05**; 336-337, 2015.

[11] C. Borri, M. Paggi. Topological characterization of anti-reflective and hydrophobic rough surfaces: are random process theory and fractal modeling applicable? *J Phys D.* **48**: No. 045301, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A PARAMETRICALLY TIME-DEPENDENT METHODOLOGY FOR RECIPROCATING CONTACT MECHANICS BETWEEN VISCOELASTIC SOLIDS

Carmine Putignano[1,2], Giuseppe Carbone[1,2], Daniele Dini[2].

[1]Department of Mechanics, Mathematics and Management,Politecnico di Bari, Bari, Italy.
e-mail address: carmine.putignano@poliba.it

[2] Department of Mechanical Engineering, Imperial College London, London, United Kingdom.

Keywords: Reciprocating motion, viscoelasticity, friction.

Abstract. *We implement an original Boundary Element methodology to study the reciprocating contact mechanics between linear viscoelastic materials. Results are shown for the case of a rigid sphere sinusoidally driven in sliding contact with a viscoelastic half-space. We observe the presence of multi-peaked pressure and displacement distributions; the hysteric friction curve is finally shown for different values of the frequency.*

1 Introduction

In modern engineering research, wide research efforts are dedicated to analyze and determine the mechanics and physics of soft linearly viscoelastic materials. The complexity of the topic is related to the strongly time-dependent and usually non-linear constitutive stress-strain relations that mark this class of materials. In particular, further complexity is added when soft bodies are brought into contact and the problem is exacerbated by the geometry of the intimately mating surfaces. As a matter of fact, a variety of publications has been dedicated to shed light on the contact mechanics of rough viscoelastic solids. Given the importance of the applications related to these issues[1, 2, 3], investigations performed across the scales, from macroscopic to atomistic levels, include analytical [4, 5] numerical [6, 7, 8, 9] and experimental [10, 11, 12, 13] studies.

In this paper, we focus on an issue that has been systematically ignored but has a particular importance: the reciprocating contact of viscoelastic materials, where the relative motion between the contacting bodies is periodically inverted. There is a countless number of engineering applications, ranging from the macro- to the nano- scales, where a periodic inversion of the motion direction is present. Earthquake viscoelastic dampers are a classic example [15, 15]: the design of these devices mostly relies on practical and empirical guidelines, and no tool for quantitative predictions is available. This lack of a robust theoretical framework affect different components, like all the sealing systems in mechanical applications with an alternate motion [14]. Indeed, enhancing performances and efficiency is infeasible without an accurate knowledge of the interfacial stresses and, consequently, of the dissipated power. Finally, reciprocating contacts have prominence also at different scales and in different contexts, like biology and biotechnology ([17],[18]). Skin, ocular system, joints, spine and vertebrae are some of the examples where viscoelastic soft contact occurs in the human body.

In this work, we present a Boundary Element Methodology that, for the first time, allow studying reciprocating contact mechanics between linearly viscoelastic solids and provide predictions of the response of the contacting surfaces in terms of stresses, strain and friction. The paper is outlined as follows. Section II describes the mathematical formulation which the numerical methodology relies on. Section III focuses on the reciprocating sliding contact of a sphere over a viscoelastic layer. Final remarks are included to comment on the relevance of the theory and of the results.

2 Formulation

We implement a Boundary Element Methodology to determine the contact solution and, specifically, stresses, strains and dissipation. It is well known that the general contact problem between a rigid indenter and a linearly viscoelastic slab may be formulated as:

$$u\left(\mathbf{x},t\right)=\int_{-\infty}^{t}d\tau\int d^{2}x\mathcal{J}\left(t-\tau\right)\mathcal{G}\left(\mathbf{x}-\mathbf{x}'\right)\dot{\sigma}\left(\mathbf{x}',\tau\right), \tag{1}$$

where $\mathbf{x}$ is the in-plane position vector, t is the time, $u\left(\mathbf{x},t\right)$ is the normal surface displacement of the viscoelastic slab, $\sigma\left(\mathbf{x},t\right)$ is the normal interfacial stress, $\mathcal{G}\left(\mathbf{x}\right)$ and $\mathcal{J}\left(t\right)$ are respectively the elastic Green's function and the creep material function. The relevance of such an approach is related to its generality: no assumption is made *a priori* on the shape of the contact domain and on the specific form of the linear viscoelastic response. As a consequence, the method can be employed for any kind of contact punch (conditions, like periodic boundaries and finite values of contacting layers thickness, can be easily managed [22, 8]) and for any kind

of linear viscoelastic material, spanning from skin tissues to rubber-based composites.

However, tackling directly Eq. (1) is extremely challenging: because of the necessity of performing discretization both in time and space, the computational cost is huge, thus resulting infeasible with the computational technologies currently available. Consequently, when focusing our attention on the reciprocating contacts, we want to reduce the computational complexity of Eq. (1) without loosing its generality in terms of contact geometry and material properties. To this end, we assume that the interfacial normal stress distribution obeys the law $\sigma(\mathbf{x},t)=\sigma[\mathbf{x}-\xi_0\sin(\omega t)]$, i.e. that the shape of normal stress distribution is fixed but moves on the viscoelastic half-space with a sinusoidal law of amplitude $|\xi_0|$ and angular frequency ω. This enables us to rewrite Equation (1) as :

$$u(\mathbf{x},t)=\int d^2x' G(\mathbf{x}-\mathbf{x}',t)\,\sigma(\mathbf{x}')\,. \tag{2}$$

where, as shown in detail in Ref. [23], $G(\mathbf{x},t)$ is equal to:

$$G(\mathbf{x},t)=\sum_{k=-\infty}^{+\infty}\frac{A_k(\mathbf{x})}{E(k\omega)}e^{ik\omega t}. \tag{3}$$

where $A_k(\mathbf{x})$ can be written as

$$A_k(\mathbf{x})=(2\pi)^{-1}\int_{-1}^{1}ds\mathcal{G}(\mathbf{x}-s\xi_0)\,B_k(s) \tag{4}$$

with $B_k(s)$ being equal to $B_k(s)=(-i)^k T_k(s)\,B_0(s)$. $T_k(s)$ is the Chebyshev polynomial of the first kind and $B_0(s)=2(1-s^2)^{-1/2}$, for $|s|\leq 1$ and 0 otherwise.

As mentioned above, the function $G(\mathbf{x},t)$ relies on a precise assumption : we assume that the shape of the stress field at the interface, whose general form is $\sigma(\mathbf{x},t)=\sigma[\mathbf{x}-\xi_0\sin(\omega t),t]$, does not change during the reciprocating motion, *i.e.* $\sigma(\mathbf{x},t)=\sigma[\mathbf{x}-\xi_0\sin(\omega t)]$. Such a condition holds true whenever $a_0/|\xi_0|\ll 1$, where a_0 the characteristic dimension of the contact region, and is equivalent to require that $|\partial\sigma/\partial t|/(|\xi_0\cdot\nabla\sigma|\,\omega)\ll 1$ (the reader is referred to Ref. [23] for more details).

The solution of Eq. (2) can be pursued by employing the iterative technique developed in Ref. [21] for elastic contacts, thus providing contact areas, stresses, strains and, ultimately, the dissipated energy.

3 Results and discussion

In order to demonstrate the main features of the reciprocating contact, we study the contact of a rigid sphere of radius R undergoing reciprocating sliding against a viscoelastic material characterized by one relaxation time (being the ratio between the high frequency modulus and the low frequency $E_\infty/E_0=11$ and the Poisson ratio $\nu=0.5$). We assume that the center $\mathbf{x}(t)$ of the sphere moves on the viscoelastic half-space following the law $\mathbf{x}(t)=[\xi_0\sin(\omega t),0]$. The dimensionless angular frequency of the reciprocating motion is $\omega\tau=5$, being τ the relaxation time of the viscoelastic material.

For different values of $\omega t\in[-\pi/2,\pi/2]$, Figure 2 shows the evolution of the dimensionless displacements, $u(x)/R$, at the centre of the contact as a function of x/ξ_0 and for a specific dimensionless applied normal load $F_n/R^2E_0^*=0.014$, and $\xi_0/R=1$. An arrow refers, in each case, to the current position of the sphere. At $\omega t=-\pi/2$ the sphere has just reached the left dead-point and starts moving from left to right. Upon reversal of the sliding direction,

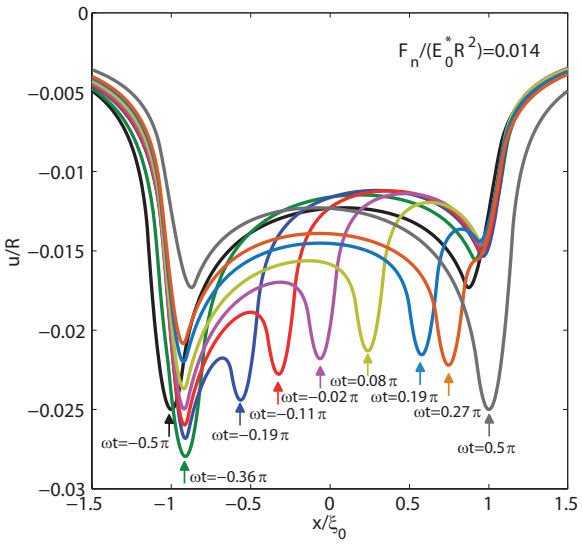


Figure 1: The dimensionless normal displacements $u(x, y = 0)/R$ as a function of the dimensionless abscissa x/ξ_0 for a constant dimensionless normal force $F_n/R^2E_0^* = 0.014$, for an amplitude $\xi_0/R = 1$ and for several values of the dimensionless time $\omega t \in [-\pi/2, \pi/2]$.

a marked increase of the dimensionless penetration at the center of the sphere is observed: although the speed is increasing, it is still too low to provoke a significant stiffening of the material, and the sphere is moving over a region of the viscoelastic half-space that has not yet had the time to relax after the previous contact of the rigid body. As the sliding speed increases, a non negligible stiffening of the material and a marked decrease of the penetration are observed (see displacement in correspondence to the arrow). Furthermore, additional deformation peaks appear: this is the result of the interplay between the deformations, induced by the indenter as it moves to the right, and the original not yet fully relaxed footprints left by the sphere at preceding times. For $0 < \omega t < \pi/2$, the sliding speed begins to decrease and the material softens again, thus leading to an increase of penetration. It is now possible to justify the occurrence of three different deformations peaks within the track when the sphere is moving between the two dead-ends: one corresponds to the current position of the sphere and the other two are located close to the left and right dead-points respectively, and are the result of the material inability to fully recover the viscoelastic deformations during a period of time comparable to the period $T = 2\pi/\omega = 6.28$ s of the reciprocating motion (recall that the relaxation time is $\tau = 5$ s).

The interplay of the sphere footprints, which occur at the dead-points of the reciprocating motion, has a significant effect on the interfacial normal stress distribution. This is clear in Figure 2, where contour plots of the pressure distribution are shown. Let us first notice that at $\omega t = -\pi/2$, *i.e.* when the sliding speed is equal to zero, the contact area as well as the interfacial normal stress distribution are characterized by an asymmetric shape. This is due to the viscoelastic time-delay which prevents the material to relax immediately when the sliding speed vanishes. As the sphere starts moving to the right, such a peak cannot disappear suddenly but has to show a gradual decrease. At the same time, since the punch is travelling towards the right, as already observed in steady-state viscoelastic contacts moving at constant velocity [7], a peak in the pressure distribution has to be originated also at the leading edge . Finally, at the center of the distribution, where we have the maximum of the displacement field in the contact area, the pressure must still resemble the classical elastic Hertzian solution. All this process strongly affects the evolution of the pressure distribution at the interface with the presence of

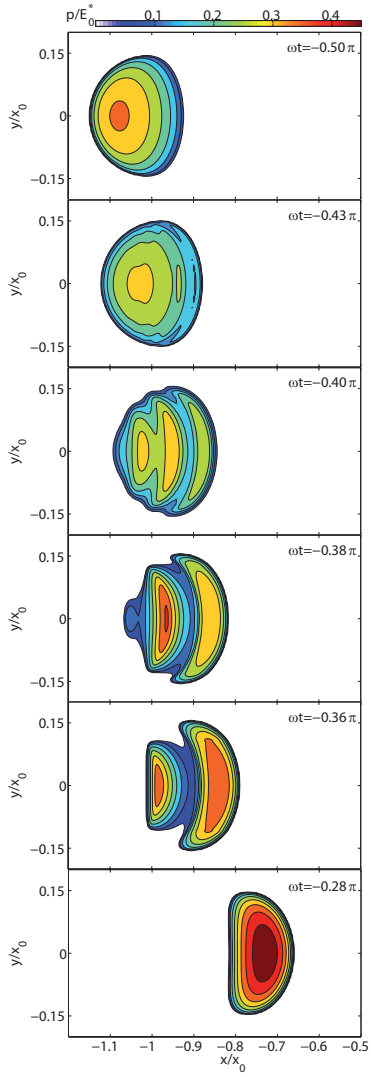


Figure 2: The shape of the contact area and the contour plots of the normalised contact pressure distributions, p/E_0^*, for several values of ωt .

multiple pressure peaks shown by the snapshots taken at $\omega t = -0.40\pi, -0.38\pi, -0.36\pi$.

Finally, in Figure 3, the reduced tangential force, F_t/F_n, easily calculated once pressures and displacements are known [7], is plotted as a function of the dimensionless abscissa x/ξ_0, which identifies the position of the sphere along the path, for different values of the parameter $\Xi = \omega\tau$. In detail, fixed the parameter $\Pi = \tau/t_0$ with t_0 being equal to $t_0 = a_0/\omega\tau$ -a_0 is the Hertzian radius-, for $\Xi = 0.1$ the material has the possibility to relax before a single reciprocating cycle is completed; consequently, the solution resembles the steady-state viscoelastic sliding contact and the tangential force F_t/F_n always opposes the sphere speed at each point along the path. However, as Ξ is increased the relaxation of the material involves time scales comparable to the time period of the reciprocating motion; in this case, there exist regions on the sphere track, specifically those close to the dead-points, where F_t/F_n has the same direction as the sliding speed. This is perfectly consistent with the results presented in Figure 2.

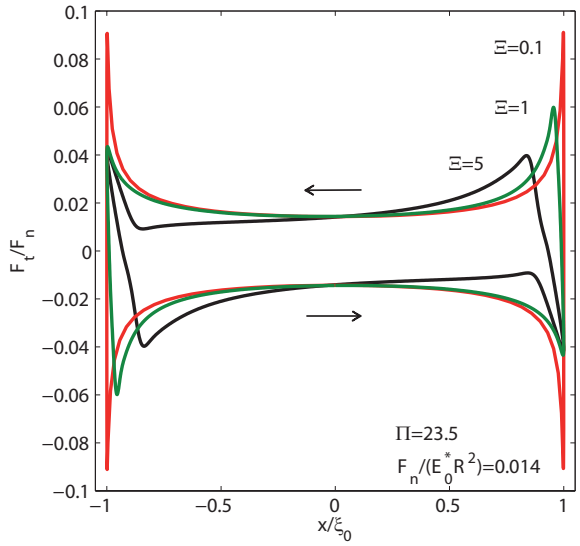


Figure 3: The ratio between the tangential and the normal force F_t/F_n as a function of the dimensionless abscissa x/ξ_0 for different values of Ξ . Arrows refer to the hysteresis cycle direction.

4 Conclusion

In this work, we present a Boundary Element Methodology capable of providing the full solution of the reciprocating contact problem between a rigid punch and a linear viscoelastic solid. The periodic features, intrinsically marking the problem, enables the parametric calculation of the contact solution for each time step without any necessity of employing the solution in the previous time interval. By implementing such a parametrically time-dependent approach, we obtain the full numerical convergence in each moment of the cycle and, interestingly, also when the punch inverts its motion.

For the simple case of a sphere in contact with a viscoelastic layer, we show that the behavior of the system is characterized by three peaks in the displacement distribution and, consequently, by multi-peaked pressure distributions. Finally, when calculating the dissipated energy, we observe that, for large values of the parameter Ξ, the tangential force F_t may have the same direction of the sliding speed.

Acknowledgements

CP gratefully acknowledges the support of Marie Curie Intra-European Fellowship SOFT-MECH (Grant No. 622632).

REFERENCES

[1] Arzt E. , Gorb S., and Spolenak R., From micro to nano contacts in biological attachment devices. Proceedings of the National Academy of Sciences, 16, 100 , 19 , 10603–10606, (2003).

[2] Bao G. , Suresh S., Cell and molecular mechanics of biological materials, Nature Materials 2, 715 - 725, (2003).

[3] Johannessen W., Vresilovic E. J., Wright A. C., Elliott D. M., Intervertebral Disc Mechanics Are Restored Following Cyclic Loading and Unloaded Recovery. Annals of Biomedical Engineering, 32, 1, 70-76, (2004).

[4] Hunter S.C. , The rolling contact of a rigid cylinder with a viscoelastic half space . Trans. ASME, Ser. E, J. Appl. Mech. 28, 611–617 (1961).

[5] Persson B.N.J., Theory of rubber friction and contact mechanics, Journal of Chemical Physics, 115, 3840 -3861, (2001).

[6] Dapp, W.B., Lücke, A., Persson, B.N.J., Müser, M.H., Self-affine elastic contacts: Percolation and leakage, Physical Review Letters, 108 (24), 244301, (2012).

[7] Carbone G. , Putignano C., A novel methodology to predict sliding/rolling friction in viscoelastic materials: theory and experiments., Journal of Mechanics and Physics of Solids, Journal of the Mechanics and Physics of Solids, 61 (8), 1822-1834, (2013).

[8] Putignano C. , Carbone G. , Dini D., Mechanics of Rough Contacts in Elastic and Viscoelastic Thin Layers, International Journal of Solids and Structures, 69–70, 507–517, (2015).

[9] Carbone G. , Putignano C.. Rough viscoelastic sliding contact: theory and experiments. Physical Review E, 89, 032408, (2014).

[10] Grosch K. A. , The Relation between the Friction and Visco-Elastic Properties of Rubber, Proceedings of the Royal Society of London. Series A, Mathematical and Physical, 274-1356, 21-39, (1963).

[11] Putignano C., Reddyhoff T., Carbone G., Dini D., Experimental investigation of viscoelastic rolling contacts: a comparison with theory. Tribology Letters , 51, 105–113, (2013).

[12] Putignano C., Le Rouzic J., Reddyhoff T. , Carbone G., Dini D. , A Theoretical and Experimental Study of Viscoelastic Rolling Contacts Incorporating Thermal Effects, Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, vol. 228, no. 10, 1112-1121, (2014).

[13] Putignano C., Reddyhoff T. , Dini. D. The influence of temperature on viscoelastic friction properties. Tribology International, doi:10.1016/j.triboint.2016.03.018, (2016).

[14] Thatte A., Salant R. F., Effects of multi-scale viscoelasticity of polymers on high-pressure, high-frequency sealing dynamics, Tribology International, 52, 75–86, (2012).

[15] Shukla A. and Datta T. , Optimal Use of Viscoelastic Dampers in Building Frames for Seismic Force. Journal of Structural Engineering, 125 (4), 401–409, (1999).

[16] Takewaki I. Building Control with Passive Dampers: Optimal Performance-based Design for Earthquakes, John Wiley & Sons, (2011).

[17] Angelini T. E. , Dunn A. C., Uruena J. M., Dickrell D. J., Burris D. L. and Sawyer W. G. , Cell Friction, Faraday Discussion , 156, 31–39, (2012).

[18] Dunn A. C., Cobb J. A., Kantzios A. N., Lee S. J. , Sarntinoranont M. , Tran-Son-Tay R. , Sawyer W. G. , Friction Coefficient Measurement of Hydrogel Materials on Living Epithelial Cells, Tribology Letters, 30, 1,13-19, (2008).

[19] Christensen R. M., Theory of viscoelasticity, Academic Press, New York, (1982).

[20] Johnson, K.L.J., Contact Mechanics, Cambridge University Press, (1985).

[21] Putignano C., Afferrante L. , Carbone G., Demelio G. , The influence of the statistical properties of self-affine surfaces in elastic contact: a numerical investigation, Journal of Mechanics and Physics of Solids, 60, 5, 973–982, (2012).

[22] Carbone G., Mangialardi L., Analysis of adhesive contact of confined layers by using a Green's function approach, The Journal of the Mechanics and Physics of Solids, **56** (2), 684-706 (2008).

[23] Putignano C., Carbone G.,Dini D. : Theory of Reciprocating Contact for Viscoelastic Solids Accepted on Physical Review E (2016).

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

STATISTICAL AND GEOMETRICAL CHARACTERISTICS OF RANDOMLY ROUGH SURFACES USED FOR CONTACT SIMULATIONS.

Rafael Schouwenaars[1], Carlos G. Figueroa[1], Víctor H. Jacobo[1] and Armando Ortiz[1]

[1] Departamento de Materiales y Manufactura
Facultad de Ingeniería Edif. O
Universidad Nacional Autónoma de México
Avenida Universidad 3000, CDMX
e-mail: raf_schouwenaars@yahoo.com, vicjacobo@yahoo.com.mx, armandoo@unam.mx

Keywords: Contact mechanics, Fractal surface, Finite elements, Midpoint algorithm, Spectral method.

Abstract. *Over the last decades, it has been established that rough surfaces of technological materials can be described by means of random processes with fractal character. This means that the numerical simulation of contact mechanics will depend on the size of the simulated surface or the lower cut-of distance (i.e. mesh size) of the simulation; moreover, each simulation only presents a single case out of an infinite spectrum of possible results. This work explores the problem of statistical convergence by focusing on the surface characteristics which are relevant for contact mechanics applications, i.e. the amount of simulations required to obtain a reasonable estimate of the surface characteristics as a function of the size of the simulated surface. This analysis is applied to three classical methods for the generation of fractal surfaces, i.e. the Midpoint Displacement algorithm, the generalised Weierstrass-Mandelbrot function and a Random Spectral method. Apart from geometrical characteristics, the simulated surfaces will be evaluated by means of a numerical modification of the Greenwood-Williamson model. It will be seen that the Weierstrass-Mandelbrot function has several shortcomings which may affect advanced contact simulations. More generally, the statistical variation between individual surface simulations introduces a considerable spread on the results, as seen in the distribution of calculated fractal dimensions or load-contact area curves.*

1 INTRODUCTION

The random nature of the roughness of engineering surfaces and its role in the determination of contact problems has been recognised since the middle of the 20th century. Archard [1] postulated a model consisting of superimposed elastic spheres of decreasing size to approximate a rough elastic surface. Greenwood and Williamson (GW) [2] in 1966 formulated their contact model in which randomly distributed asperities were modelled as spherical caps. Their model was random but not fractal. Likewise, in 1969, Whitehouse and Archard [3] provided an advanced statistical analysis of the asperity distribution for a randomly rough surface modelled as white noise.

Independently, Longuet-Higgins [4] in 1957 analysed the random superposition of waves of different frequencies as a model for the sea surface and provided an analysis of most of the important statistical properties of such a surface, which are still of interest in the field of contact mechanics. Reference to fractals was not made in this early research, as the concept of fractal dimensions was brought to the attention of a broader public only in 1967 [5]. The recognition of the fact that many engineering surfaces show fractal characteristics affects the aforementioned analyses in several ways. Greenwood concluded that for a fractal surface, one point out of three would become an asperity [6] (this effect is seriously mitigated if two-dimensional height maps are considered instead of linear roughness profiles). The model by Archard, on the other hand, is fractal but not stochastic. Finally, the statistical analysis by Longuet-Higgins relies on the existence of the image moments of the power spectrum of the surface. For fractal surfaces, which are non-stationary random processes, these moments do not exist, corresponding to the fact that statistical parameters such as the root mean square (RMS) roughness depend on the size of the measured area (for a finite measured or simulated area, this inconvenience does not generate any practical problems).

While the fractal nature of surfaces may affect classical models to some degree, the known properties of the statistical functions describing them, such as the power spectrum and autocorrelation function open the possibilities for formulating mathematical models which take this aspect into account, as was done by Majumdar and Bhushan [7,8] for elasto-plastic materials and by Persson for viscoelastic materials [9,10]. However, such models often include additional assumptions in the calculations which affects their mathematical soundness and to some degree their precision [11]. An exception is the mathematical analysis of the contact of the Weierstrass profile [12], but this result is limited to the two-dimensional case.

As a consequence, most researchers try to incorporate the fractal nature of surfaces through numerical simulation of both the fractal surface and the contact problem [7,13,14]. It shall be noted that often, such simulations are based on surface profiles obtained along a line and not on the entire measured or simulated surface, which is the topic treated in this paper. In some occasions, complete surfaces are generated, but only to extract a large number of linear scans for statistical analysis [15-17]. Taking into account the widespread availability of fully 3D optical profilers which provide complete height maps, such approaches are becoming more and more anachronistic.

Full 3-D analysis has received more attention in recent times. On the one hand, there are improved versions of the original GW-model [18,19], on the other hand, large-scale numerical methods are used by means of finite elements [20,21], boundary elements [22-24], and combinations of various techniques in a multiscale approach [24,25]. The main complication associated to these simulations is the elevated computational cost associated to the solution of large systems of equations or to the iterations required for incremental solutions. Very specifically, Putignano et al. [22-24] describe the need for mesh refinement to acquire numerical

convergence, i.e. they investigate the point to which the mesh must be locally refined beyond which no significant increase in precision is obtained.

The present work takes an entirely different approach to the problem of convergence. While some finite element results will be presented to illustrate the problems to be addressed, the focus of the text is on the statistical aspects introduced by the measurement or simulation of random surfaces. A single example of a measured surface will be compared by simulated surfaces produced by the midpoint algorithm (MP), the random spectral method (RS) and the generalised Weierstrass-Mandelbrot (WM) function, as explained in section 2. The geometrical characteristics used to analyse these surfaces will be outlined and illustrated by an example of each of the techniques proposed. Then, a statistical analysis will be presented of some of these characteristics made on 100 repetitions for the MP and RS methods. To clearly exemplify the importance of such analysis in contact mechanics, a numerical version of the GW-model is applied to each of the simulated surfaces and the statistics of the results will be analysed.

2 METHODS

Three methods were considered to create the randomly rough surfaces. The first one is the random midpoint algorithm with random additions only to the midpoints, as described by Saupe [26] and used in contact simulations by Hyun et al. [20,21]. The second method (RS) is based on the analysis of random surfaces by Longuet-Higgins [4], who analised surfaces produced by an infinite cosine series where the amplitude of the terms in the series scales as a known function (the power spectrum) of the wavelength and the phase angles are randomly distributed in the interval [-π,π]. The method is described by the following formula:

$$z(x,y) = \frac{2\pi(b-a)}{N}\sum_{n=1}^{N} c(r)\cos[r(2\pi x\cos\theta + 2\pi y\sin\theta + \varphi)]$$

where z is the surface height, N the number of terms in the series, r the length of the wave vector, a and b the lower and upper cut-of length for the wave vector, θ the angle between the wave vector and the X-axis and φ a random phase angle. The amplitude $c(r)$ obeys the following law:

$$c(r) = r^{-7+2D}$$

The angles θ and φ are randomly chosen in the interval [-π,π]. The value of r is randomly chosen in the interval $[a,b]$ in such a way that the wave vectors are uniformly distributed in frequency space. D is the fractal dimension. All results in the present work correspond to N=1000.

For the WM-surface, the following formula is used [13,14,27]:

$$z(x,y) = L\left(\frac{G}{L}\right)^{D-2}\sqrt{\frac{\log\gamma}{M}}\sum_{m=1}^{M}\sum_{n=0}^{N}\gamma^{(D-3)n}\cos\Big[\varphi - \cos\left(2\pi\gamma^{n}\sqrt{x^2+y^2}\cos\left(\arctan\left(\frac{y}{x}\right)-\frac{\pi\, m}{M}\right)\right)+\varphi\Big]$$

L is the length of the simulated surface, which is taken equal to 1 in this work. G is called the fractal roughness. The lacunarity γ [26] was set equal to 1.5 as suggested in literature [7]. The random angle φ is chosen from a uniform distribution on [-π,π]. M is taken equal to 9 and N=14.

All surfaces were calculated on a unit square with 256×256 subdivisions. The upper and lower cut-off for the RS method were taken equal to 1 and 2^8 respectively. The mean value of the simulated height was subtracted from the height values in each point and the height values

divided by the standard deviation. Therefore, all surfaces have the same RMS roughness of 1. The power spectra were calculated as the absolute value of the Fourier transform of the surface. In each point, the derivatives were determined by fitting the following polynomial to the point and 8 surrounding points (neighbours and corners):

$$f(x,y) = f(x_0,y_0) + f_{10}(x-x_0) + f_{01}(y-y_0) + \frac{f_{20}}{2}(x-x_0)^2 + f_{11}(x-x_0)(y-y_0) + \frac{f_{02}}{2}(y-y_0)^2$$

The f_{ij} in this formula evidently are the first and second partial derivatives of the function in the neighbourhood of the point (x_0,y_0). Asperities were determined as points which were higher than the eight surrounding points. The asperity radius was calculated as the geometric mean of the eigenvalues of the Hessian at the corresponding point. The fractal dimension of the surface was calculated by the triangular prism method [28,29].

To illustrate the relevance of statistical variations in the field of contact mechanics, a numerical version of the GW-model (nGW) was used for each simulated surface. Load and contact area are calculated for each individual asperity by first defining the impression depth of an asperity:

$$h_i(z) = \begin{cases} z_0 - z & z < z_0 \\ 0 & z \geq z_0 \end{cases}$$

where z_0 is the original vertical position of the asperity and z the actual position of the contacting plane. Then the load and contact area of an individual asperity are given by:

$$p_i(z) = \frac{4}{3} r_i^{1/2} h_i^{3/2}$$

$$a_i(z) = \pi r_i h_i$$

and the evolution of the total load and area are simply obtained by summing the above formulas for all asperities at a given number of discrete values for z. In the results, non-dimensional areas (contact area / total area) and non-dimensional loads (total load × total area / Young's modulus) are used. As a consequence of setting the standard deviation of the roughness to 1 (for an area of 1×1), the roughness studied here is very high as compared to what is found on engineering surfaces, but this artefact is not relevant for the results.

Although the main topic of this paper relates to the statistics of surfaces generated by numerical algorithms, a brief reference will be made to a measured surface and to the finite element modelling of contact problems to provide a reference and justification for the study. The measured surface was arbitrarily chosen from a set of available results obtained in a NANOVEA white light achromatic optical profiler and corresponds to a cold-rolled and recrystallised AA6016T4 aluminium alloy. This material is intended for deep-drawing applications in the automotive industry but is not subject to a specific surface modification process and therefore shows a fractal roughness pattern, inherited from the roughness of the rolls.

A finite element mesh was generated in ABAQUS which has 244×244 (3^5+1) surface nodes. Three layers of uniform cube elements are used underneath the surface before the mesh is allowed to coarsen with a factor of 3 in each subsequent two layers of elements below the surface. All the important stress concentrations during the simulations are limited to the upper three, fine-meshed layers of the model. The simulations were performed under an explicit integration scheme; care was taken to maintain the kinetic energy of the model below 5% of the total energy and to avoid dynamic effects at the moment of first contact between the bodies.

Roughness was introduced on the surfaces of both contacting bodies by means of the MP-method with a RMS roughness equal to 1/10 of the element height in the upper three layers of the model. The upper contacting body was modelled as an elastic solid with a Young's

modulus of 210 GPa and Poisson coefficient of 0.3, the lower body had the same elastic properties but was assigned a yield stress of 210 MPa and a strain hardening exponent of 0.2, using Hollomon's equation for the elastoplastic contact simulations. The friction coefficient between both surfaces was set equal to 0 in one set of simulations and to 0.1 in a second one, to study the interaction between this effect and the effect of plastic deformation of the asperities. Simulations were executed by approaching both surfaces to a distance equivalent to a bearing area of 40% and then sliding horizontally at a fixed vertical distance.

3 FINITE ELEMENT RESULTS

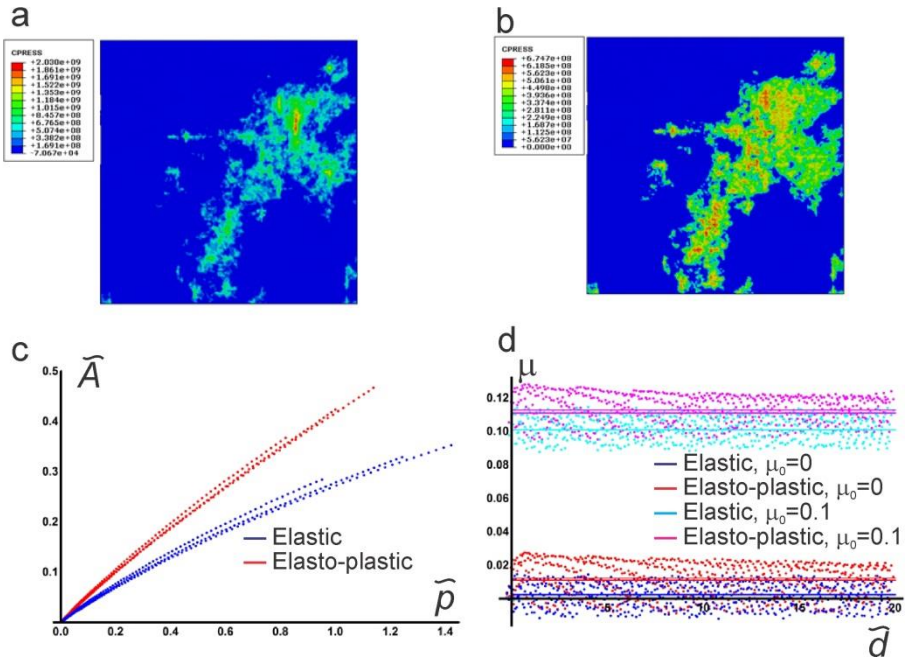


Fig. 1. a) contact area for a fractal surface generated by the midpoint algorithm at 40% bearing area for an elastic and b) elastoplastic strain hardening medium. c) presents the load-displacement curve, d) gives the friction coefficient for movement in the horizontal direction for the elastic and elasto-plastic models at 40% bearing area. The friction coefficients used between the contacting surfaces in the FEM were 0 and 0.1; for the elasto-plastic contact, the resulting friction coefficient is increased due to energy dissipation in the plastic asperities.

The finite element results are discussed first because they illustrate the reason why the statistical simulations are essential. Elastic simulations take approximately four hours on a on a workstation with an Intel® Xeon® CPU E5-2687W processor at 3.40GHz and 64 GB RAM, plastic simulations twice that time, with the horizontal sliding portion representing the main portion of the calculations. The difference in contact area between the elastic (Fig 1a) and elasto-plastic simulations is evident in the load-contact area curves as well (Fig. 1c), with the area increasing more rapidly for the latter case. The effect of the statistical nature of the simulations is clearly visualised in the spread on the curves in Fig. 1c.

One of the long-term goals of the authors is to research the interaction between the friction coefficient μ used in the finite element model (which can be considered as a micro-friction coefficient at the level of the individual asperities) and the one obtained by dividing the horizontal force during sliding by the vertical force (a macro-friction coefficient at the level of the entire body). This is illustrated in Fig. 1d. Here, a very large dispersion is observed in the individual points of the distinct simulations, as the interaction between individual asperities is visualised during the sliding process. The average values differ relatively less but show clear

variation from simulation to simulation. It is also clear that anisotropy of the surfaces would affect the results in sliding, even though its effect in normal contact may be minor.

4 SURFACE SIMULATIONS

Fig. 2 presents graphical examples for individual surfaces. The standard deviations for the height distributions have been scaled to 1 and the fractal dimension of the simulated surfaces was set equal to the measured one, i.e. D=2.8. As a consequence, the topography of the four examples looks very similar (left hand column). With respect to the power spectra (central column), the "grainy" appearance of the spectrum of the measured surface is important to notice. This appears to be a characteristic of "real" surfaces and is very well reproduced by the MP-spectrum and to a lesser degree by RS. In the latter case, an increase of N increases the similarity with the measured data. The spectrum of the WM-surface shows a deterministic pattern of individual peaks and can therefore not be considered as a realistic simulation of measured data. Finally, referring to the autocorrelation functions, it is seen that the measured surface presents a complex pattern which is not found in the simulated ones. This topic has to be researched further.

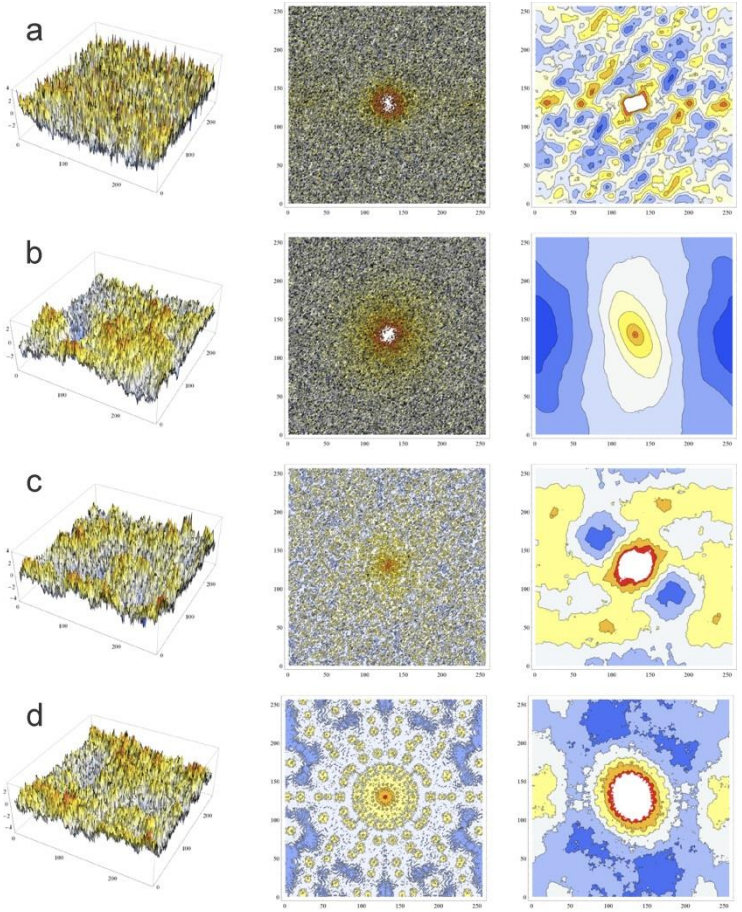


Fig. 2. Graphical representation of the functions describing the surfaces, from left to right the surface topography, the power spectrum and the autocorrelation function for a) a measured surface, b) a surface simulated by the random midpoint algorithm, c) by the random spectral method and d) by the modified Weierstrass-Mandelbrot function.

The statistical characteristics of the four surfaces presented in fig. 2 are analysed in figs. 3-6. Only some of the many possible surface metrics determined during simulation are presented. With respect to the height distributions, height distributions of the asperities and distribu-

tion of the partial derivatives of the surfaces, it is seen that the simulated surfaces follow a Gaussian distribution as predicted by theory [4], while the measured ones do not. Only a slight correlation between asperity radius and height is seen in the measured data, which is almost absent in the simulated ones. In any case, this correlation is probably too weak to be taken into account expressly in contact models. The measured surface is anisotropic, as demonstrated by the difference in the distribution of the partial derivatives and the orientation distribution of the surface gradients. This is a logical result for a cold-rolled and recrystallised material.

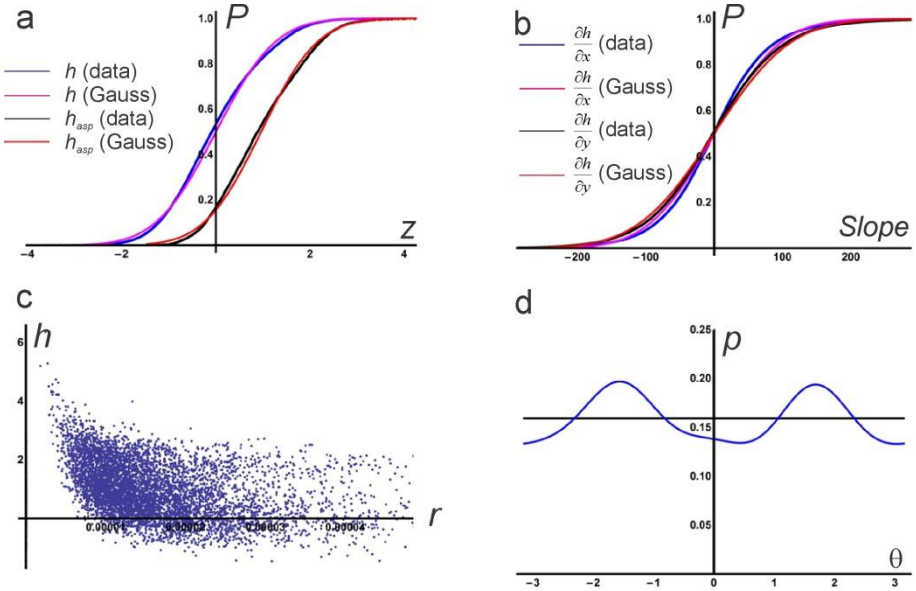


Fig. 3. Statistical parameters of measured surface. a) gives the cumulative distribution function (cdf) of the surface height (all points) and asperity height as well as the ideal cumulative Gaussian distribution with the same mean and standard deviation. b) shows the cdf of the partial derivatives, c) plots asperity heights vs. asperity radius and d) shows the porbability density function (pdf) of the directions of the gradients.

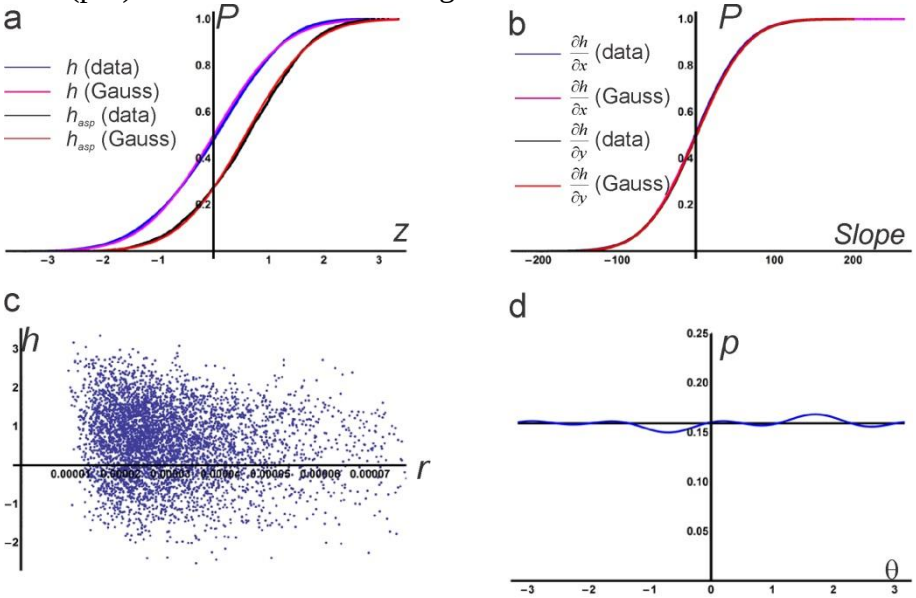


Fig. 4. Statistical parameters of a surface generated by the midpoint algorithm. a) gives the cdf of the surface height (all points) and asperity height as well as the ideal cumulative Gaussian distribution with the same mean and standard deviation. b) shows the cdf of the par-

tial derivatives, c) plots asperity heights vs. asperity radius and d) shows the pdf of the directions of the gradients.

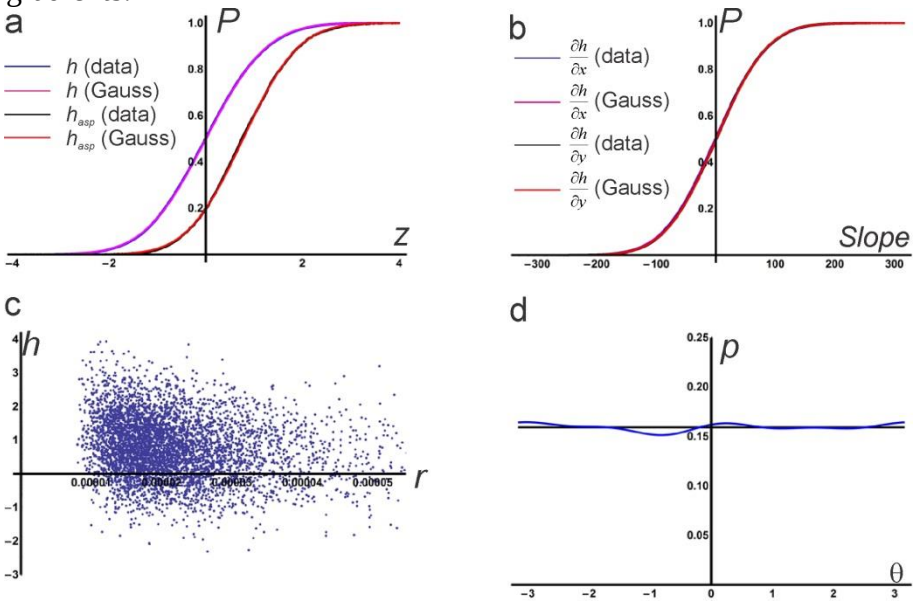


Fig. 5. Statistical parameters of a surface generated by the random spectral method. a) gives the cdf of the surface height (all points) and asperity height as well as the ideal cumulative Gaussian distribution with the same mean and standard deviation. b) shows the cdf of the partial derivatives, c) plots asperity heights vs. asperity radius and d) shows the pdf of the directions of the gradients.

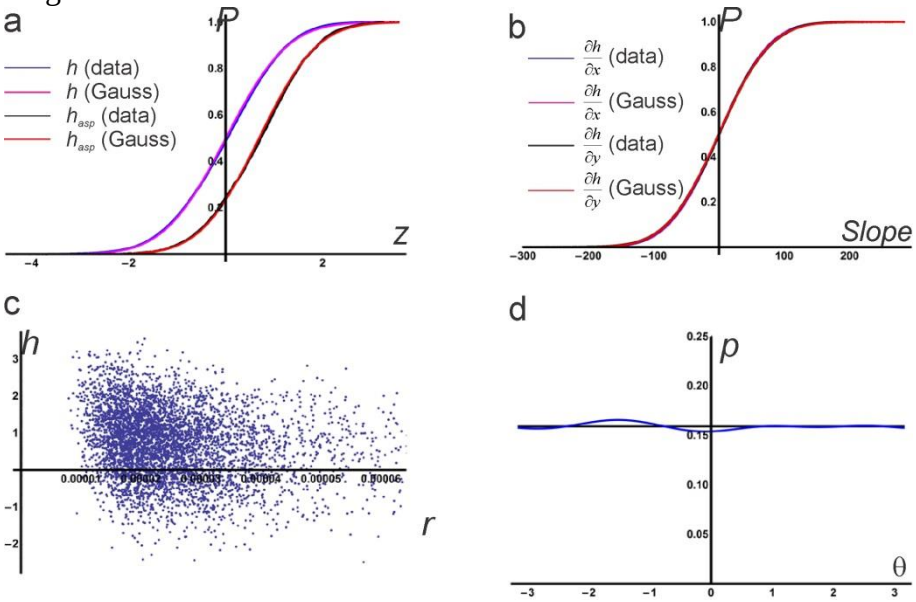


Fig. 6. Statistical parameters of a surface generated by the generalised Weierstrass-Mandelbrot function. a) gives the cdf of the surface height (all points) and asperity height as well as the ideal cumulative Gaussian distribution with the same mean and standard deviation. b) shows the cdf of the partial derivatives, c) plots asperity heights vs. asperity radius and d) shows the pdf of the directions of the gradients.

5 STATISTICAL RESULTS

Results for the statistical simulation of 100 surfaces by the MP and RS methods are analysed here, using a fractal dimension D=2.5. Both the algorithms used for the calculation of the surface as the triangular prism method for the determination of its fractal dimension can induce systematic or statistic variations of the fractal dimension D, as illustrated in fig. 7. The results are biased toward higher values of D and show some variation with respect to the average value. It would certainly be interesting to research the origin of the bias, but the statistical spread is probably a natural result caused by the random nature of the processes involved. It must be compared in future research to the variation observed on real surfaces.

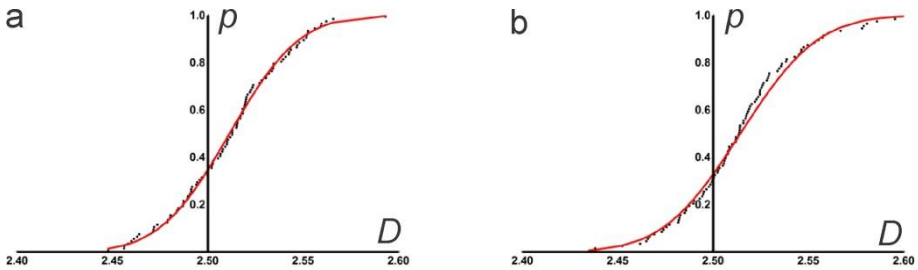


Fig. 7. a) Distribution of the fractal dimension for 100 simulations by the MP-method and b) for the RS-method. The Gaussian distribution is superimposed as a reference.

To illustrate the effect of random variation in the field of contact mechanics, 20 simulation points obtained by the nGW-model for each of the 100 surfaces are plotted in fig. 8. Considerable spread is observed around the average curves. Differences between the average curves are small; to identify such differences quantitatively, more simulations are required to allow determining confidence intervals as a function of the number of simulations.

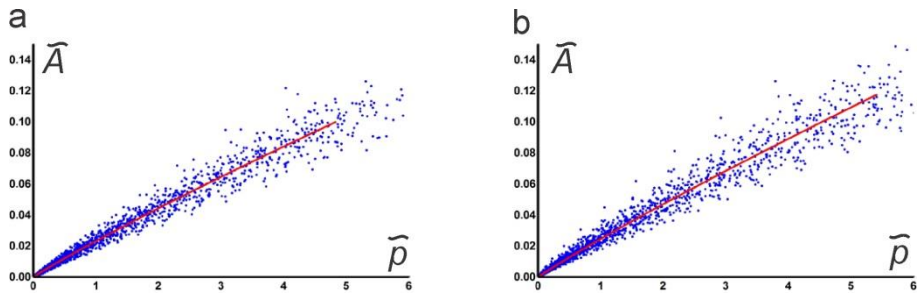


Fig. 8. Results of the nGW-calculations on 100 surfaces simulated by a) the MP-method and b) the RS-method. The red line provides the average of all simulations.

6 SUMMARY AND CONCLUSIONS

- Based on the observation of simulated surfaces alone, the three surface generation algorithms studied here cannot be distinguished and produce isotropic surfaces with a Gaussian distribution of heights and asperity heights. It shall be noted that the experimental measurement used as an example in this paper was anisotropic and showed non-Gaussian distributions, indicating that the fact that the simulated surfaces follow predicted theoretical behaviour does not mean that they are a realistic representation of physical surfaces.
- The random midpoint algorithm produces a power spectrum which closely resembles the power spectrum of measured surfaces and does not show individual frequency peaks. The random spectral method is capable of producing similar spectra if a sufficiently high

number of harmonic components is used. The power spectrum of the bivariate generalised Weierstrass-Mandelbrot surfaces is unrealistic. This is not a problem for contact models which neglect the spatial correlation of the contact areas, such as the Greenwood-Williamson model but may severely affect more advanced models. Moreover, the calculation of the Weierstrass-Mandelbrot function is numerically inefficient: with only 126 harmonics, it requires three times more computation time than the random spectral method with 1000 harmonics. This function should therefore not be used in contact simulations. The midpoint algorithm is more efficient than the spectral method, but may introduce anisotropy effects which have to be further investigated before final conclusions can be reached.

- Although the numerical Greenwood-Williamson model used in this paper is primitive as compared to the standards of modern contact mechanics, it allows identifying the essential phenomena with respect to the problem of statistical variation. Considering the spread on the results, one may consider using less precise methods for contact simulations and instead focus on a higher number of simulations, as long as these are unbiased with respect to the average values obtained. Alternatively, one may increase the size of the simulated surface to capture a wider range of random features and therefore obtain a better average. However, considering that the surfaces studied are non-stationary random processes, increasing the size will also increase statistical variation, while computational cost may increase in a non-proportional manner. This is a topic which has received very little attention and must be addressed in future research.

ACKNOWLEDGMENTS

Financial support of DGAPA under grant IN114215 is acknowledged. The authors wish to thank Miguel Silva and Christian Daniel Maya who participated as students in various aspects of the present work.

REFERENCES

[1] Archard, J. F. Elastic deformation and the laws of friction. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 243, 190-205 1957.

[2] Greenwood, J. A., & Williamson, J. B. P. Contact of nominally flat surfaces. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 295 300-319 1966.

[3] Whitehouse, D. J., & Archard, J. F. The properties of random surfaces of significance in their contact. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 316, 97-121. 1970.

[4] Longuet-Higgins, M.S. The statistical analysis of a random, moving surface. Philosophical transactions of the Royal Society A, 249, 321-387, 1957.

[5] Mandelbrot, B. B. How long is the coast of Britain? Statistical Self-similarity and Fractional dimension. Science, 156, 636-638 1967.

[6] Greenwood, J. A., & Wu, J. J. Surface roughness and contact: an apology. Meccanica, 36, 617-630, 2001.

[7] Majumdar, A., & Bhushan, B. Role of fractal geometry in roughness characterization and contact mechanics of surfaces. Journal of Tribology, 112, 205-216, 1990.

[8] Majumdar, A., & Bhushan, B. Fractal model of elastic-plastic contact between rough surfaces. Journal of Tribology, 113, 1-11, 1991

[9] Persson, B. N. J. "Elastoplastic contact between randomly rough surfaces." Physical Review Letters 87 116101, 2001.

[10] Persson, B. N.J. "Theory of rubber friction and contact mechanics." The Journal of Chemical Physics 115 3840-3861, 2001.

[11] Manners, W., and J. A. Greenwood. "Some observations on Persson's diffusion theory of elastic contact." Wear 261, 600-610, 2006.

[12] Ciavarella, M., Murolo G., Demelio G., Barber, J.R. Elastic contact stiffness and contact resistance for the Weierstrass profile. Journal of the Mechanics and Physics of Solids, 52, 1247-1256, 2004.

[13] Yan W. and Komvopoulos K. "Contact analysis of elastic-plastic fractal surfaces." Journal of Applied Physics 84 3617-3624, 1998.

[14] Yin, X., and Komvopoulos, K. An adhesive wear model of fractal surfaces in normal contact. International Journal of Solids and Structures 47 912–921, 2010

[15] Tao, Q., Lee, H.P. and Lim, S.P. Contact mechanics with various models of roughness descriptions. Wear 249, 539-545, 2001.

[16] Zavarise, G., Borri-Brunetto, and Paggi, M. On the reliablility of microscopical contact models. Wear 257, 229-245, 2004.

[17] Kogut, L. and Jackson, R.L. A comparison of Contact Modelling Utilizing Statistical and Fractal Approaches. Journal of Tribology, 128, 213-217, 2006.

[18] Ciavarella, M., Delfine, V., & Demelio, G. A “re-vitalized” Greenwood and Williamson model of elastic contact between fractal surfaces. Journal of the Mechanics and Physics of Solids, 54, 2569-2591, 2006.

[19] Ciavarella, M., Greenwood, J. A., & Paggi, M. Inclusion of “interaction” in the Greenwood and Williamson contact theory. Wear, 265, 729-734, 2008

[20] Hyun, S., Pei, L., Molinari, J. F., & Robbins, M. O. Finite-element analysis of contact between elastic self-affine surfaces. Physical Review E, 70, 026117, 2004.

[21] Pei, L., Hyun, S., Molinari, J. F., & Robbins, M. O. Finite element modeling of elasto-plastic contact between rough surfaces. Journal of the Mechanics and Physics of Solids, 53, 2385-2409, 2005.

[22] Putignano, C., Afferrante, L., Carbone, G., & Demelio, G. A new efficient numerical method for contact mechanics of rough surfaces. International Journal of Solids and Structures, 49, 338-343, 2012

[23] Putignano, C., Afferrante, L., Carbone, G., & Demelio, G. The influence of the statistical properties of self-affine surfaces in elastic contacts: A numerical investigation. Journal of the Mechanics and Physics of Solids, 60, 973-982, 2012.

[24] Putignano, C., Afferrante, L., Carbone, G., & Demelio, G. P. A multiscale analysis of elastic contacts and percolation threshold for numerically generated and real rough surfaces. Tribology International, 64, 148-154, 2013

[25] Anciaux, G., & Molinari, J. F. Contact mechanics at the nanoscale, a 3D multiscale approach. International journal for numerical methods in engineering, 79, 1041-1067, 2009.

[26] Saupe, T. Algorithms for random fractals. In: The science of fractal images, 71-136, 1988.

[27] Ausloos, M. and Berman, D.H. A multivariate Weierstrass-Mandelbrot function. Proceedings of the Royal Society of London A, 400, 331-350, 1985.

[28] Clarke C.K. Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. Computers & Geosciences, 12, 713-722, 1986.

[29] Zou, G, Lam, N.S.N, A comparison of fractal dimension estimators based on multiple surface generation algorithms. Computers & Geosciences, 31, 1260-1269, 2005.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NUMERICAL OPTIMIZATION OF NEAR-ROAD VEGETATION BARRIERS

Luděk Beneš, Viktor Šíp

Dept. of Technical Mathematics, Faculty of Mechanical Engineering, CTU in Prague
Karlovo nám 13, 1 21 35 Praha 2, Czech republic
e-mail: beneslu1@fs.cvut.cz

Keywords: ABL, Pollution dispersion, Vegetation, Road dust.

Abstract. *The contribution deals with the numerical simulation of the influence of near-road vegetation barriers on the dustiness along the highway. Vegetation barrier close to the road has been proposed to reduce the amount of PM10 and PM75 in nearby areas.*

The flow in the Atmospheric Boundary Layer (ABL) is assumed to be incompressible, yet the density is not constant due to the gravity. The flow is described by the RANS equation for viscous incompressible flow with variable density. The two equations turbulence model is used for the closure of this set of equations. Three effects of the vegetation should be considered: effect on the air flow, i.e. slowdown or deflection of the flow, influence on the turbulence levels inside and near the vegetation, and filtering of the particles present in the flow. Deposition velocity reflects four main processes by which particles depose on the leaves: Brownian diffusion, interception, impaction and gravitational settling.

The numerical scheme is based on the finite volume method and artificial compressibility method. For the convective terms the AUSM+up scheme is used. Second order accuracy is achieved via the linear reconstruction, where gradients are calculated using least squares approach. For the viscous fluxes diamond type scheme is used. Resulting set of ODE equations is integrated using BDF2 method.

49 variants of the vegetation differ in the density, width and height were simulated. The influence of the mentioned parameters were examinated. Main processes affecting the dustiness in modeled cases are emphasized.

1 Introduction

Increasing level of dustiness is one of the major problems mainly in the populated areas. The atmospheric particulate matter (PM) has a significant negative impact on the human health, as was demonstrated in the numerous studies. Apart from industry and local heating, vehicular traffic is one of the major sources of particulates. Near road vegetation barriers were proposed as a means of the reduction of a harmful PM in the atmosphere. Their effectivity is influenced by a number of parameters: atmospheric conditions, properties of the particulates, vegetation type or its position. Experimental studies of this flow both on-site and in wind tunnels are complicated and expensive. One of the appropriate ways how to understand the phenomena of vegetation flows and particle transport are the numerical simulations.

The effect of the vegetation on the air flow alone has been subject of the research for a long time. Although number of studies dealing with near-road flow field, vegetation and pollutant dispersion have been published, research on deposition of the particulate matter on the vegetation is still ongoing. Computationally efficient RANS models of horizontally homogeneous canopy are often used in studies of vegetation barriers [1]. To capture detailed temporal evolution of the flow field, LES models of horizontally homogeneous canopy were also used [2]. Advances in high-performance computing allowed to simulate complex models of detailed 3D canopy architecture resolved to the level of stems and branches [3]. Problems connected with the modeling of the stratified fluid are solved in [4, 5].

Several reviews on related topics are available, dealing with dry deposition on canopies [15] or focusing on a relation between air pollution and vegetation in an urban setting [13, 10]. Among modeling studies on particle dispersion through vegetation barrier are those investigating in 2D and 3D computational domains [17, 22, 19]. Problem of dust transport in real-world situations has been numerically investigated in [7, 8].

In this paper we investigate influence of the vegetation block over the highway notch. Aim of this study is to determine effects of the near-road barrier under varying vegetation parameters, namely width and height of the block and its density. We use simplified 2D model of a horizontally homogeneous vegetation block with original LAD profile. Main objective is to investigate influence of the barrier on pollutant concentration behind the barrier.

2 Numerical model

2.1 Physical and mathematical model

We are interested in the modeling of the flow in the lowest part of ABL, approximately 200 meters over the ground. Velocities here vary in the order of ones or tenth m/s, the flow can be modeled as incompressible, but due to the gravity force the density isn't constant, the fluid is stratified. The second significant problem is the presence and modeling of the vegetation (forest or bushes), which has a major influence on the air flow. Three effects of its should be considered: effect on the air flow, i.e. slowdown or deflection of the flow, influence on the turbulence levels inside and near the vegetation, and the filtering of the particles present in the flow.

2.1.1 Equations for the fluid motion

This type of flow can be described by the Reynolds-averaged Navier-Stokes (RANS) for viscous, incompressible, stratified fluid. These equations are rewritten using Boussinesque hypothesis. The pressure p and potential temperature θ are split into background component in

hydrostatic balance and fluctuations, $p = p_0 + p'$ and $\theta = \theta_0 + \theta'$. Boussinesq approximation stating that changes in density are negligible everywhere except in the gravity term is utilized. Resulting set of equations is written here

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla(p'/\rho_{\text{ground}}) = \nu_E \nabla^2 \boldsymbol{u} + \boldsymbol{g} + \boldsymbol{S_u}, \tag{2}$$

$$\frac{\partial \theta}{\partial t} + \nabla \cdot (\theta \boldsymbol{u}) = \frac{\nu_E}{\text{Pr}} (\nabla \cdot (\nabla \theta)). \tag{3}$$

Vector $\boldsymbol{u}$ stands for velocity, ρ_{ground} is the air density ρ at the ground level, $\nu_E = \nu_L + \nu_T$ is the sum of the laminar and turbulent kinematic viscosity, $\boldsymbol{g} = (0, g\frac{\theta'}{\theta_0}, 0)$ is the gravity term, $\boldsymbol{S_u}$ represent the momentum sink due to the vegetation and $\text{Pr} = 0.75$ is the Prandtl number.

2.1.2 Turbulence

For the closure of previous equations, standard $k-\epsilon$ model is used. Equations for turbulence kinetic energy k and dissipation ϵ are as follows:

$$\frac{\partial \rho k}{\partial t} + \nabla \cdot (\rho k \boldsymbol{u}) = \nabla \cdot \left(\left(\mu_L + \frac{\mu_T}{\sigma_k} \right) \nabla k \right) + P_k - \rho\epsilon + \rho S_k, \tag{4}$$

$$\frac{\partial \rho \epsilon}{\partial t} + \nabla \cdot (\rho \epsilon \boldsymbol{u}) = \nabla \cdot \left(\left(\mu_L + \frac{\mu_T}{\sigma_\epsilon} \right) \nabla \epsilon \right) + C_{\epsilon_1} \frac{\epsilon}{k} P_k - C_{\epsilon_2} \rho \frac{\epsilon^2}{k} + \rho S_\epsilon. \tag{5}$$

The model is completed by a relation between k, ϵ and the turbulent dynamic viscosity μ_T, $\mu_T = C_\mu \rho \frac{k^2}{\epsilon}$. In the equations above μ_L is the laminar dynamic viscosity, P_k is the production of the turbulence kinetic energy, and S_k and S_ϵ are sources of k and ϵ respectively. Both could be written as the sum of a road traffic part and a part due to the vegetation, $S_k = S_k^r + S_k^v$, $S_\epsilon = S_\epsilon^r + S_\epsilon^v$. Sources due to the road traffic are modeled by the model from [6]. Sinks and sources caused to the vegetation are described in the next section.

Following constants of the model are used: $\sigma_k = 1.0$, $\sigma_\epsilon = 1.167$, $C_{\epsilon_1} = 1.44$, $C_{\epsilon_2} = 1.92$ and $C_\mu = 0.09$.

2.1.3 Particle transport

Pollution dispersion is calculated using transport equation for non dimensional mass fraction w,

$$\frac{\partial \rho w}{\partial t} + \nabla \cdot (\rho w \boldsymbol{u}) - (\rho w u_s)_y = \nabla \cdot \left(\frac{\nu_E}{\text{Sc}} \nabla \rho w \right) + \rho f_c + S_w. \tag{6}$$

Here f_c is the source term and S_w is the vegetation deposition term. Based on the review and the discussion in [23], the Schmidt number $\text{Sc} = 0.72$ was used. The settling velocity u_s of a spherical particle with the diameter d and density ρ_p is given by the Stokes' equation, $u_s = (d^2 \rho_p g C_c)/(18\mu)$, with the correction factor $C_c = 1 + \frac{\lambda}{d}(2.34 + 1.05\exp(-0.39d/\lambda))$, where $\lambda = 0.066$ μm is the mean free path of the particle in the air [9].

2.1.4 Vegetation

Vegetation model plays an important role. We suppose vegetation as horizontally homogenous, its properties are described by vertical *Leaf area density* (LAD) profile - foliage surface area per unit volume - and a leaf type and size. We have used original LAD profile shown in Figure 1. This basic LAD profile is multiplied by the coefficient representing the density of vegetation.

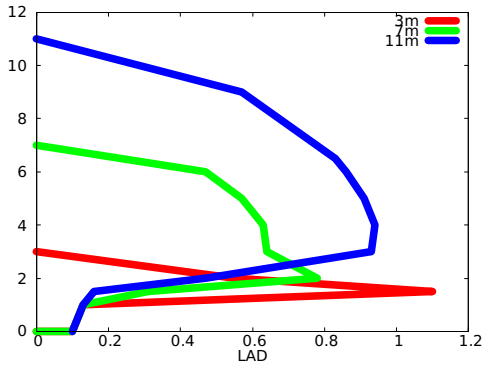


Figure 1: LAD profile

The effects of the vegetation can be divided into three parts: first, it is drag induced by the vegetation and this caused sink in the momentum equation, (2)

$$\boldsymbol{S}_u = -C_d \mathrm{LAD} |\boldsymbol{u}| \boldsymbol{u}$$

Here $C_d = 0.3$ is the drag coefficient [11].

Secondly, it is the influence on the turbulence levels. Following [11], we model this term as

$$S_k^v = C_d \mathrm{LAD}(\beta_p |\boldsymbol{u}|^3 - \beta_d |\boldsymbol{u}| k), \quad S_\epsilon^v = C_{\epsilon_4} \frac{\epsilon}{k} S_k^v,$$

in Eqs. (4) and (5) with constants $\beta_p = 1.0$, $\beta_d = 5.1$ and $C_{\epsilon_4} = 0.9$.

And lastly, it is a particle sink term in Eq. (6), $S_w = -\mathrm{LAD} u_d \rho w$. The term is proportional to the deposition velocity u_d. Deposition velocity reflects four main processes by which particles depose on the leaves: Brownian diffusion, interception, impaction and gravitational settling. Its value generally depends on wind speed, particle size and vegetation properties. In this study, the model from [16] derived for broadleaf trees was used.

2.2 Numerical method

For the numerical solution the finite volume method in connection with artificial compressibility method were used. The continuity equation (1) is rewritten in the next form,

$$\frac{1}{\beta} \frac{\partial p'}{\partial t} + \nabla \cdot \boldsymbol{u} = 0. \tag{7}$$

The choice of the parameter β is discussed e.g. in [14], here we have used $\beta = 1000$.

Transformed set of equations is discretized using the finite volume method on unstructured grid. For the convective terms the AUSM+up scheme [12], designed for all speed flows, is used. Second order accuracy is achieved via the linear reconstruction, where gradients are calculated

using least squares approach. To prevent artificial overshooting, Venkatakrishan limiter [24] is utilized. For discretization of viscous terms a dual mesh is used (diamond type scheme).

The resulting set of ordinary differential equations is solved using an implicit BDF2 method of the second order. In every time step, first the system of the Navier-Stokes equations (2, 3, 7) is solved, followed by the system of the $k - \epsilon$ equations (4, 5) and then by the system of the passive scalar equations (6). Values of turbulent viscosity, coupling together turbulence equations with the Navier-Stokes equations, are taken from the previous time step.

Each of these nonlinear systems is solved by the Newton method. Inner linear systems are solved using matrix-free GMRES solver. The linear systems are preconditioned by ILU(3) preconditioner. To construct the preconditioner matrices, Jacobian of the system is needed. Its evaluation via finite differences is a costly operation, which we reduce by two complementing approaches: via matrix coloring, which exploit the sparseness of the Jacobian, and by calculating the preconditioner matrices (as well as the Jacobians) only every 20 iterations.

3 Numerical results

Large number of parameters play role in model presented above. Sensitivity analysis were performed in our previous studies [20], [21], where the influence of various types of meteorological conditions and geometrical configurations were tested. Based on this computations, following test case was chosen as a representative.

3.1 Case settings

The computational domain is $350\ m$ long ($x\ \in <\ -50, 300\ >\ m$) and $100\ m$ high and represents highway notch. Schematic sketch of the domain is shown in Figure 2. Road is situated between $20\ m$ and $45\ m$, side slopes in the angle $45°$ are $4\ m$ heigh and are placed on both sides of the highway in $x\ \in <\ 16, 20\ >\ m$ and $x\ \in <\ 50, 54\ >\ m$. Four sources of pollutant, representing the road, are placed between 23 m and 42 m from the inlet at the height of 0.8 m. Vegetation block is placed downstream from the road above the notch and starts in $x = 55$m.

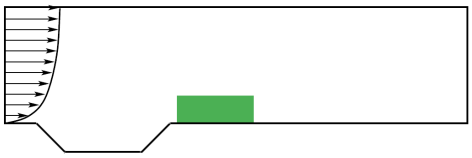


Figure 2: Sketch of the domain (not to scale)

We model the particles of diameter 10 µm and 75 µm with density 1000 kg/m^3. Each source of the pollutant has the intensity 1 µg/s. No resuspension of the particles fallen on the ground is allowed. Density of the traffic is set to 16 passenger cars and 4 heavy duty vehicles per minute in each of the four lanes.

The ABL is supposed as weakly stable stratified ($\partial T/\partial y$ = 0 K/m). As was shown in [20], the stratification plays a minor role in this type of computation. Computational domain is relatively short and the mechanical turbulence caused by passing cars plays the dominant role in the initial mixing of pollutants.

49 different geometrical configuration were simulated. We have tested all combinations of following parameters of vegetation plus case without the vegetation.

• Width: 50m, 80m, 110m, 140m.
• Hight: 3m, 7m, 11m.
• Density 0.25, 0.5, 1.0, 1.5.

3.2 Boundary conditions

• **Inlet:** logarithmic wind profile is prescribed at the inlet with u_{ref} = 5 m/s at height y_{ref} = 10 m, $v = 0$. Roughness parameter z_0 is set to 0.1 m. Homogeneous Dirichlet b.c. for concentration is satisfied. Pressure is extrapolated from the domain, temperature is set to the $T = 20°$.
• **Top:** velocity on the top is given by the Dirichlet b.c., pressure is extrapolated from the domain, concentration $C = 0$ and temperature $T = 20°$.
• **Bottom:** on the ground, the non–slip boundary conditions for velocity component are prescribed. Homogeneous Neumann b.c. for pressure and temperature are supposed, all particles fallen on the ground stay here for all time.
• **Outlet:** Homogeneous b.c. for velocity components, temperature and concentration are supposed. Pressure is prescribed by the barometric formula.

For the turbulence equations, boundary conditions and wall functions according to [18] are used.

3.3 Computational mesh

Computational domain consist of app. 3410^3 cells and it's generated by the *snappyHexMesh*. The mesh is exponentially refined. While on the top of the domain the cell size is app. $2x3$ m, close to the ground and forest it is app. $0.3x0.45$ m. On the ground the boundary layer is made from 3 cells, the first cell is 0.06 m thick.

3.4 Computational results

Typical flow fields are demonstrated in Figure 3. On the left hand side the situation with relatively sparse forest is shown, the flow is decelerated without recirculation. On the other hand, in the case of dense vegetation, massive recirculation zone was developed behind the block.

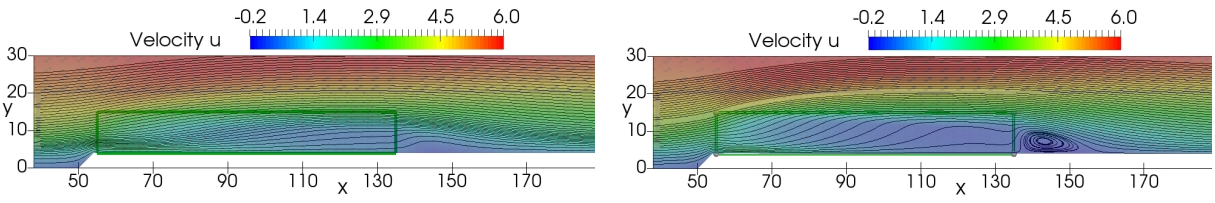


Figure 3: Streamlines for the case of forest 80m wide and 11m high. Density factor 0.5 left and 1.5 right.

The PM concentration in the fixed point $x = 250$m and 3m above the ground is shown in Fig.4 (PM10) and Fig.5 (PM75). In both cases the dependency on the width of vegetation is significant for low vegetation, for high is close to the constant and short forest is sufficient. Unsurprisingly, efficiency is higher for heavier particles, where concentration reduces to $1/9$, whereas to $7/10$ for lighter ones.

Different processes and behavior we can study in Figures 6 - 9, where the horizontal and vertical distributions of pollutant are shown. The horizontal concentrations are drown along the line 3m above the ground, the vertical along the line at $x = 250$m. For the heavier particles

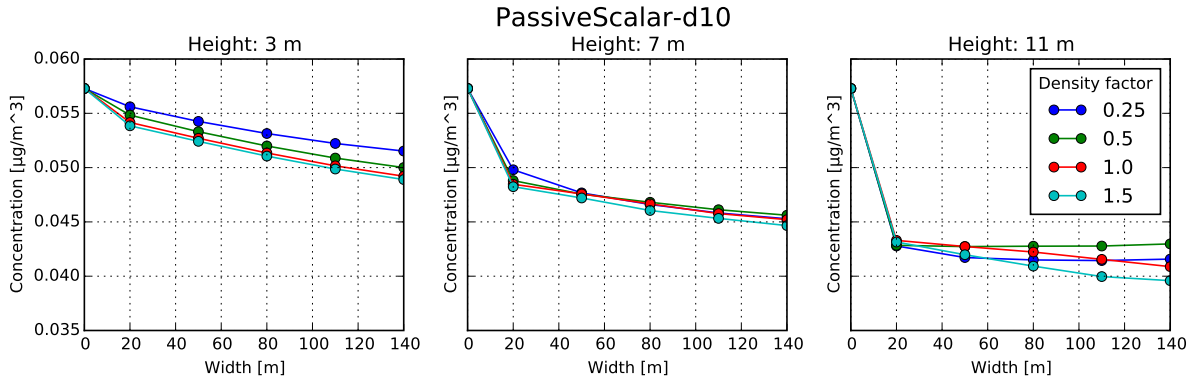


Figure 4: Concentration at $x = 250$m, 3m above the ground for $d = 10$µm

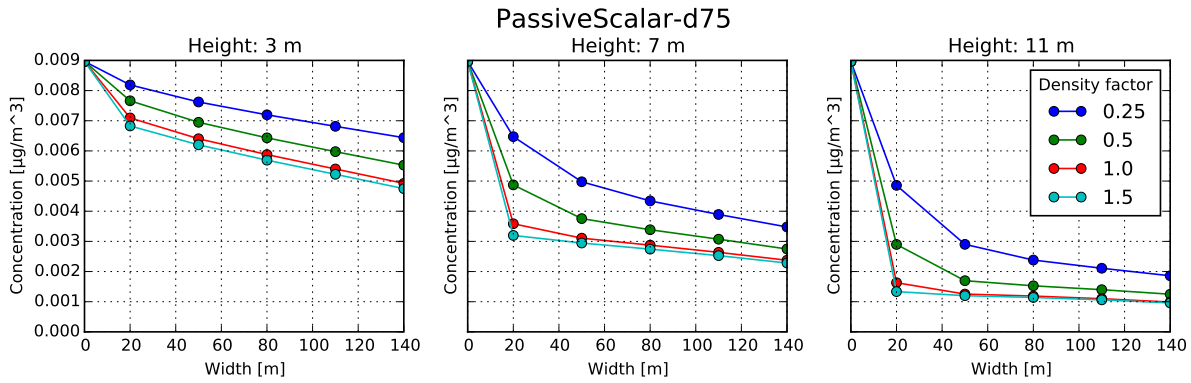


Figure 5: Concentration at $x = 250$m, 3m above the ground for $d = 75$µm

the reduction of concentration is monotonous, see Fig.7. The reduction significantly depends on forest heights, Fig. 9, for low vegetation it is insignificant, for height 11m the reduction for sparse forest is close to the 50%, for dense $80 - 90\%$, for wider forest is close to 90% independently on the density of vegetation. The main effect is the deceleration of the flow inside the vegetation and the consequent longer time for deposition of particles.

The case of lighter, PM10 particles, is more complex, Figs. 6, 8. The monotonical decreasing of concentration is visible for low or sparse vegetation only. The overall effect is however small in those cases. The profile of concentration inside the forest in the case of higher and dense vegetation is different. The recirculation zone close to the forest, the deposition velocity is small (5.5mm/s in comparison to 300mm/s for PM75), the particles accumulate inside the forest. On the other hand the deflection of the flow by the obstacle and increase of the turbulence which leads to spread of the particles to the higher parts of the atmosphere play the major role here. As it is well illustrates in Fig. 8, the reduction of the concentration close to the ground is caused mainly by the transportation of the dust at higher altitudes.

4 Discussion

A method for evaluation of the effects of vegetation barriers on pollution dispersion was developed and used for modeling of the PM10 and PM75 concentration emitted from the highway. In this study effects of height, width and density of leafy vegetation were examined.

Basic processes playing important roles in the deposition and transportation of the particles were identified and shown. The height of vegetation is more important then its length, for higher

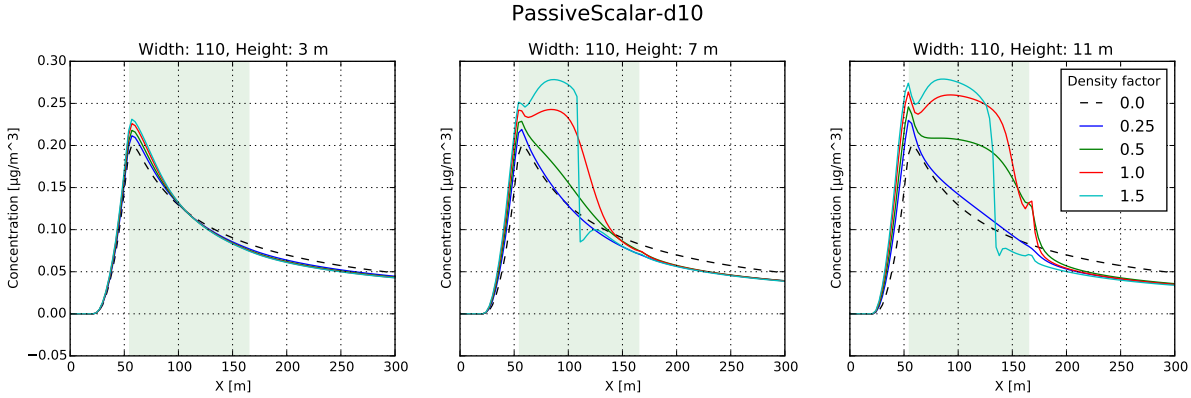


Figure 6: Horizontal profiles of concentration 3m above the ground for $d = 10$µm and block width 110m

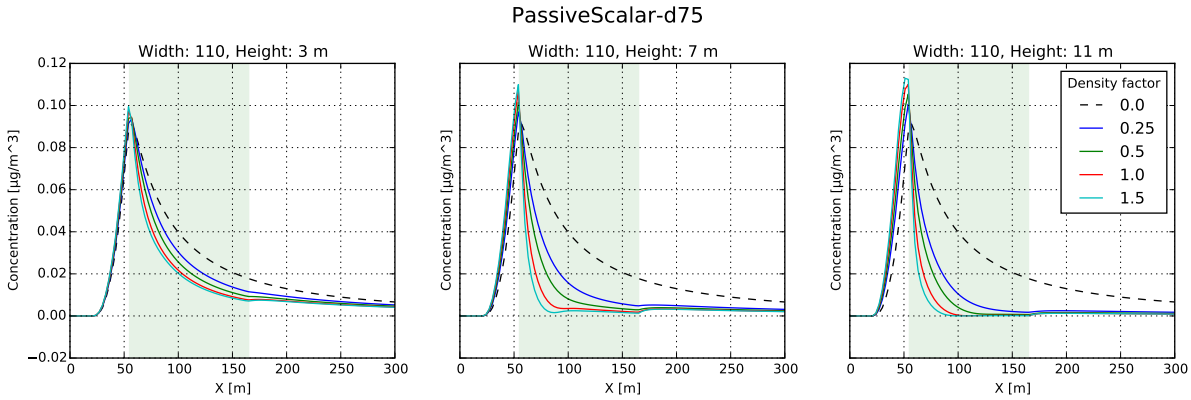


Figure 7: Horizontal profiles of concentration 3m above the ground for $d = 75$µm and block width 110m

forest the dependence on length is insignificant, the low forest is ineffective. The deposition plays significant role for heavier PM75 particles and the vegetation is very efficient due to deceleration of the flow. In the case of lighter PM10 particles, the significant effect is deflection of the flow and increasing of turbulence by the vegetation, which leads to the spreading of pollutant to the higher parts of atmosphere.

It is necessary to say, the results are very sensitive to exact capture of the fluid flow and parametrisation of the vegetation plays significant role.

5 Acknowledgment

This work was supported by the grant SGS13/174/OHK2/3T/12 of the Czech Technical University in Prague..

REFERENCES

[1] G.G. Katul, L. Mahrt, D. Poggi, and C. Sanz: One- and two-equation models for canopy turbulence. Bound. Layer Meteor. 113 (2004), 81-109.

[2] L. Fitzmaurice, R.H. Shaw, K.T. Paw U, and E.G. Patton: Three-dimensional scalar microfront systems in a large-eddy simulation of vegetation canopy flow. Bound. Layer Meteor. 112 (2004), 107-127.

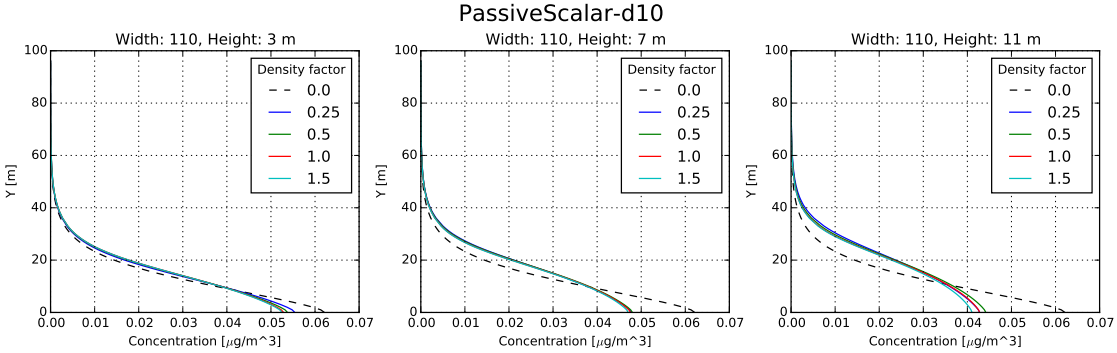


Figure 8: Vertical profiles of concentration at $x = 250$m for $d = 10$μm and block width 110m

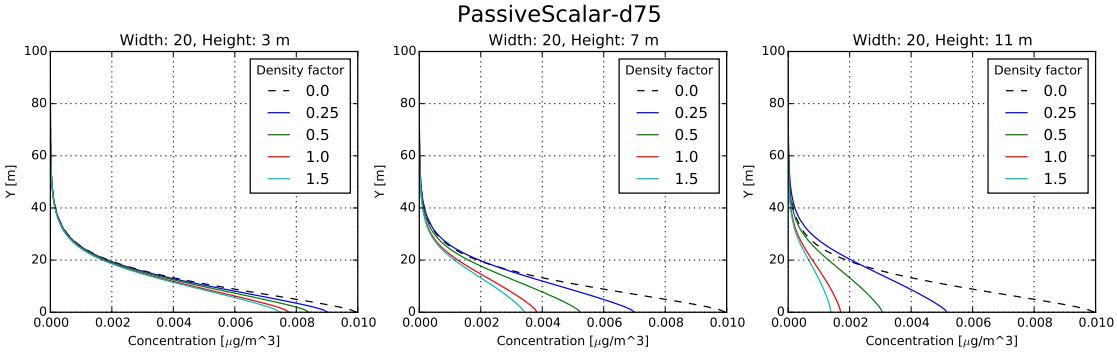


Figure 9: Vertical profiles of concentration at $x = 250$m for $d = 75$μm and block width 20m

[3] A.M. Endalew, M. Hertog, M.A. Delele, K. Baetens, T. Persoons, M. Baelmans: H. Ramon, B.M. Nicolaï, and P. Verboven. CFD modelling and wind tunnel validation of airow through plant canopies using 3D canopy architecture. Int. J. Heat Fluid Flow 30 (2009), 356-368.

[4] L. Beneš, T. Bodnár, and K. Kozel.: Comparison of two numerical methods for the stratified flow. *COMPUTERS & FLUIDS,*, 46(1):148–154, 2011. ISSN 0045-7930, DOI: 10.1016/j.compfluid.2011.02.003.

[5] T. Bodnár, L. Beneš, Ph. Fraunié, K. Kozel: Application of compact finite-difference schemes to simulations of stably stratified fluid flows. Applied Mathematics and Computation 219(7), 2012, 3336–3353.

[6] D. BÄUMER, B. VOGEL, AND F. FIEDLER, *A new parameterisation of motorway-induced turbulence and its application in a numerical model*, Atmos. Environ. **39**:31 (2005), 5750–5759.

[7] L. Matejicek, Z. Janour, L. Benes, T. Bodnar and E. Gulikova: Spatio-temporal modelling of dust transport over surface mining areas and neighbouring residential zones. Sensors 8 (2008), 3830-3847.

[8] L. Beneš, T. Bodnár and K. Kozel: Numerical simulation of the neutrally stratified ABL flow over complex geometry. In International Conference of Computational Methods in Sciences and Engineering, American Institute of Physics (2009).

[9] W. HINDS, *Aerosol technology: Properties, behavior, and measurement of airborne particles*, Wiley, 1999.

[10] S. JANHÄLL, *Review on urban vegetation and particle air pollution - deposition and dispersion*, Atmos. Environ. **105** (2015), 130–137.

[11] G. KATUL, L. MAHRT, D. POGGI, AND C. SANZ, *One- and two-equation models for canopy turbulence*, Bound. Layer Meteor. **113** (2004), 81–109.

[12] M.-S. LIOU, *A sequel to AUSM, part II: AUSM+-up for all speeds*, J. Comput. Phys. **214** (2006), 137–170.

[13] T. LITSCHKE AND W. KUTTLER, *On the reduction of urban particle concentration by vegetation - a review*, Meteorol. Z. **17** (2008), 229–240.

[14] F. MULDOON AND S. ACHARYA, *A modification of the artificial compressibility algorithm with improved convergence characteristics*, Int. J. Numer. Meth. Fl. **55**:4 (2007), 307–345.

[15] A. PETROFF, A. MAILLIAT, M. AMIELH, AND F. ANSELMET, *Aerosol dry deposition on vegetative canopies. Part I: Review of present knowledge*, Atmos. Environ. **42** (2008), 3625–3653.

[16] A. PETROFF, L. ZHANG, S. PRYOR, AND Y. BELOT, *An extended dry deposition model for aerosols onto broadleaf canopies*, J. Aerosol Sci. **40**:3 (2009), 218–240.

[17] M. RAUPACH, N. WOODS, G. DORR, J. LEYS, AND H. CLEUGH, *The entrapment of particles by windbreaks*, Atmos. Environ. **35** (2001), 3373–3383.

[18] P. RICHARDS AND R. HOXEY, *Appropriate boundary conditions for computational wind engineering models using the k-ϵ turbulence model*, J. Wind Eng. Ind. Aerodyn. **46 & 47** (1993), 145–153.

[19] J. STEFFENS, Y. WANG, AND K. ZHANG, *Exploration of effects of a vegetation barrier on particle size distributions in a near-road environment*, Atmos. Environ. **50** (2012), 120–128.

[20] V. ŠÍP AND L. BENEŠ, *Modelling the effects of a vegetation barrier on road dust dispersion*, Appl. Mech. Mater. **821** (2016), 105–112.

[21] V. ŠÍP AND L. BENEŠ, *CFD Optimization of a Vegetation Barrier*, Proceedings of **ENUMATH** (2016), Springer (in press).

[22] A. TIWARY, H. MORVANB, AND J. COLLS, *Modelling the size-dependent collection efficiency of hedgerows for ambient aerosols*, Aerosol Science **37** (2005), 990–1015.

[23] Y. TOMINAGA AND T. STATHOPOULOS, *Turbulent Schmidt numbers for CFD analysis with various types of flowfield*, Atmos. Environ. **41** (2007), 8091–8099.

[24] V. VENKATAKRISHNAN, *Convergence to steady state solutions of the euler equations on unstructured grids with limiters*, J. Comput. Phys. **118** (1995), 120–130.

Equation Chapter 1 Section 1ECCOMAS Congress 2016
Equation Chapter 1 Section 1 VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

LARGE EDDY SIMULATIONS OF MAGNETIC FIELD EFFECT ON TURBULENT FLOW IN A SQUARE DUCT

Jie Mao[1], Kunlei Zhang[1], Zhongquan Tan[1] and Ke Liu[2]

[1] School of Mechanical Engineering, Hangzhou Dianzi University
Hangzhou, China
e-mail: maojie@hdu.edu.cn, zhangkunlei@126.com, tzqfly2009@163.com

[2] School of Energy & Power Engineering, Dalian University of Technology
Dalian, China
sdqdllk@163.com

Keywords: Magnetohydrodynamic duct flow, Turbulent, Large Eddy Simulation

Abstract. *Numerical simulation of magnetohydrodynamic turbulent duct flow is important in the application of liquid metal fusion blanket. We develop a large eddy simulation magneto-hydrodynamic turbulent flow solver based on finite volume method and dynamic Smagorinsky subgrid model in OpenFOAM environment. The induced magnetic field is negligible as the magnetic Reynolds number is small. The electric potential Poisson equation has been used to solve the electromagnetic field. The consistent and conservative method has been used to solve the electric potential, the electric current and the Lorentz force. Periodic boundary condition has been used at the inlet and outlet. We investigate the effects of applied magnetic field on mean velocity field, fluctuation velocity, and turbulent kinetic energy of the liquid metal fluid flow. The secondary flow generated by the anisotropic turbulence stress has been shown obviously. As the Hartmann number increases the turbulent kinetic energy decreases, which shows the turbulence suppression effects of the external magnetic field.*

1 INTRODUCTION

The study of magnetohydrodynamic (MHD) turbulent flow is essential in the application of liquid metal pumps, MHD power generators, and liquid metal blankets of thermal nuclear fusion reactors[1]. Turbulent MHD duct flows have been investigated extensively by numerical simulation and experiment in these years [2-8]. Zikanov [9] reviewed the laminar-turbulent transition of MHD flows in duct, pipe, and channel.

Numerical investigation of MHD duct flow has been implemented by direct numerical simulation (DNS) [2], large eddy simulation (LES) [5, 7, 10] and Reynolds-averaged Navier-Stokes (RANS) method [11]. Researchers concerned the effects of the external magnetic field, wall conductance ratio, the cross-section of the duct on the MHD turbulent flow. Large eddy simulation is a promising method to investigate the MHD turbulent flow with limited grid and calculation time. There are different subgrid-scale(SGS) models in the simulation of MHD turbulent flow such as Smagorinsky model (SM), coherent structure Smagorinsky model (CSM) and dynamic Smagorinsky model (DSM) [6, 7]. The study of Kobayshi [6] and Kransnov [4] shows that the DSM can predict the MHD flow transformation accurately.

We have developed MHD solver with dynamic Smagorinsky model in OpenFOAM environment. MHD turbulent flow in a insulating square duct has been simulated to investigate the effect of the external magnetic field on the turbulent flow.

2 GOVERNING EQUATIONS AND NUMBERICAL METHODS

2.1 Governing equations

The incompressible MHD duct flow subject to an uniform magnetic field parallel to insulated walls is governed by the mass conservative equation, the Navier-Stokes equation with Lorentz force, the Ohm's law, electric current conservation and the electric potential Poisson equation,

$$\nabla \bullet \mathrm{u} = 0\,, \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} = -\nabla p + \nabla\cdot\nu\nabla\mathbf{u} + \frac{1}{\rho}\mathbf{j}\times\mathbf{B}\,, \tag{2}$$

$$\mathrm{j} = \sigma(-\nabla\phi + \mathrm{u}\times\mathrm{B})\,, \tag{3}$$

$$\nabla\cdot\mathrm{j} = 0\,, \tag{4}$$

$$\nabla^2\phi = \nabla\cdot(\mathbf{u}\times\mathbf{B})\,. \tag{5}$$

Here $\mathbf{u}$ is the velocity vector, t is the time, p is the pressure divided by the density, ν is the kinematic viscosity, $\mathbf{j}$ is the electric current density, ρ is the density, $\mathbf{B}$ is the applied external magnetic field, σ is the electrical conductivity and ϕ is the electric potential.

In the magnetohydrodynamic flow, there are several important dimensionless parameters. The magnetic Reynolds number $Re_m = \mu\sigma L v_0$, μ is the magnetic permeability. In fusion blanket application, the magnetic Reynolds number is less than 1. Therefore, we could neglect the induced magnetic field in the Ohm's law. The Reynolds number $Re = v_0 D/\nu$, D is hydrodynamic diameter, v_0 is character velocity, which in general is the mean velocity of the

flow. The interaction parameter or Stuart number $N = \sigma LB^2 / \rho v_0$. The Hartmann number is defined as $Ha = BL\sqrt{\sigma/\rho\nu} = \sqrt{Re \cdot N}$.

The spatial filtering of unsteady Navier-Stokes equation is:

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\overline{u}_i \overline{u}_j\right) = -\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j \partial x_j} - \frac{\partial \tau_{ij}}{\partial x_j} + \frac{1}{\rho}\left[\mathbf{j} \times \mathbf{B}\right]_i, \tag{6}$$

where τ_{ij} is the subgrid-scale stresses,

$$\tau_{ij} = \overline{u_i u_j} - \overline{u}_i \overline{u}_j = \tau_{ij}^* + \frac{1}{3}\delta_{ij}\tau_{kk}. \tag{7}$$

2.2 Dynamic Smagorinsky SGS model

In LES, the SGS stress is modeled as a single SGS turbulence model. Germano [12] proposed the dynamic Smagorinsky SGS model. In the DSM, τ_{ij}^* is modeled with the filter width $\overline{\Delta}$ as:

$$\tau_{ij}^* = -2\nu_t \overline{S_{ij}} = -2\left(C_{SGS}\Delta_1\right)^2 \overline{|S|}\,\overline{S_{ij}}. \tag{8}$$

Here, C_{SGS} is the Smagorinsky constant, the velocity-strain tensor for the resolved component $\overline{S_{ij}}$ and its magnitude $\overline{|S|}$ are defined by

$$\overline{S_{ij}} = \frac{1}{2}\left(\frac{\partial \overline{u_j}}{\partial x_i} + \frac{\partial \overline{u_i}}{\partial x_j}\right), \tag{9}$$

$$\overline{|S|} = \sqrt{2\overline{S_{ij}}\,\overline{S_{ij}}}. \tag{10}$$

The model parameter of the DSM is determined using a least square procedure proposed by Lilly [13] with an average in homogeneous directions,

$$C_{SGS}^2 = \frac{1}{2}\frac{\left\langle L_{ij} M_{ij} \right\rangle}{\left\langle M_{ij} M_{ij} \right\rangle}. \tag{11}$$

The angular brackets $\langle\ \rangle$ indicate an average procedure over the cell face. L_{ij} and M_{ij} are given by:

$$L_{ij} = \widehat{\overline{u}_i \overline{u}_j} - \hat{\overline{u}}_i \hat{\overline{u}}_j \tag{12}$$

$$M_{ij} = \Delta_1^2 \widehat{\overline{|S|}\,\overline{S}_{ij}} - \Delta_2^2 \hat{\overline{|S|}}\,\hat{\overline{S}}_{ij} \tag{13}$$

$$\Delta_1^3 = \Delta_x \Delta_y \Delta_z, \qquad \Delta_2 = 2\Delta_1. \tag{14}$$

Here, Δ_x, Δ_y, Δ_z are element sizes in x, y, z directions, respectively.

The Van Driest function is used for the wall damping corrections, which is expressed as:

$$f = 1 - \exp(-y^+ / 26). \tag{15}$$

2.3 Numerical model and boundary conditions

An incompressible electric conducting flow in a square duct subject to external uniform magnetic field has been simulated. The duct is electric insulating. The magnetic field is applied in y direction. The x, y and z are streamwise, Hartmann layer and side layer directions, respectively. The calculation domain is $(2\pi,1,1)$. The Reynolds number is $\mathrm{Re}=6400$, where $D=1$. The Hartmann number equals 8, 20 and 25.6. Therefore, the corresponding ratios of $R=Re/Ha$ are 800, 320 and 220 respectively.

At the wall, no-slip boundary condition is applied:

$$\mathbf{u}=0. \tag{16}$$

As the wall is electric insulating, no electric current penetrates the wall,

$$\partial\phi/\partial n=0, \tag{17}$$

here n is the normal direction of the wall.

At the inlet and outlet, periodic boundary condition is used for the velocity, the pressure and the electric potential.

The domain is divided by $256\times96\times96$ grids. A uniform grid is used along the streamwise direction and nouniform grids are used along y, z direction. The spatial resolutions of Δy^+, Δz^+ are ranged from $1.03\sim8.224$ with stretching 3%. The time step is 1.0×10^{-3} seconds, and the statistic properties are obtained by averaging 500 seconds data after the flow reaches a statistically steady state.

Consistent and conservative methods on a rectangular collocated grid [14] have been used to solve the electric current, the electric Potential equation and Lorentz force. The the mass conservative equation and the momentum equation are solved by the PISO algorithm in OpenFOAM.

In order to keep a constant flow rate, a constant pressure gradient corrected by the flow rate at the cross section is added in the Navier-Stokes equation.

The initial field is perturbed by adding 2% random fluctuation of the average velocity in three directions.

3 RESULTS AND DISCUSSION

The instantaneous axial velocity contours with transverse velocity vectors and the time-averaged axial velocity with transverse velocity vectors at the cross section at the middle of the duct are presented in Figure 1 and Figure 2 for $Ha=20$. Figure 1 shows that there is a high speed axial velocity core in the centre of the duct. The secondary velocities are generated especially near the parallel duct wall, which accompanies the reduction of the axial velocity. Figure 2 shows the time statistical information of the velocity. It shows that the secondary velocities direct from centre to the four corners and form eight vortexes at the corners. As the external magnetic field is along y direction, the velocity component along z is reduced remarkably because of the Lorentz force. The vortex formed by the secondary flow is bigger near the Hartmann walls than that near the side walls.

Figure 3 shows the mean axial velocity along horizontal bisector. With the same Reynolds number, when the Hartmann number is small, the MHD effect is not apparently, the difference of the mean velocity distributions along y, z is quite small. With the external magnetic field increasing, the MHD effect results that the distribution of the mean axial velocity along the parallel external magnetic field is more circular than that along the vertical magnetic field direction.

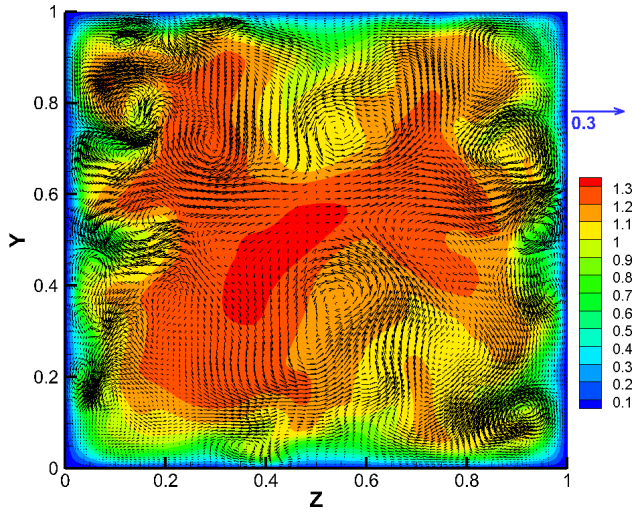


Figure 1 Instantaneous axial velocity contours and secondary velocity vectors for *Re*=6400, *Ha*=20.

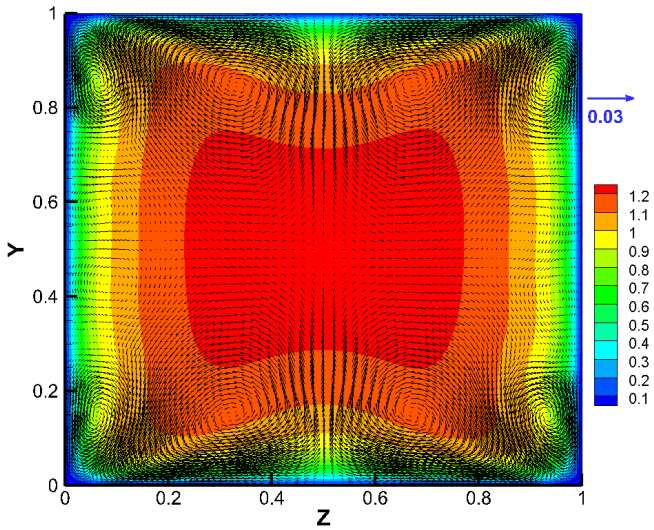


Figure 2 Mean axial velocity contours and secondary velocity vectors for *Re*=6400, *Ha*=20.

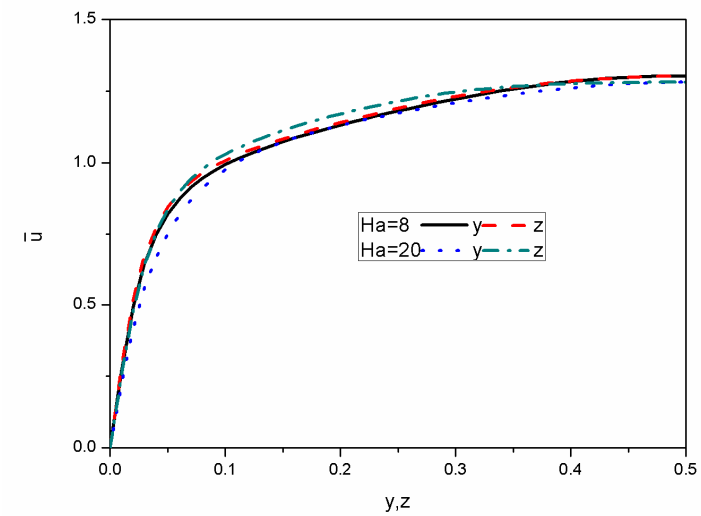


Figure 3 Mean axial velocity along horizontal bisector.

Figure 4 shows the turbulent kinetic energy along horizontal bisector. As $Ha = 8$, the turbulent kinetic energy is almost the same along y, z. The effect of the magnetic field is not obvious. As Hartmann number increases, the turbulent kinetic energy decreases obviously. Furthermore, the turbulent kinetic energy along y is remarkably small than that along z directions. This is caused by the MHD effects. The main induced electric current returns through the Hartmann layer, while the main induced electric current parallel to the external magnetic field in the side layers. The MHD effect results that the fluctuations of the velocity are suppressed by the Lorentz force.

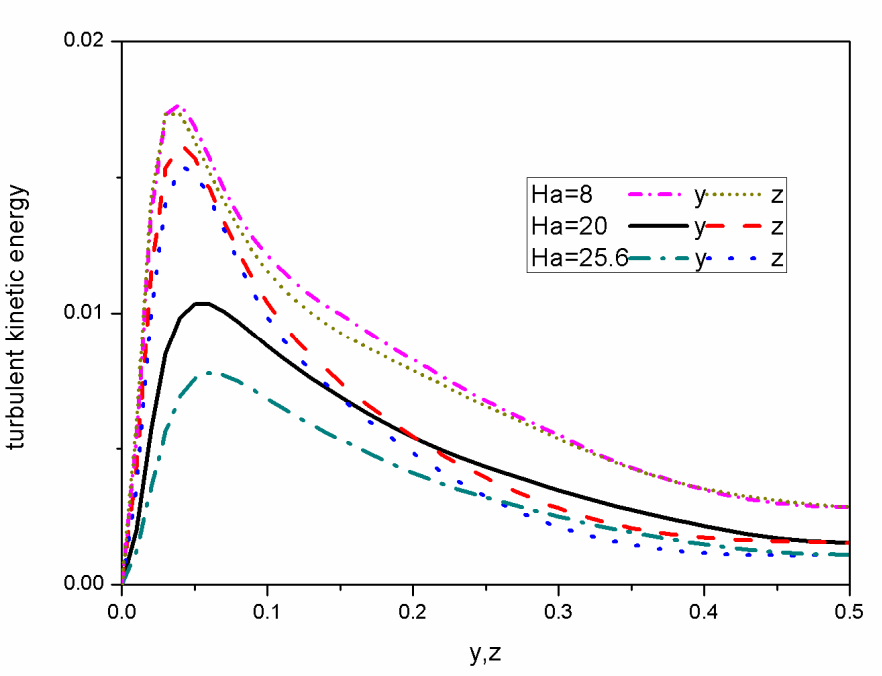


Figure 4 Turbulent Kinetic energy along horizontal bisector.

4 CONCLUSIONS

- A MHD turbulent LES solver has been developed in OpenFOAM environment.
- Dynamic Smagorinsky model has been used to simulate the MHD square duct flow with electric insulating walls.
- MHD duct flows with the same Reynolds number and different Hartmann numbers have been numerically simulated to investigate the effect of the external magnetic field on the turbulent flow.
- External magnetic field depresses the turbulent fluctuation especially on the Hartmann layer direction.

ACKNOWLEDGEMENTS

The authors are grateful to Ville Vuorinen at Aalto University, Humberto Medina at Coventry University for the helpful discussion and comments.

This work is supported by the National Nature Science Foundation of China (NSFC) under Grant No. 11375049 and the National Magnetic Confinement Fusion Science Program of China under Grant No. 2014GB125003.

REFERENCES

[1] U. Müller and L. Bühler, Magnetofluiddynamics in channels and containers. *Springer Verlag*,2001.

[2] D. Krasnov, O. Zikanov, and T. Boeck, Numerical study of magnetohydrodynamic duct flow at high reynolds and hartmann numbers. *Journal of Fluid Mechanics*, **704**, 421-446, 2012.

[3] D. Krasnov, P. Parepalli, O. Zikanov, and T. Boeck, Effect of wall conductivity on current distribution and turbulence in a channel flow with spanwise magnetic field. *European Journal of Mechanics - B/Fluids*, **30**(4), 421-427, 2011.

[4] D. Krasnov, O. Zikanov, J. Schumacher, and T. Boeck, Magnetohydrodynamic turbulence in a channel with spanwise magnetic field. *Physics of Fluids*, **20**(9)2008.

[5] H. Kobayashi, Large eddy simulation of magnetohydrodynamic turbulent duct flows. *Physics of Fluids*, **20**(1), 015102, 2008.

[6] H. Kobayashi, Large eddy simulation of magnetohydrodynamic turbulent channel flows with local subgrid-scale model based on coherent structures. *Physics of Fluids*, **18**(4), 045107, 2006.

[7] H. Kobayashi, The subgrid-scale models based on coherent structures for rotating homogeneous turbulence and turbulent channel flow. *Physics of Fluids* **17**(4), 045104, 2005.

[8] R. Chaudhary, S. P. Vanka, and B. G. Thomas, Direct numerical simulations of magnetic field effects on turbulent flow in a square duct. *Physics of Fluids*, **22**(7), 075102, 2010.

[9] O. Zikanov, D. Krasnov, T. Boeck, A. Thess, and M. Rossi, Laminar-turbulent transition in magnetohydrodynamic duct, pipe, and channel flows. *Applied Mechanics Reviews*, **66**(3), 030802, 2014.

[10] B. Knaepen and P. Moin, Large-eddy simulation of conductive flows at low magnetic reynolds number. *Physics of Fluids*, **16**, 1255, 2004.

[11] S. Smolentsev, M. Abdou, and N. Morley, Application of the ‘‘ k – ε ’’ model to open channel flows in a magnetic field. *International Journal of Engineering Science*, **40**, 693-711, 2002.

[12] M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, A dynamic subgrid‐scale eddy viscosity model. *Physics of Fluids A: Fluid Dynamics (1989-1993)*, **3**(7), 1760-1765, 1991.

[13] D. K. Lilly, A proposed modification of the germano subgrid‐scale closure method. *Physics of Fluids A: Fluid Dynamics (1989-1993)*, **4**(3), 633-635, 1992.

[14] M.-J. Ni, R. Munipalli, N. B. Morley, P. Huang, and M. a. Abdou, A current density conservative scheme for incompressible mhd flows at a low magnetic reynolds number. Part i: On a rectangular collocated grid system. *Journal of Computational Physics*, **227**(1), 174-204, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

OPTIMAL DESIGN AND ADDITIVE MANUFACTURING OF NOVEL REINFORCING ELEMENTS FOR COMPOSITE MATERIALS

F. Fabbrocino[1], I. Farina[1,2], A. Amendola[3], L Feo[3], F.Fraternali[3]

[1]Pegaso University
Department of Engineering
Piazza Trieste e Trento, 48, 80132 Naples, Italy
francesco.fabbrocino@unipegaso.it

[2]University of Naples Parthenope
Department of Engineering
Centro Direzionale di Napoli, isola C4, 80143- Naples, Italy
ylenia.farina@gmail.com

[3]University of Salerno
Department of Civil Engineering
Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
adamendola@gmail.com, l.feo@unisa.it, f.fraternali@unisa.it

Keywords: Fibres, Fibre/matrix bond, Fracture toughness, Surface analysis, Additive manufacturing

Abstract. *We experimentally investigate on the use of additive manufacturing technologies for the design and fabrication of innovative reinforcing elements of novel composite materials. We perform short-beam shear tests on cement mortar specimens reinforced with additively manufactured reinforcing fibers made of photopolymers or a titanium alloy. The fracture toughness, shear capacity and first crack strength of the examined materials are estimated based on the provisions of different international standards for construction materials. We also characterize the surface morphology of the examined fibers through microscopy analyses before and after testing. The given results highlight that the microscopic or macroscopic nature of the surface roughness of the analyzed fibers greatly influences the energy absorption capacity of the final materials, while the nature of the fibers' material (metallic/polymeric) is of central importance in terms of strength properties. The present study represents a first step in the direction of designing reinforcing elements with hierarchical structure to form fabrics, fibers and coatings of groundbreaking reinforcements for next generation composites, profiting from the rapid prototyping capabilities of additive manufacturing technologies at different scales.*

1 INTRODUCTION

The nascent field of structural "metamaterials", i.e. artificial materials showing unconventional properties mainly derived by their geometric design is growing rapidly and attracting increasing attention from many research areas (refer, e.g., to [1] and references therein). In recent years, additive manufacturing (AM) has become the most common technique for fabricating materials that exhibit unusual behaviors, which are not found in natural materials. Several additive manufacturing methods have been proposed in this field, with resolution ranging from the centimeter- to the nanometer-scale. Worth mentioning here are: polyjet 3D printing technologies; electron beam melting; x-ray lithography; deep ultraviolet lithography; soft lithography; two-photon polymerization; atomic layer deposition; and projection micro-stereolithography, among other available methods (refer to [2][3][4][5][6] and references therein). However, the potential of rapid prototyping in the design of novel mechanical metamaterials is not completely understood at present, and there is an urgent need for research exploring the suitability of such techniques for the manufacture of real life engineering materials.

Most natural shapes exhibit hierarchical organization of matter and fractal geometries, which provide increased surface area for the same volume of material. Recent studies have shown that hierarchical composites showing multiscale fibers coated with carbon nanotubes feature enhanced interlaminar shear strength (ILSS), a key property of composite materials, which can be weak in the presence of smooth matrix-fiber interfaces [7][8]. In such materials, the development of rough fracture surfaces near to the surface of the reinforcing elements, along with crack deflection mechanisms, enhances matrix toughness. The increase in the surface roughness of the reinforcing elements, as compared to smooth interfaces, delays the matrix failure and improves surface energy dissipation [7]. In addition, the pull-out of fine-scale features of the reinforcements bridges the matrix, significantly contributing to the enhancement of composite strength and toughness [8].

The present study investigates the use of additively manufactured reinforcing elements with multiscale geometry for the reinforcement of cementitious mortars. Fibers with structural hierarchy originating from their geometric design are manufactured from computer-aided design (CAD) data, employing additive manufacturing techniques based on polymeric and metallic materials. A first source of hierarchical architecture arises from using fibers covered by lattices on a smaller scale [9]. A second strategy employs meshes with fractal geometry (Figure 1) [10]. Reinforcements with hierarchical structure are combined with a cement mortar, in order to obtain advanced composites with enhanced ILSS, and enhanced first-crack strength and fracture toughness.

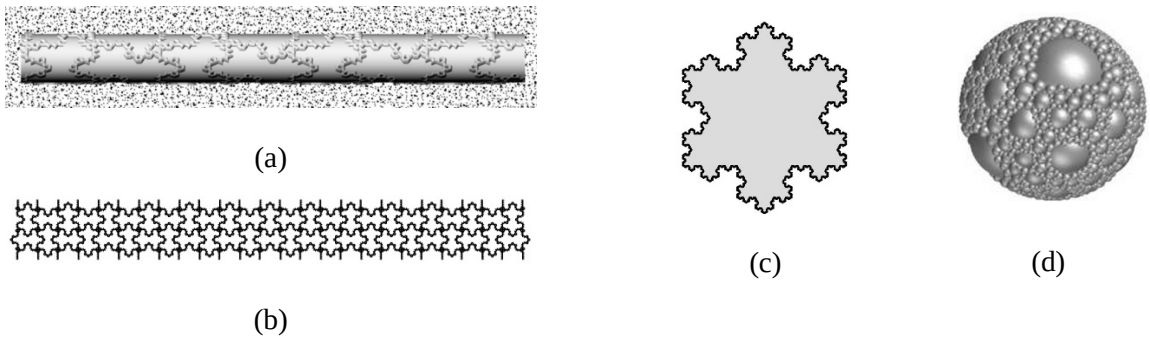


Figure 1: Lattices with fractal geometry used to form reinforcements of composite materials: (a) fiber coating; (b) fabric; (c) fiber cross-section; (d) junction element/fiber embossment (Apollonian sphere packing) [10].

We begin in Sect. 2 by describing the preparation of the mortar and fibers analyzed in the present study. Next, we analyze the first-crack strength and toughness of fiber-reinforced mortar specimens through three point bending tests (Sect. 3). We continue by analyzing the morphology of the fibers' surface in the virgin state and after their pull-out from the matrix (Sect. 4). We end in Sect. 5 with concluding remarks and an outline of future work to be carried out to deepen the potential of AM for the optimal design of novel composite materials.

2 CEMENT MORTAR REINFORCEMENT WITH 3D PRINTED FIBERS

Let us examine the fiber reinforcement of a pre-mixed cement-based mortar matrix of class M5 [11] with mechanical properties shown in Table 1. We employ reinforcing fibers fabricated through two different AM technologies: PolyJet 3D printing of liquid photopolymers through the Objet500 Connex commercial printer by Stratasys®, and electron beam melting (EBM) of high-strength metallic materials through the Arcam EBM S12 facility available at the Department of Materials Science and Engineering of the University of Sheffield.

We printed six polymeric fibers in the photopolymer transparent resin Fullcure 720 (see Table 1 for mechanical properties), with 7.5 mm diameter and 100 mm length, Three of these fibers show smooth lateral surface (hereafter denoted as "Pol_S" fibers, see Figure 2a), while another three are coated on the lateral surface with a fractal lattice based on the Koch snowflake ("Pol_R" fibers, see Figure 1a and Figure 2b).

Metallic fibers were also produced, employing EBM to manufacture four cylindrical fibers in the titanium alloy Ti6Al4V [6], with 7.5 mm diameter, 160 mm length, and smooth lateral surface ("Ti_S" fibers, see Figure 3a,c and). We also employed EBM to produce four 7.0 mm diameter Ti6Al4V fibers, coated with a 0.75 mm × 0.75 mm grid of cylindrical embossments. These cylinders exhibit 0.20 mm diameter and 0.50 mm length ("Ti_R" fibers, see Figure 3b,d and Figure 4: Cross-section (a) and 3D view (b) of a portion of the Ti_R fiber).

A microscope characterization of the surface morphology of the examined fibers is given in Sect. 4. Such fibers were employed to reinforce prismatic specimens of a cement mortar with square cross-section, 40 mm width and 160 mm length. Mortar specimens were manufactured by adding 180 cc of water for each kg of the pre-mixed cement mortar, according to manufacturer's recommendations. We set the mortar cover of Pol_S and Pol_2 fibers equal to 20 mm (effective depth equal to 200 mm: fibers placed at mid-height), and that of Ti_S and Ti_R fibers equal to 7 mm (effective depth equal to 33 mm). The above fiber placements were aimed at reproducing pure crack-bridging reinforcement in the case of the low modulus Pol_S and Pol_R fibers, and combined shear-flexure reinforcement in the case of the high modulus Ti_S and Ti_R fibers (see Table 1). All the specimens were cured at room temperature for 28 days before testing. We use the nomenclature Pol_R, Pol_S, Ti_R, and Ti_S to denote the mortar specimens reinforced with the corresponding 3D printed fibers, and unreinforced mortar specimens by the symbol UNR. We manufactured three Pol_R, Pol_S, Ti_R and Ti_S specimens each, and four UNR specimens.

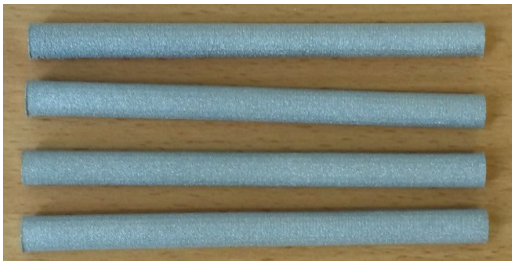


(a)

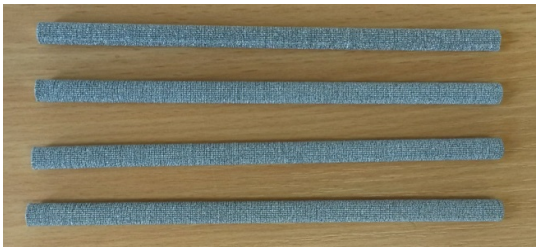


(b)

Figure 2: Photographs of Pol_S (a) and Pol_R (b) fibers.



(a)



(b)

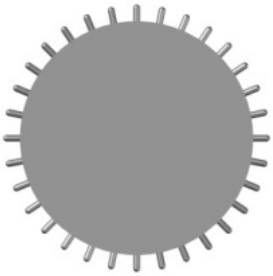


(c)

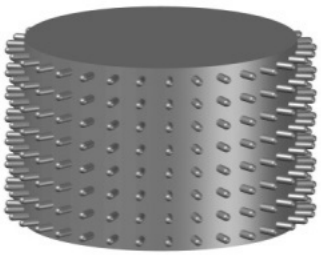


(d)

Figure 3: Top and side views of Ti_S (a,c) and Ti_R (b,d) fibers.



(a)



(b)

Figure 4: Cross-section (a) and 3D view (b) of a portion of the Ti_R fiber.

Class M5 mortar	
mass density [kg/m^3]	1515
compressive strength [MPa]	5
flexural strength [MPa]	1
Fullcure 720 Stratasys® photopolymer	
mass density [kg/m^3]	1.18-1.19
tensile strength [MPa]	50-65
modulus of elasticity [MPa]	2000-3000
Ti6Al4V titanium alloy	
mass density [kg/m^3]	4430
tensile strength [MPa]	950
modulus of elasticity [GPa]	120

Table 1: Mechanical properties of mortar and fibers.

3 THREE POINT BENDING TESTS

We studied the mechanical response of the examined fiber-reinforced mortars by carrying out three- point bending (TPB) tests in displacements control, with 0.25 mm/min loading rate. For each examined specimen, we first determined the applied load versus mid-span deflection curve, and next we computed the first crack strength, shear capacity and material toughness according to the methods specified in the international standards for construction materials. The load-deflection curves obtained for Pol_S and Pol_R specimens are illustrated in Figure 5 and Figure 6, respectively, while the analogous curves competing to Ti_S and Ti_R specimens are given in Figure 7 and Figure 8, respectively. Figure 9- Figure 11 illustrate the synchronization of frames taken from in-situ videos of the TPB tests and the load-deflection curve of some Pol_R, Ti_S and Ti_S specimens [12]. Finally, Figure 12 provides pictures of the configurations after TPB testing of selected specimens.

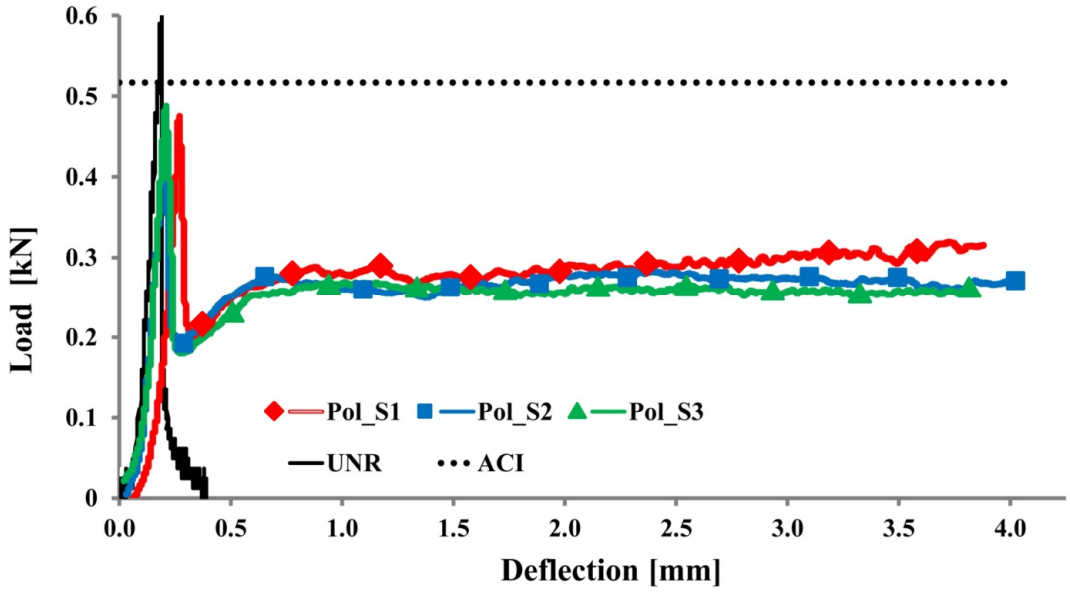


Figure 5: Force vs. deflection curves of Pol_S specimens (UNR: averaged response of UNR specimens).

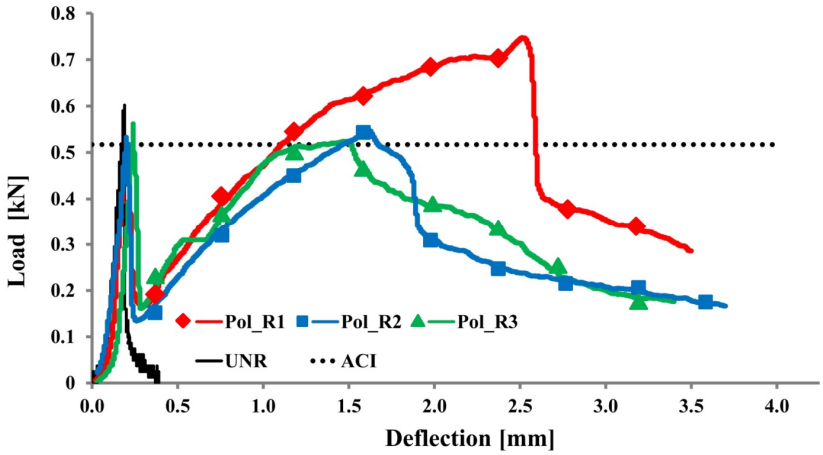


Figure 6: Force vs. deflection of Pol_R specimens (UNR: averaged response of UNR specimens).

The results in Figure 5 and Figure 6 highlight that the maximum load carried by UNR specimens is approximatively equal to 0.6 kN, and is reached just before crack onset. Such specimens exhibit brittle failure, and fast snapping to a collapsed configuration with zero residual strength after crack onset (see Fig. 12a). In contrast to this, Pol_S specimens exhibit residual load carrying capacity approximately constant and equal to 0.25 kN in the post-cracking regime (Figure 5). The load-deflection curves of Pol_R specimens show a marked load drop after crack onset, which is followed by a hardening branch, and next by a second load drop and a softening branch leading to specimen failure (Figure 6). In such specimens, we observed a shear-type failure affected by diagonal cracks propagating from the point of application of the vertical load (see Figure 9 and Fig. 12c), while in Pol_S specimens we observed a flexural-type failure due to the vertical propagation of the central crack up to failure (no diagonal cracks, see Fig. 12b).
Figure 7 and Figure 8 illustrate the results of TPB tests on Ti_S and Ti_R specimens, respectively. Both Ti_S and Ti_R specimens show force-deflection response characterized by a load drop at crack onset; a slightly hardening response after such a load drop; a second load drop and a final plateau or softening branch leading to specimen failure.

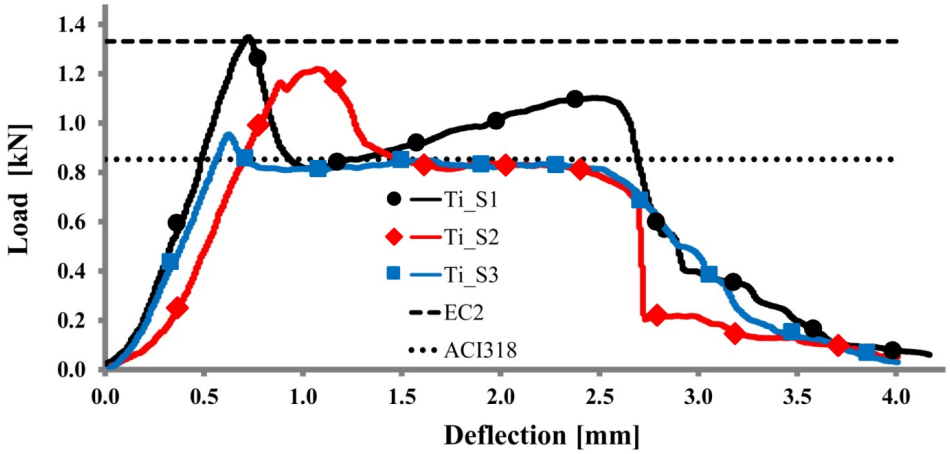


Figure 7: Force vs. deflection curves of Ti_S specimens.

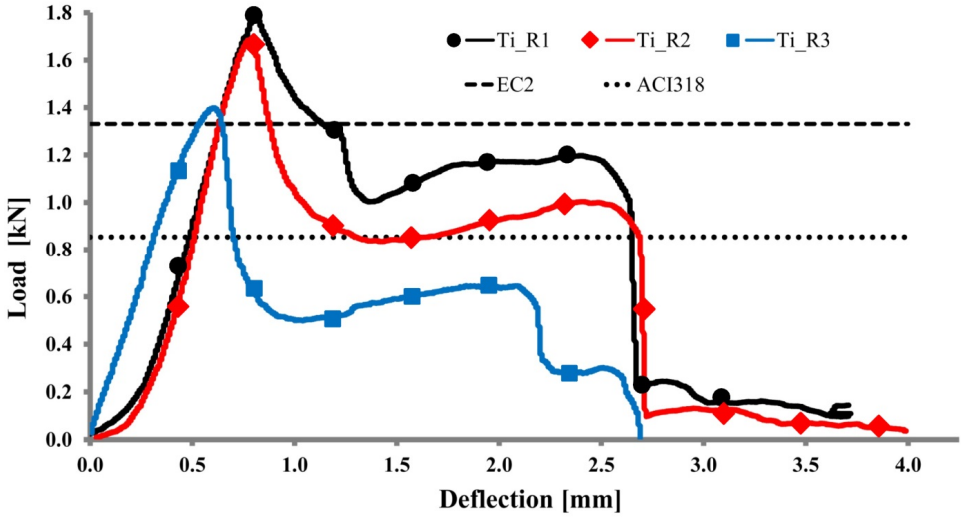


Figure 8: Force vs. deflection curves of Ti_R specimens.

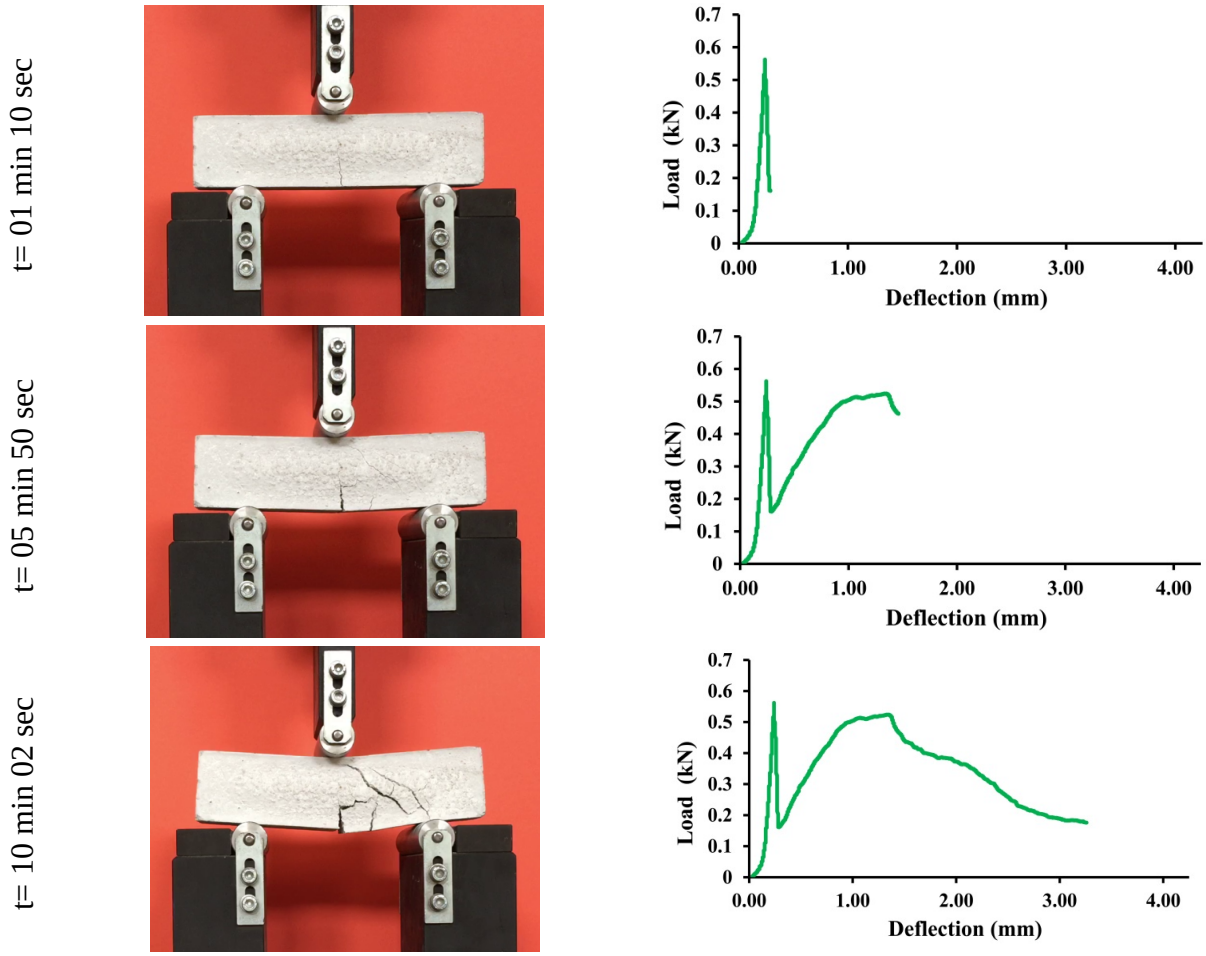


Figure 9: Synchronization between frames from an in-situ video of a TPB on the Pol_R specimen #3 (left) and the load-deflection curve of the same specimen (right).

The maximum load carried by Ti_S specimen (at crack opening) is about equal to 1.1 kN on average (Figure 7), while the maximum load carried by Ti_R specimens is about equal to 1.6 kN on average (Figure 8). It is worth noting that both Ti_S and Ti_R specimens exhibit maximum load carrying capacity that is more than twice that of Pol_S (~0.45 kN, cf. Figure 5), Pol_R (~0.55 kN, cf. Figure 6) and UNR (~0.60 kN, cf. Figure 5-Figure 6) specimens. Both Ti_S and Ti_R specimens exhibited shear-type failure, as shown in Figure 10, Figure 11, and Figure 12d-e. The optical microscopy analyses presented in Sect. 4 show that Ti_S fibers exhibit considerable surface roughness (not included in the CAD design), as well as Ti_R fibers, due to material processing defects (like, e.g., internal porosity) common in the EBM process [6].

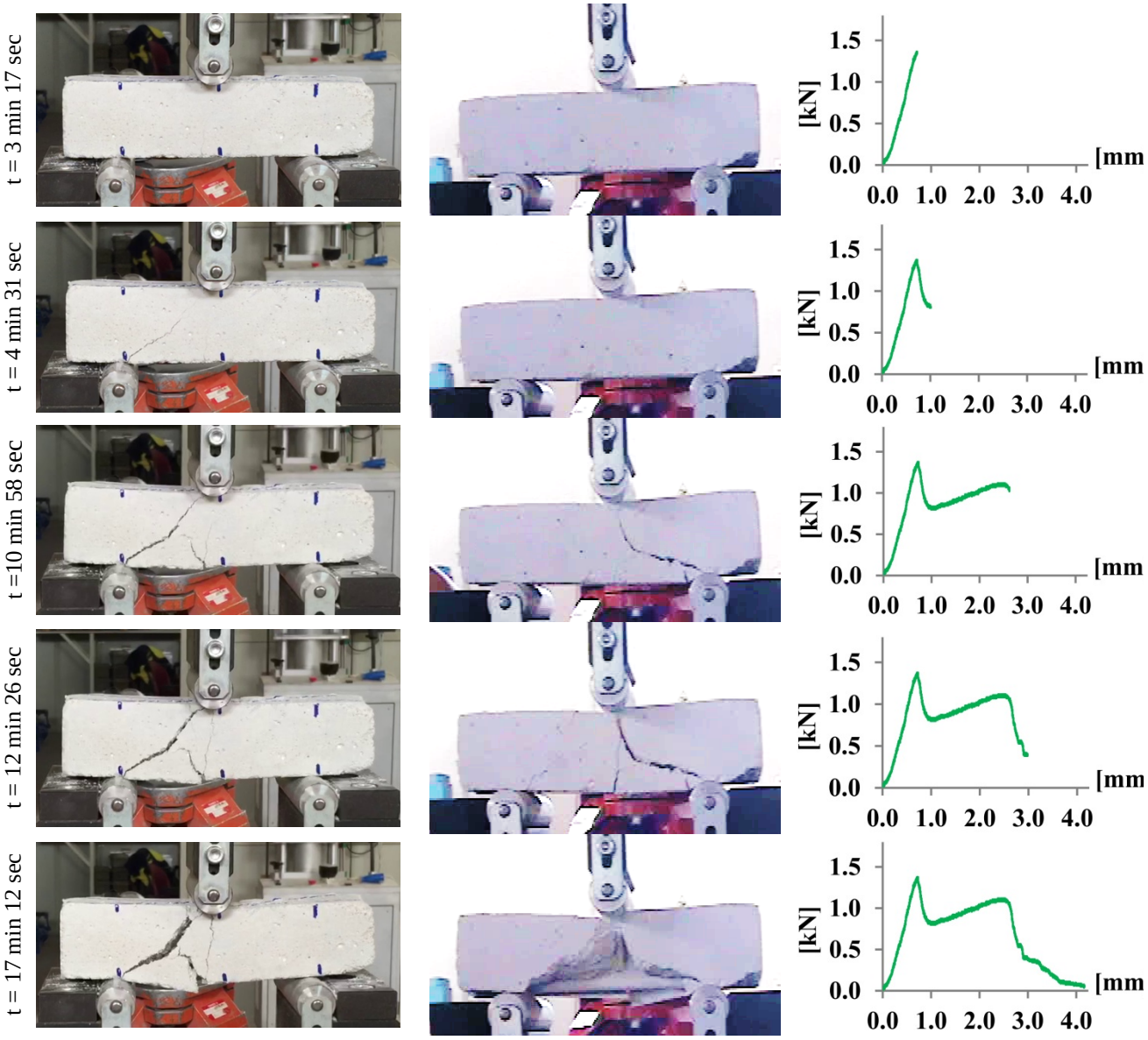


Figure 10: Synchronization of frames from in-situ videos of a TPB on the Ti_S specimen #1 (left: front view, center: back view) and the load-deflection curve of the same specimen (right).

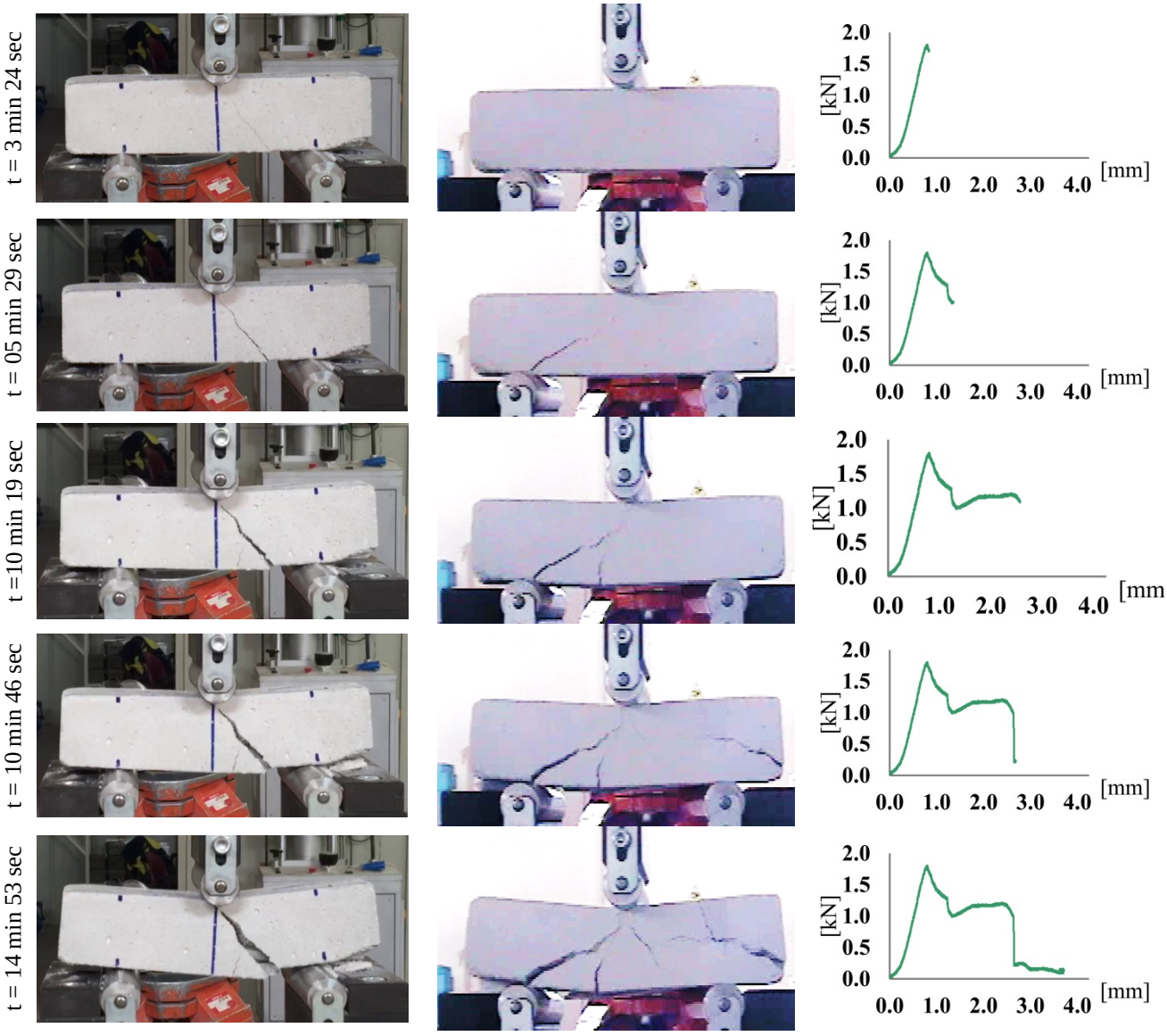
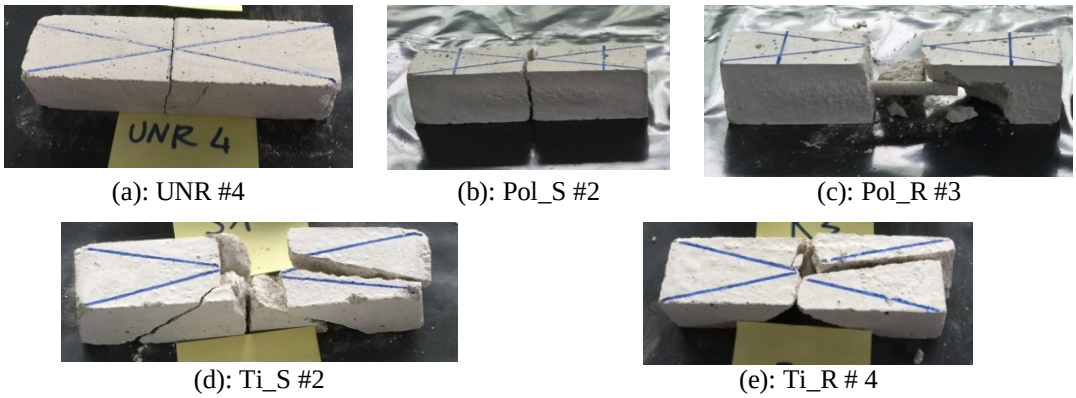


Figure 11: Synchronization of frames from in-situ videos of a TPB on the Ti_R specimen #1 (left: front view, center: back view) and the load-deflection curve of the same specimen (right).



(a): UNR #4 (b): Pol_S #2 (c): Pol_R #3

(d): Ti_S #2 (e): Ti_R # 4

Figure 12: Pictures of different specimens taken after the completion of TPB tests.

3.1 Fracture toughness

By extending the provisions of ASTM C1018 [13] to the current materials, we characterize the fracture toughness of the fiber-reinforced mortars examined in the present study through the following index:

$$I = \frac{A\left(\delta = 3\bar{\delta}\right)}{A_{UNR}\left(\bar{\delta}\right)}, \tag{1}$$

where $\bar{\delta}$ denotes the mid-span deflection in correspondence with the first crack load of the generic specimen; $A\left(\delta = 3\bar{\delta}\right)$ denotes the area under the load-deflection curve from the origin up to $\delta = 3\bar{\delta}$; and $A_{UNR}\left(\bar{\delta}\right)$ denotes the mean value of the area under the load-deflection curve of UNR specimens up to crack opening. The above index represents a measure of the relative energy absorption capacity of the fiber-reinforced materials against the unreinforced mortar. Table 2 gives the toughness index I of the examined mortar specimens, and the mean values of such a quantity that we recorded for each different material.

specimen	I	mean value		specimen	I	mean value
Pol_S_1	3.12	3.12		Ti_S_1	39.08	41.63
Pol_S_2	3.24			Ti_S_2	51.59	
Pol_S_3	3.01			Ti_S_3	34.23	
Pol_R_1	2.36	2.87		Ti_R_1	68.15	51.15
Pol_R_2	3.13			Ti_R_2	52.87	
Pol_R_3	3.12			Ti_R_3	32.44	

Table 2: Fracture toughness indices of the examined specimens.

The results in Table 2 show that Pol_S and Pol_R specimens exhibit a considerable energy absorption capacity, as compared to unreinforced specimens, while Ti_S and Ti_R specimens exhibit extremely large values of I, as compared both to the unreinforced mortar analyzed in the present study, and the fiber reinforced mortars and concretes analyzed in [14][15][16][17][18][19][20]. It is worth noting, however, that most of specimens analyzed in the present work exhibit flexural collapse under deflections lower than $5.5\,\bar{\delta}$, which make it impossible to compute the toughness index I_{10} and the residual strength factor $R_{5,10}$ [14][15] for such materials. In other words, the materials examined in the present study exhibit large toughness and residual strength in the first post-crack regime [16][17], but brittle response for very large deflections (increased brittleness in additively manufactured Ti6Al4V is frequently observed at high strains, due to the low defect tolerance of the alloy [21]). We remark, however, that the deflection regime $\delta \le 3\bar{\delta}$ characterizes a wide range of real-life applications dealing with small or moderately large strains of construction materials.

3.2 Shear capacity and first crack strength

As we already noticed, Pol_R, Ti_S and Ti_R specimens exhibited shear failure under TPB tests, with propagation of diagonal cracks from the point of application of the external load down to the hinge supports (cf. Figure 9-12) [22][23]. On extending results for reinforced concretes without shear reinforcements [22][23] to the present mortars, we hereafter analyze the provisions of different international standards for the shear capacity of such mate-

rials. The shear capacities predicted by ACI 318 [24], BS 8110-1[25], and EC2 [26] are respectively given by the following formulas

$$V_{Rd}^{ACI} = 0.17\phi_c \sqrt{f_{ck}}\, bd \tag{2}$$

$$V_{Rd}^{BS} = \frac{0.79}{\gamma_M}\left(100\frac{A_s}{bd}\right)^{\frac{1}{3}}\left(\frac{400}{d}\right)^{\frac{1}{4}}\left(\frac{f_{ck}}{25}\right)^{\frac{1}{3}} bd \tag{3}$$

$$V_{Rd}^{EC2} = \min\left\{\frac{0.18}{\gamma_c}k\left(100\frac{A_s}{bd}\, f_{ck}\right)^{1/3},\ 0.035\left(1+\sqrt{\frac{200}{d}}\right)^{\frac{3}{2}}\sqrt{f_{ck}}\right\} bd \tag{4}$$

where: A_s and d are the fiber cross sectional area and the effective depth of the cross-section (cross-section height minus the mortar cover), respectively; f_{ck} denotes the characteristic compressive strength of the mortar declared by the manufacturer ($f_{ck} = 5.00$ MPa, cf. Table 1); and it results: $\gamma_M = 1.25$, $\phi_c = 0.85$, $\gamma_c = 1.5$.

Table 3 compares the predictions of the shear capacities of the examined materials with the mean values of the experimental capacities $\overline{V}_{\exp}$ corresponding to one half of the peak loads observed in Figure 5 - Figure 8 (averaged among all specimens). The results in Table 3 show that ACI 318 [24] best matches experimental results for Pol_R and Pol_S specimens (0.26 kN vs 0.27-0,.30 kN), while EC2 [26] instead best matches the experimental shear capacity of Ti_S and Ti_R specimens (0.67 kN vs 0.65-0.83 kN). It should be remarked that the above codes do not account for the surface microstructure and roughness of the reinforcing elements.

specimens	d	A_s	$\overline{V}_{\exp}$	V_{Rd}^{BS}	V_{Rd}^{ACI}	V_{Rd}^{EC2}	$\bar{f}_{cr}$
	$[mm]$	$[mm^2]$	$[kN]$	$[kN]$	$[kN]$	$[kN]$	$[MPa]$
Pol_S	20	44.18	0.27	0.79	0.26	0.53	1.25
Pol_R	20	44.18	0.30	0.79	0.26	0.53	1.42
Ti_S	33	38.48	0.65	1.15	0.42	0.67	1.45
Ti_R	33	38.48	0.83	1.15	0.42	0.67	1.97

Table 3: Comparison between experimental and theoretical predictions of shear capacity and first crack strength.

We now investigate on the first crack strength of the analyzed materials by computing the maximum tensile stress f_{cr} carried by the mortar in correspondence with the load F_{cr} that produces crack initiation. The bending moment associated with such a load is trivially equal to $M_{cr} = F_{cr}L/4$, L denoting the clear length of the specimen ($L = 100\ mm$). On assuming linear elastic behavior of the material up to crack initiation and homogenized properties for the fiber-reinforced cross-section [26], we obtain the mean values of f_{cr} given in Table 3. Such results correspond to assuming the fiber-mortar homogenization factor equal to E_f / E_m, where E_f is the Young modulus of the fiber material (see Table 1), and E_m is the Young modulus of the mortar. The latter has been computed as $E_m \cong 1000\, f_{ck}$ [27], obtaining: $E_m \sim 5000$ MPa. The results in Table 3 show that all the fiber reinforced specimens exhibit first crack strength greater than the flexural strength of the mortar declared by the manufac-

turer (1.0 MPa, cf. cf. Table 1). The maximum value of such a property is exhibited by Ti_R specimens, (1.97 MPa), which is considerably larger than the first crack strength of Ti_S (1.45 MPa), Pol_S (1.25 MPa) and Pol_R (1.42 MPa) specimens.

4 SURFACE MORPHOLOGY OF 3D PRINTED FIBERS BEFORE AND AFTER TESTING

Optical microscopy was employed to investigate on the morphology of the fiber surface at the virgin state and after their pull-out from the matrix. Figure 12 shows optimal microscope images of the examined fibers taken with the optical microscope Olympus SZ-PT using 1500× magnification, before their insertion into the matrix (a), and after fiber pull-out (b). Grayscale versions of the images in Figure 12 were obtained through ImageJ (http://imagej.nih.gov/ij/), a public domain Java image processing software. Each image in Figure 12 was captured and converted from color to grayscale, by application of a grey value threshold and elliptical fit (Figure 13).

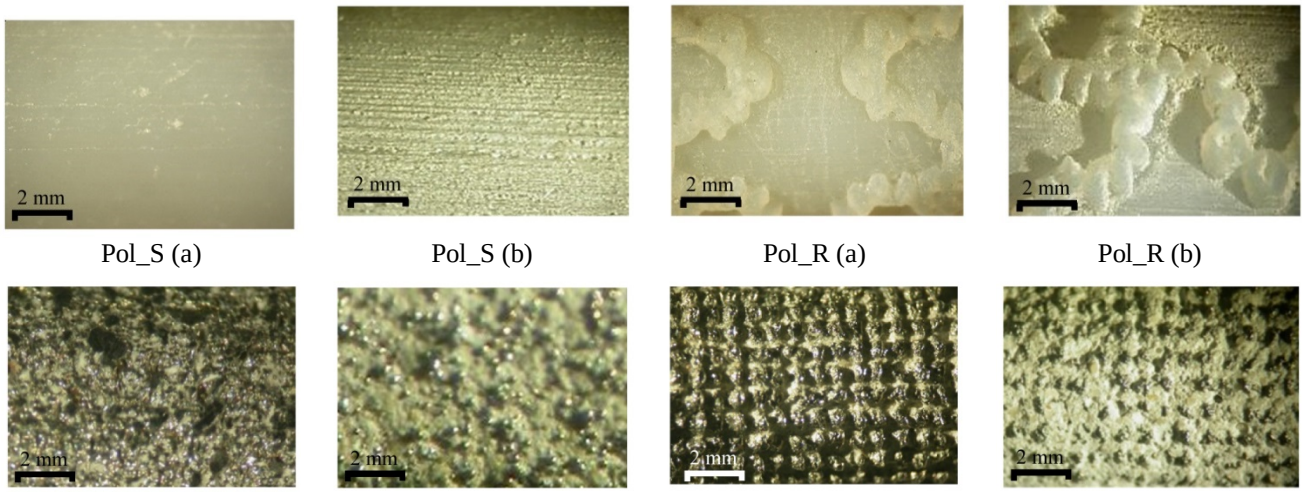


Figure 12: Optical microscope images (1500× magnification).of the examined fibers before test (a) and after pull-out (b).

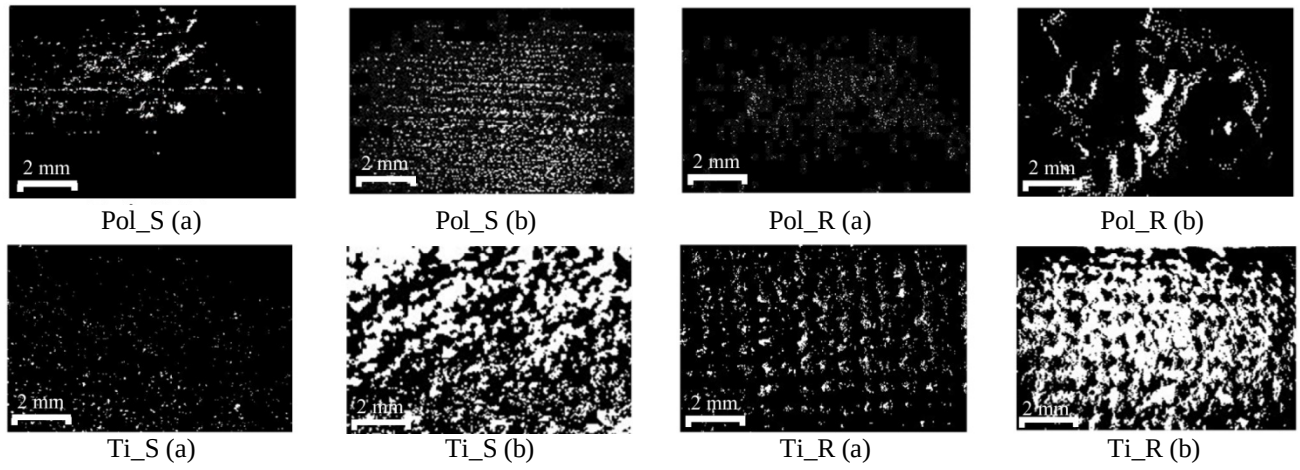


Figure 13: Grayscale images of the examined fibers before test (a) and after pull-out (b).

The images provided in Figure 12-Figure 13 show the presence of mortar particles attached to the pulled-out fibers. Large size mortar particles (with maximum features up to ~1.0 mm) are

visible on the surface of Ti_R and Pol_R fibers, while finer scale particles are visible on the surface of the Ti_S fiber. Small size mortar particles are also visible on the surface of the Pol_S fiber after pull-out. The presence of large size mortar particles attached to the pulled-out fibers is an indicator of high roughness of the fracture surface at the fiber-matrix interface. Such a phenomenon proves that a layer of matrix covers the fiber during the pull-out, causing enhanced interfacial bonding. The development of a rough fracture surface within such a layer, and the crack-deflection mechanisms observed in Figure 9-Figure 11 indicate improvement of matrix toughness in presence of shear failure. Overall, we observe significant increases in the interlaminar shear strength of mortars reinforced with Pol_R, Ti_S and Ti_R fibers, as compared to the unreinforced material.

5 CONCLUDING REMARKS

We have investigated the use of 3D printed fibers for the flexural reinforcement of a cement mortar, employing polymeric (Fullcure 720 Stratasys® photopolymer) and metallic (Ti6Al4V titanium alloy) fibers manufactured with different technologies and surface morphologies.

Three point bending tests on fiber-reinforced mortar specimens have highlighted crack patterns indicating shear failure in the case of reinforcements with high surface roughness fibers (Pol_R, Ti_S and Ti_R fibers), and flexural failure in the case of reinforcements with smooth (Pol_S) fibers, as well as in the case of unreinforced specimens (Sect. 3). The results of Sect. 4 show that the fiber surface geometry strongly influences interfacial bonding, while the chemical nature of the fiber and mortar materials markedly affects the shear capacity and the flexural strength of the fiber-reinforced material (cf. Sect. 3.2).

All the results presented in this study indicate that an optimized fiber surface design may significantly increase the energy absorption capacity of fiber-reinforced mortars. They represent a first step in the direction of designing reinforcing elements with hierarchical structure to form fabrics, fibers and coatings of groundbreaking reinforcements for next generation composites, profiting from the rapid prototyping capabilities of AM technologies at different scales.

Future extensions of the present study (see, e.g., [28]) will focus on nano-, micro- and macro-scale lattice reinforcements for a wide range composite materials, through a closed-loop approach including the computational design and the additive manufacturing of physical models via innovative, multimaterial deposition techniques. Such materials will be used, e.g., for the fabrication of innovative materials, and the structural retrofitting of existing buildings and historical constructions [29][30][31][32].

An experimental characterization phase will implement and verify the theoretical predictions. The design phase will employ multiscale approaches to the toughness and strength of composite materials reinforced with multiscale fibers and/or fabrics, as a function of the microstructure [33][34]. Mechanical models of the composite material will be formulated at the meso-scale, through discrete and/or finite element approaches, assuming different constitutive equations for the bulk phases and the reinforcement-matrix interface (via variational fracture [34][35]). The optimal topologies of the reinforcing elements will be researched using structural optimization procedures [36][37][38], on employing minimum weight, maximum composite strength, and/or maximum fracture toughness as goal functions.

REFERENCES

[1] M.H. Lu, L. Feng, Y.F. Chen, Phononic crystals and acoustic metamaterials. *Materials. Today*. 12, 34-42, 2009.

[2] M. Kadic, T. Bückmann, N. Stenger, M. Thiel, M. Wegener, On the practicability of pentamode mechanical metamaterials. *Applied Physics Letters,*100, 191-901, 2012.

[3] M. Schittny, T. Bückmann, M. Kadic, M. Wegener, Elastic measurements on macroscopic three-dimensional pentamode metamaterials. *Applied Physics Letters,*103, art. no. 231905, 2013.

[4] L.R. Meza, S. Das, J.R. Greer. Strong, lightweight, and recoverable three-dimensional ceramic nanolattices, *Science*, 345, 1322-26, 2014.

[5] Zheng X et al. Design and optimization of a light-emitting diode projection micro-stereolithography three-dimensional manufacturing system. *Review of Scientific Instruments*, 83, art. no. 125001, 2012.

[6] A. Amendola, E.H. Nava, R. Goodall, I. Todd, R.E. Skelton, F. Fraternali. On the additive manufacturing and testing of tensegrity structures, *Composite Structures*, 131, 66-71, 2015.

[7] J. Li, Z. Wu, C. Huang, L. Li. Multiscale carbon nanotube-woven glass fiber reinforced cyanate ester/epoxy composites for enhanced mechanical and thermal properties. *Composite Science and Technology*, 104, 81-88, 2014.

[8] H. Mei, Q. Bai, T. Ji, H. Li, L. Cheng. Effect of carbon nanotubes electrophoretically-deposited on reinforcing carbon fibers on the strength and toughness of C/SiC composites. *Composite Science and Technology*. 103, 94-99, 2014.

[9] N.A. Fleck, V.S. Deshpande, M.F. Ashby, Micro-architectured materials: past, present and future, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.*,466, 2495–516, 2010.

[10] F. Fraternali, F. Fabbrocino, I. Farina. *Multiscale Structural Element ("Elemento Strutturale a Geometria Multiscala"*) Italian Patent Application No.102015000044896, (http://www.fernandofraternaliresearch.com/publications/P1115IT00_Brevetto_Elemento_Multiscala_2015.pdf, accessed on Oct. 28, 2015).

[11] EN 998-2, 2003. *Specification for mortar for masonry - Part 2: Masonry mortar*.

[12] J.R. Raney, F. Fraternali, A. Amendola, C. Daraio. Modeling and in situ identification of material parameters for layered structures based on carbon nanotube arrays. *Composite Structures*, 93, 3013–18, 2011.

[13] ASTM C1018. *American Society for Testing and Materials. Standard test method for flexural strength of fiber reinforced concrete*. Book of ASTM Standards. Barr Harbor Drive (West Conshohocken, PA, United States): ASTM; 1998.

[14] F. Fraternali, I. Farina, C. Polzone, E. Pagliuca, L. Feo, On the use of R-PET strips for the reinforcement of cement mortars. *Composites. Part B, Engineering*, 46, 207-10, 2013.

[15] S. Spadea, I. Farina, A. Carrafiello, F. Fraternali, Recycled nylon fibers as cement mortar reinforcement. *Construction and Building Materials*, 80, 200-9, 2015.

[16] F. Fraternali, V. Ciancia, R. Chechile, G. Rizzano, L. Feo, L. Incarnato. Experimental Study of the Thermo-Mechanical Properties of Recycled PET Fiber Reinforced Concrete. *Composite Structures*, 93, 2368–74, 2011.

[17] F. Fraternali, S. Spadea, V.P. Berardi, Effects of recycled PET fibers on the mechanical properties and seawater curing of Portland cement-based concretes. *Construction and Building Materials*, 61, 293-302, 2014.

[18] E. Cuenca, J. Echegaray-Oviedo, . Serna P Influence of concrete matrix and type of fiber on the shear behavior of self-compacting fiber reinforced concrete beams. *Composites. Part B, Engineerin*g,,75, 135-47, 2015.

[19] H. Kim, G. Kim, N. Gucunski, J. Nam, J. Jeon, Assessment of flexural toughness and impact resistance of bundle-type polyamide fiber-reinforced concrete. *Composites. Part B, Engineerin*g,78, 431-46, 2015.

[20] A. Martins, G. Vasconcelos, R. Fangueiro, F. Cunha, Experimental assessment of an innovative strengthening material for brick masonry infills. *Composites. Part B, Engineerin*g, 80, 328-42, 2015.

[21] E. Hernandez-Nava, C.J. Smith, F. Derguti, S. Tammas-Williams, F. Leonard, P.J. Withers, I. Todd, R. Goodall, The effect of density and feature size on mechanical properties of isostructural metallic foams produced by additive manufacturing, Acta Materialia, 85, 387–95, 2015.

[22] G.A. Rombach, M. Kohl, V.H. Nghiep, Shear Design of Concrete Members without Shear Reinforcement - A Solved Problem? *Procedia Engineering*, 14, 134–40, 2011.

[23] T. Ikegawa, H. Saito, H. Ohuchi, H. Kitoh, H. Tsunokake, Flexural and Shear Failure Tests of Reinforced Concrete Beams with Low Grade Recycled Aggregate. *Memoirs of the Faculty of Engineering*, Osaka City University, 50, 29-36, 2009.

[24] ACI Task Committee 318. *Building code requirements for structural concrete (ACI 318-08) and commentary*. American Concrete Institute. Farmington Hills, 2008.

[25] BS8110-1. *Structural use of concrete, part 1: Code of practice for design and construction*. British Standards Institute. London, 1997.

[26] EN 1992-1-1:2004. *Eurocode 2. Design of concrete structures - Part 1-1: General rules and rules for buildings.*

[27] EN 1996-1-1:2005. *Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures.*

[28] I. Farina, F. Fabbrocino, G. Carpentieri, M. Modano, A. Amendola, R. Goodall, L. Feo, F. Fraternali, On the reinforcement of cement mortars through 3D printed polymeric and metallic fibers. *Composites. Part B, Engineerin*g, 90, 76-85, 2016.

[29] F. Fraternali, G. Carpentieri, M. Modano, F. Fabbrocino, R.E. Skelton, A tensegrity approach to the optimal reinforcement of masonry domes and vaults through fiber-reinforced composite material. *Composite Structures*, 134, 247-54, 2015.

[30] F. Fabbrocino, I. Farina, V.P. Berardi, A.J.M. Ferreira, F. Fraternali, On the thrust surface of unreinforced and FRP-/FRCM-reinforced masonry domes. *Composites. Part B, Engineerin*g, 83, 297-305, 2015.

[31] F. Fraternali, G. Carpentieri, A. Amendola, R.E. Skelton, V.F. Nesterenko, Multiscale tunability of solitary wave dynamics in tensegrity metamaterials, *Applied Physics Letters*, 105 (20), art. no. 201903, 2014.

[32] A. Bossio, F. Fabbrocino, G.P. Lignola, A. Prota, G. Manfredi, Simplified model for strengthening design of beam-column internal joints in Reinforced Concrete Frames, *Polymers*, 7 (9), 1732-1754, E-ISSN: 2073-4360, DOI: 10.3390/polym7091479, 2015.

[33] Y. Li, M. Zhou, Prediction of fracture toughness of ceramic composites as function of microstructure: I. Numerical simulations, *Journal of the Mechanics and Physics of Solids*, 61(2), 472-88, 2013.

[34] F. Fraternali, Free Discontinuity Finite Element Models in Two-Dimensions for In-Plane Crack Problems, *Theoretical and Applied Fracture Mechanics*, 47, 274-82, 2007.

[35] B. Schmidt, F. Fraternali, M, Ortiz, Eigenfracture: An eigendeformation approach to variational fracture. *Multiscale Modeling and Simulation*, 7(3), 1237-66, 2009.

[36] F. Fraternali, A. Marino, T. El Sayed, A. Della Cioppa, On the structural shape optimization through variational methods and evolutionary algorithms. *Mechanics of Advanced Materials and Structures*, 18, 224-43, 2011.

[37] R.E. Skelton, F. Fraternali, G. Carpentieri, A. Micheletti, Minimum mass design of tensegrity bridges with parametric architecture and multiscale complexity. *Mechanics Research Communications*,58, 124-32, 2014.

[38] G. Carpentieri, M. Modano, F. Fabbrocino, F. Fraternali, Optimal design and dynamics of truss bridges, *COMPDYN 2015 - 5th ECCOMAS* Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, 1731-1740, 2015

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A TEMPERATURE SENSITIVE CRYSTAL PLASTICITY MODEL FOR THE PREDICTION OF HIGH TEMPERATURE MECHANICAL BEHAVIOUR OF MULTI-PHASE TIAL ALLOY

M. Umer Ilyas[1], **M. Rizviul Kabir**[1]

[1] Institute of Materials Research, German Aerospace Centre (DLR)
Linder Höhe, D-51147 Cologne, Germany
e-mail: umer.ilyas@dlr.de, mohammad-rizviul.kabir@dlr.de

Keywords: TiAl Alloy, Crystal Plasticity Modelling, High Temperature Behaviour, Computational Methods.

Abstract. *Intermetallic TiAl alloys for aero-engine turbine blade applications show good thermomechanical properties and excellent creep resistance. These alloys consist of multi-phase intermetallic constituents, mainly of γ-TiAl and α_2-Ti_3Al intermetallic phases, which are arranged in the form of lamellar colonies and globular grains. Characterizing these alloys for room and high temperature behaviour is an ongoing issue that needs to be solved by using a synergic approach of experimental and numerical methods.*

In the recent progress of numerical approaches for the analysis of TiAl alloys, physically motivated crystal plasticity finite element models (CPFEM) have been successfully used to describe room temperature mechanical behaviour and to predict microstructure-property correlations. Unfortunately, not much progress in numerical modelling that is able to characterize this alloy with respect to high temperature mechanical behaviour has been reported. To overcome this lacking, in the present work we propose a temperature sensitive CPFEM model, where the temperature-dependent slip rates of the crystallographic deformation modes are described as a function of history variable and physical material parameters. This model has been verified and validated for multi-phase TiAl lamellar microstructure using the experimental results of fully lamellar PST-TiAl single crystal as published in the literature. Further, for the computational analysis, unit cell based FE models are constructed with representative lamellar microstructure consisting of γ-TiAl and α_2-Ti_3Al lamellar plates. Constitutive behaviour of the phases is described by the proposed crystal plasticity model. A multiscale local-global FE approach has been used to obtain macro-scale homogenized mechanical behaviour of the local microstructure. In this presentation, we will show that our model is able to predict high temperature deformation behaviour of TiAl alloy as observed in the PST-TiAl experiments quite satisfactory. The model successfully captures the temperature sensitive yield behaviour, anisotropic response related to morphological and crystallographic orientation, and rate sensitivity of the slip deformation. Further, we will demonstrate that the activity of crystallographic slip systems can explain the local plasticity in both room and high temperature.

1 INTRODUCTION

Multi-phase TiAl alloys are used in low-pressure turbine blades of aircraft engines where the operating temperature does not exceed 700°C. Recently, considerable amount of research activity is involved in developing new generation of TiAl alloys that may show enhanced thermomechanical properties beyond 700°C temperature. Today, understanding the processing-microstructure-property correlations has become vital to the development of new material with enhanced properties.

For TiAl alloys, the microstructure-property correlations have been explored so far using detailed experiments and microstructure analysis. The experimental results and additional microstructural analysis give somewhat a qualitative picture how the microstructure and properties are correlated. However, the underlying mechanisms of deformation influenced by the local microstructure and available intermetallic phases that influence the macroscopic behaviour remain mostly unknown. Today's material design concepts are based on detail understanding of microstructure-property relationships in a quantitative manner. This requires physics based understanding of the materials deformation mechanisms, which can be obtained by advanced experimental techniques (FIB, nanoindentation) as well as can be explored by advanced numerical modelling techniques incorporating physically motivated materials constitutive behaviour.

To this goal, in past years only limited computational analysis approaches have been proposed for the investigation of TiAl alloy's deformation behaviour. The available models are based on materials micromechanics and micro-macro relationship where the crystal plasticity constitutive models have become a key numerical tool to explore the materials deformation behaviour considering physical microstructural quantities. Only room temperature deformation behaviour was focused. The deformation behaviour at high temperature, which is more relevant for TiAl alloys, has not been systematically investigated so far. Literary study also reveals a lack of proper crystal plasticity models to predict the high temperature deformation behaviour of TiAl alloys.

To overcome this lacking, we propose an extension of crystal plasticity constitutive model applicable to high temperature deformation behaviour. This was done by incorporating the rate sensitivity of the slip deformation as a function of temperature and physical quantities of TiAl alloy, such as, burgers vector, shear modulus of the crystal systems.

In the present work, we have conducted an exploratory numerical investigation to assess the proposed temperature sensitive crystal plasticity model for the room and high temperature behaviour of TiAl alloy. This model will be validated using a single crystal TiAl alloy consisting of multi-phase lamellar microstructure. This type of multi-phase single crystal is known as PST-crystal. For model validation, use of PST-TiAl alloy is ideal, as the fundamental micromechanisms of deformation for various lamellar configurations under various loading histories have been explored through extensive experimental work in past decades.

2 MATERIAL AND METHODS

2.1 Microstructure

The lamellar microstructure of TiAl alloy consists of two intermetallic phases: γ-TiAl with tetragonal crystal structure and α_2-Ti_3Al with hexagonal crystal structure (Figure 1a, b). These phases are arranged parallel to each other in the form of lamellar plates. Between two α_2-lamellar plates many γ-lamellae can be situated. The orientation relationship between these

two phases is as follows: $(0001)_{\alpha 2}$ and $\{111\}_{\gamma}$ planes and the closely packed directions, $\langle 11\bar{2}0\rangle_{\alpha 2}$ and $\langle 1\bar{1}0|_{\gamma}$ are parallel.

The γ-phases contribute to the overall deformation of the macroscopic samples, while, α_2-lamellae is hard to deform. In γ-phases, ordinary, twin, and super dislocations are active which are operative at $\{111\}_{\gamma}$-planes. With respect to available deformation modes of the γ-lamellae, they are divided into γ -matrix and γ -twin lamellae (Figure 1c).

For modelling purpose, the lamellar microstructure is described in a unit cell taking only three representative lamellae phases: α_2, γ-twin and γ-matrix, as shown in Figure 1d. The crystallographic orientation and vol. % of the phases are maintained in the unit cell. In FE model, each phase is defined by 3D linear element. Setting the unit cell periodically side by side, a fully lamellar microstructure is constructed.

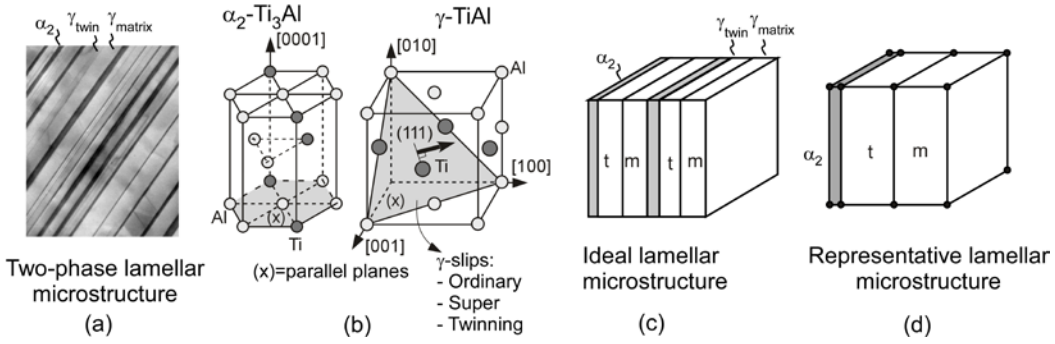


Figure 1: a) Two-phase lamellar microstructure of a PST-TiAl, b) crystal structure of intermetallic TiAl phases, c) simplified lamellar structure with regularly arranged lamellae phases (α_2, γ-matix, γ-twin), d) unit cell of lamellar structure consisting of representative volumes of a γ-matix, a γ-twin and a α_2-lamellae

To understand the mechanical behaviour of the lamellar microstructure, the deformation micro-mechanisms of the intermetallic phases with respect to lamellar orientation have been investigated in detail by many researchers. A clear understanding of these mechanisms has been obtained by analysis of PST-TiAl single crystal (Figure 1a) consisting of various lamellar configurations. Investigating the PST single crystal, very fundamentals of the alloy deformation under room and high temperature behaviour can be obtained. The alloy constituents, microstructural details and the mechanical test data can be used to develop and validate any numerical model and the model parameters influencing the elementary micro-mechanisms can be separately justified, as it will be done in the present work.

2.2 Constitutive behaviour

The constitutive behaviour of the intermetallic TiAl phases is described in a classical crystal plasticity model. In crystal plasticity model the crystallographic slip is taken as the only mechanism for plastic deformation, which occurs due to shearing of slip planes in slip directions. In this model, the shear strain rate $\dot{\gamma}^{\alpha}$ of the slip system α is formulated in terms of resolved shear stress τ^{α} and current strength of the slip systems g^{α} [1, 2]:

$$\dot{\gamma}^{\alpha} = f\left(\tau^{\alpha}, g^{\alpha}\right)$$

$$\dot{\gamma}^{\alpha} = \dot{\gamma}_0 \frac{\tau^{\alpha}}{g^{\alpha}} \left|\frac{\tau^{\alpha}}{g^{\alpha}}\right|^{\frac{1}{m}-1} \tag{1}$$

where, $\dot{\gamma}_0$ is the initial shear strain rate and *m* is the rate sensitivity parameter that controls the rate dependent constitutive behaviour of the material.

The rate sensitivity parameter for the shear strain rate is phenomenological. It generally lies between 0 and 1. Higher values of *m*, i.e. equal to or closer to one, induce viscoplastic response of the material. Lower values results in rate independent deformation of the material.

The evolution equation for the slip system resistance is described by a linear hardening law:

$$\dot{g}^{\alpha} = h_0^{\alpha} \sum_{\beta} q_{\alpha\beta} \dot{\gamma}^{\beta} \tag{2}$$

where, h_0^{α} is the hardening modulus, $q_{\alpha\beta}$ is the slip-plane hardening matrix, which describes the effect of self- and latent hardening. In literature, the values of $q_{\alpha\beta}$ for fcc crystals are in the range of 1.0–1.4. Since $q_{\alpha\beta}$ values for the particular cases of the α_2- and γ-phase have not yet been experimentally estimated, we assume $q_{\alpha\beta}$ = 1.0 for all slip combinations, and do not distinguish between self- and latent hardening.

Twinning is implemented in the UMAT as a unidirectional glide of the partial dislocation and is activated by a set of parameters same as for slip activation.

This classical description of crystal plasticity does not show much sensitivity to strain rate and temperature [3]. As an improvement, Kothari and Anand [3] proposed a physically motivated kinetic equation of shear strain rate on slip systems based on thermally activated dislocation motion [4]. In this approach, the shear strain rate takes the form:

$$\dot{\gamma}_{\alpha} = \dot{\gamma}_0 \exp\left\{ -\frac{\Delta F}{k_B T}\left[1 - \left(\frac{\tau_t^{\alpha}}{g_t^{\alpha}} \right)^{p} \right]^{q} \right\} sign\left(\tau^{\alpha}\right)$$

$$where \quad \tau_t^{\alpha} = \left[1 - \left(\frac{T}{Tc} \right)^{1/q} \right]^{1/p} \quad and \quad Tc = \frac{\Delta F}{k \ln\left(\dot{\gamma}_0 / \dot{\gamma}_0^{\alpha} \right)} \tag{3}$$

Here, k_B is Boltzmann constant, *T* is temperature, $\dot{\gamma}_0$ is the initial strain rate, and $\dot{\gamma}$ is the current strain rate. T_c is the limit temperature beyond which enough thermal energy is present for the barriers to be overcome only by the thermal activation, without the aid of external stresses.

The values of *p* and *q* control the shape of the glide profile. Their values are typically $0 \leq p \leq 1$ and $1 \leq q \leq 2$. ΔF is the activation energy, which is the free energy required to overcome the thermal obstacles during slip deformation without the help of applied stress.

A connection between the physically based formulation of Kothari [3] and the classical power law can be established by defining a physically modified rate sensitivity parameter, *m*, in terms of temperature and activation energy as follows:

$$m = \frac{\partial \ln|\tau|}{\partial \ln|\dot{\gamma}|} = \frac{k_B T}{\Delta F}\frac{1}{pq}\left(\frac{\left(\frac{T}{T_c}\right)^{\frac{1}{q}-1}}{1-\left(\frac{T}{T_c}\right)^{\frac{1}{q}}}\right) \qquad (4)$$

$$where, Tc = \left(\frac{\Delta F}{k_B.\ln\left(\frac{\dot{\gamma}_0}{|\dot{\gamma}|}\right)}\right)$$

The power law equation can be now modified as follows:

$$\dot{\gamma}^{\alpha} = \dot{\gamma}_0 \frac{\tau^{\alpha}}{g^{\alpha}}\left|\frac{\tau^{\alpha}}{g^{\alpha}}\right|^{\frac{1}{m(\dot{\gamma},T)}-1} \qquad (5)$$

Where in classical power law the rate sensitivity parameter is purely phenomenological, in current modification the evolution of shear strain rate depends on materials physical parameter via the rate sensitivity parameter *m*, which is now sensitive to slip strain rate and temperature.

The ΔF should be representative to the investigated material. The activation energy can be written in terms of shear modulus μ of the material and its burgers vector *b* as follows [3]:

$$\Delta F = A\mu b^3 \qquad (6)$$

Depending on the strength of the obstacle, the pre-factor *A* varies from 0.05 to 2.0. Frost and Ashby [5] suggested the range of *A* depending upon the strength of the material obstacles as shown in Table 1. In other words, this pre-factor describes the strength of the material obstacles.

Obstacle Strength	*A*	Description
Strong	2.0	Strong Precipitates, Dispersions
Medium	0.2 - 1.0	Weak Precipitates, Forest Dislocations
Weak	< 0.2	Lattice Resistance, Solution Hardening

Table 1: Characteristics of obstacles

In γ-TiAl alloys, the dislocations glide on {111} planes with the Burgers vectors, 0.283nm, 0.566nm, 0.4905nm [6]. For current calculations, the Burger's vector of 0.283nm is used. The shear modulus, μ for γ-TiAl alloys varies with temperature *T* as follows [6]:

$$\mu = 70.30\text{-}0.0141(T) \qquad (7)$$

For glide profile, the values of *p* and *q* can be taken as 1 for simplicity which yields the *m* equation to the following:

$$m = \frac{k_B T}{\left(A\mu b^3 - k_B T . \ln\left(\frac{\dot{\gamma}_0}{|\dot{\gamma}|} \right) \right)} \tag{8}$$

This modified power law form of viscoplastic formulation taking a physically motivated *m* parameter for temperature sensitivity has not been pursued in any published work. One advantage of using this model is that the framework of the crystal plasticity model need not be changed rigorously but can be easily implemented in the existing code.

This model has been verified for room and high temperature behaviour of PST-lamellar alloy.

2.3 Micro-macro two-scale Model

While the constitutive behaviour of the phases is obtained by the temperature sensitive CPFEM model, the overall mechanical behaviour is calculated from the numerical averaging of stresses and strains over the whole unit cell consisting of intermetallic phases. To obtain the macroscopic behaviour, for example, load vs displacement or stress vs strain curves, as a response to the local deformation micromechanics of the microstructure; a two-scale micro-macro approach is implemented. In this two-scale approach, deformation localization is done via transforming nodal deformation gradient from the macroscopic boundary to the micro-scale unit cell boundary (known as FE^2 approach for micro-macro coupling). After getting stresses on each intermetallic phase due to boundary loading, a volumetric average is performed to obtain overall stress values. At the boundary nodes a reaction force can be derived, which is used to re calculate the macroscopic load-displacement curve (converted to stress-strain curve). The approach is schematically shown in Figure 2, and a detail of this model can be found elsewhere [7,8].

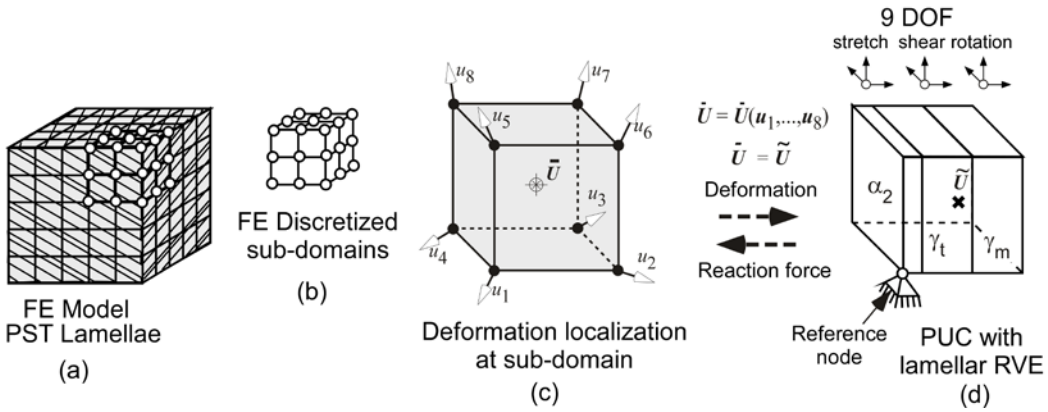


Figure 2: Micro-macro approach for homogenized multi-phase behaviour (a) discretized FE model of a PST-TiAl, (b) a macro-scale subdomain where a micro-scale unit cell is defined, (c) deformation localization at the subdomain, (d) localized deformation is transferred to the unit cell to act as boundary conditions.

In the present case, the microstructure details, e.g. presence of phases, vol. % of the crystallographic orientations or texture, are defined in the unit cell. Further, correlation of the macroscopic sample and the oriented lamellar microstructure is defined by a rotation matrix.

3 SIMULATION AND RESULTS

3.1 Parameter estimation and deformation behaviour

The fundamental parameter set has been adjusted for experimentally obtained room temperature behaviour for three basic orientations of the PST lamellar, e.g. 0°, 45° and 90°. For the numerical stress-strain response, the crystal plasticity model as proposed in section 2.3 is used. The microstructural descriptions of the multi-phase alloys with their anisotropy are defined by a unit cell as described in section 2.1 and 2.3.

To estimate the crystallographic strength and hardening parameters of the slip systems, a morphological description of all the slip systems as defined by Lebensohn et al. [9] is used as listed in Table 2. This description allows one to estimate crystallographic viscoplastic parameters with only a minimum set of experiments performed on 0°, 45° and 90° orientated lamellar PST-TiAl alloy. Details of the parameter estimation are given elsewhere [10]. In Table 2, the strength of the slip systems at a particular deformation mode is denoted as g_0, and Q_{so} is a strengthening factor for super dislocations.

Slip type	Lamellar deformation modes		
	Longitudinal	Mixed	Transverse
Ordinary	g_0^{long}	g_0^{mixed}	g_0^{trans}
Super	$g_0^{long}.Q_{SO}$	$g_0^{mixed}.Q_{SO}$	$g_0^{trans}.Q_{SO}$
Twin	g_0^{long}	g_0^{mixed}	g_0^{trans}

Table 2: Morphological classification of slip systems with respect to the lamellar deformation modes

According to the experimental evidence of slip system activity and stress analysis of Kishida et al. [11], we assume that for all slip systems the ordinary and twinning stresses are the same, super dislocation strength is higher. Now a series of numerical simulations has been performed for fitting the experimental stress-strain behaviour of 0°, 45° and 90° oriented lamellae. The experimental curves are taken from the work of Uhlenhut [12].

In Figure 3a the stress-strain behaviour obtained via simulation fitting is shown along with the experimental curves. Both yield and hardening behaviour have been captured very well with the proposed CPFEM model. Even the anisotropic yield for 0°, 45° and 90° lamellar orientation is satisfied very well, see Figure 3b.

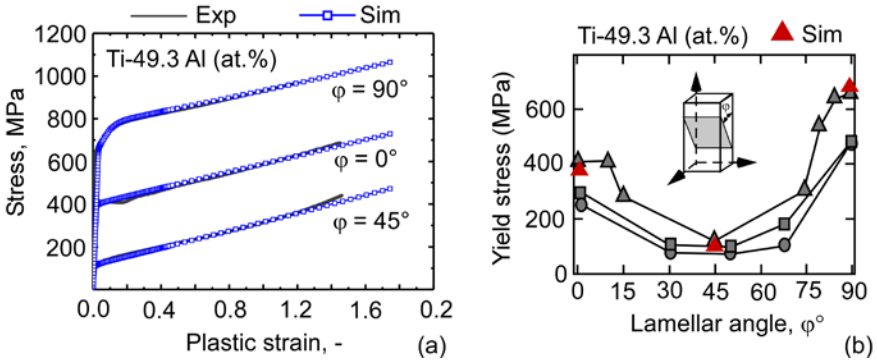


Figure 3: (a) Room temperature compression stress-strain curves for oriented PST crystal, (b) Yield anisotropy of PST-TiAl lamellar alloy

The crystal strengths, g_0 are obtained from these fitted curves, which are listed in Table 3. These parameters will be taken as the fundamental crystal plasticity parameters for TiAl lamellar structure and will be used to predict high temperature behaviour.

Slip type	Lamellar deformation modes		
	Longitudinal	Mixed	Transverse
Ordinary	75	235	260
Super	75	235	260
Twin	75	--	260

Table 3: Strength parameters for the crystal plasticity model.

3.2 Anisotropic yield at high temperature

To examine the temperature sensitivity of the proposed CPFEM model, the load-deformation behaviour is simulated for 27, 200, 400, 600, 800 and 1000°C. For the PST-TiAl, Inui et al. [13] have performed a series of experiments for high temperature behaviour with different oriented lamellar microstructure. He uses 31° lamellar orientation for 'soft deformation' of lamellae and 0° and 90° for 'hard deformation'. For numerical modelling we have used 45° lamellar orientation for 'soft deformation' as the model parameters for this lamellar arrangement has already been verified for room temperature. For hard deformation 0° and 90° lamellar orientation is considered as before.

In Figure 4, the simulated results for yield anisotropy depending on temperature and lamellar orientation have been plotted together with the experimental results from Inui et al. [12]. A very good prediction of yield anisotropy with respect to the lamellar orientation and temperature has been obtained. An overall reduction of yield stress with the increase of temperature has been seen for both experimental and numerical results. At low temperatures, the yield stress decreases quickly with the increase in temperature. At intermediate temperature range, the decrease is gradual for all orientations. At high temperature range, the yield stress decreases rapidly with increases in temperature.

For 0° oriented lamellae, an anomaly has been identified at 800°C. At this temperature range an increase in yield stress is observed, which was however not trivial for 45° and 90° orientation. This anomalous behaviour of high yield stress is associated with the lamellar boundaries of PST TiAl, which impose additional obstacles during dislocation movements crossing the lamellar boundaries [13].

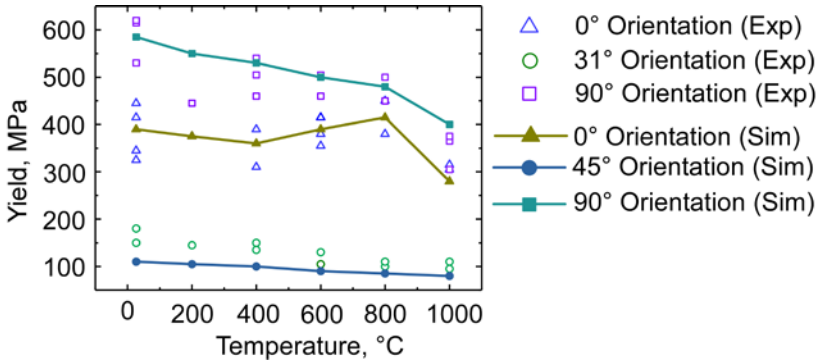


Figure 4: (a) Temperature dependency of yield behaviour for oriented lamellar grains, (b) Temperature dependency of the CRSS ratio of ordinary and super dislocation (Qso) as predicted from numerical analysis

3.3 Activity of slips at room and high temperature

The CPFEM model also allows us to study the activation of different slip modes, which are important for deformation behaviour in individual phases at different temperatures and differently oriented lamellae configurations. Following, we have calculated the relative activity of the slips at room (27°C) and 800°C temperature, and for 0°, 45°, and 90° oriented lamellae. The activity of slips has been calculated at the onset of yield. The predictions are shown in Figure 5.

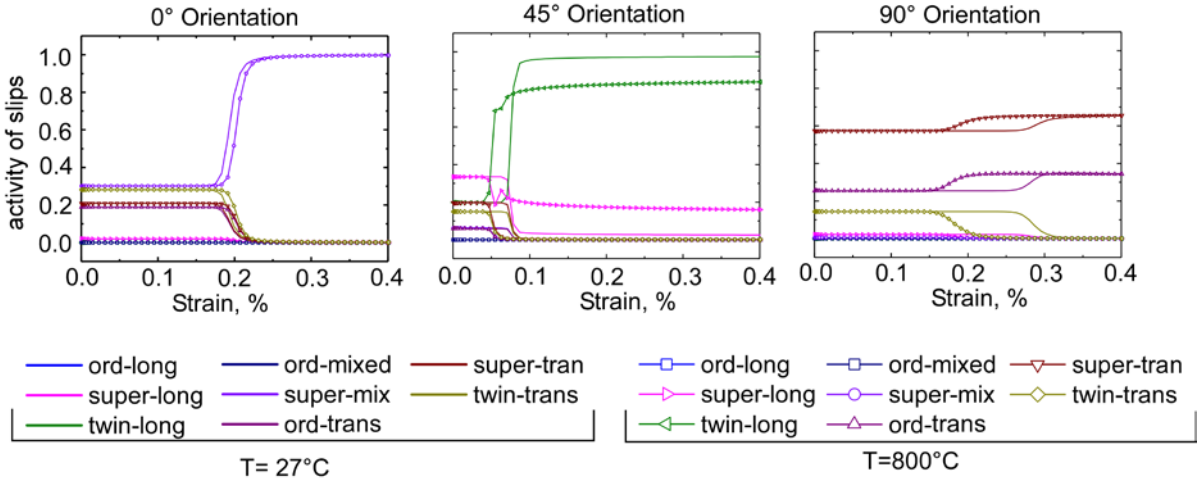


Figure 5: Activity of slips for 0°, 45° and 90° oriented lamellae at room temperature and 800°C

For 0° lamellar orientation, we observe that the super dislocations are dominant for plastic deformation at both room (27°C) and high temperature (800°C). For 45° lamellar orientation, the deformation occurs mainly due to twinning at room temperature. But at high temperature some super dislocations are also active in addition to the twinning. For the 90° lamellar orientation, both super and ordinary dislocations are dominant for slip deformation. For 45° and 90° lamellar orientation, at high temperature these slip modes are activated earlier compared to room temperature, which allows an early softening of the material. For 0° lamellar orientation, the super dislocation slips are activated later at 800°C than at room temperature, which is responsible for the anomalous behaviour of 0° oriented lamellae. The results seem consistent with the experimental observations of Kishida [11] for room temperature and Inui [13] for high temperature behaviour.

4 SUMMARY AND CONCLUSION

The crystal plasticity model proposed in this work incorporates the rate sensitivity of individual slips systems by a temperature dependent function proposed by Kothari and Anand [3]. In classical analysis of high temperature behaviour using crystal plasticity model the temperature sensitivity is obtained only by using a phenomenological parameter *m*, which was insensitive to slip rates and temperature sensitive. The main objective in this work was to show an alternative formulation of power law equation of crystal plasticity to predict the temperature sensitivity, which is more physics based.

This model has been validated using available data from the literature. We have validated the model for room temperature deformation behaviour capturing anisotropy due to oriented lamellar configurations. Further, high temperature yield anisotropy as well as temperature dependent yield behaviour has been predicted, which matches the experiments very well. The

model also explains the high temperature deformation behaviour with respect to the activity of slips systems for different oriented lamellar configuration.

Further, we have determined slip strength parameters for PST-TiAl alloy from the experimental fitting of room temperature deformation behaviour. Due to temperature sensitive rate equation in the crystal plasticity model, the strength parameters become temperature sensitive. Using this temperature sensitive strength parameter set the model is able to predict the high temperature behaviour of PST-TiAl as determined experimentally in [13].

The results presented in this work clearly state that this new formulation of CPFEM can be successfully used to predict room and high temperature deformation behaviour of fully lamellar PST-TiAl alloy. This model also shows huge potentials to understand the microstructure property correlations of the intermetallic systems. In future, further analysis will be conducted to validate this model for a wide range of TiAl alloy microstructure subjected to different temperatures.

REFERENCES

[1] J.W. Hutchinson, Bounds and self-consistent estimates for creep of polycrystalline materials. *Proceedings of the Royal Society of London*, A 348, 101–127, 1976.

[2] R.J.Asaro, A. Needleman, Texture development and strain hardening in rate dependent polycrystals. *Acta Metallurgica 33*, 923–953, 1985.

[3] M. Kothari, L. Anand, Elasto-viscoplastic constitutive equations for polycrystalline metals: application to Tantalum. *Journal of the Mechanics and Physics of Solids*, 46, 51-83, 1998.

[4] U.F. Kocks, A.S Argon, M.F Ashby, *Thermodynamics and Kinetics of Slip, Progress in Material Science*. Pergamon Press, 1975.

[5] M.F. Ashby, H.J. Frost, *Deformation Mechanism Maps-The Plasticity and Creep of Metals and Ceramics*. Pergamon Press, Oxford, UK, 1982.

[6] F. Appel, J.D.H. Paul, M. Oehring, *Gamma Titanium Aluminide Alloys: Science and Technology*. Wiley VCH, Weinheim, 2011.

[7] M. Werwer, A. Cornec. Numerical simulation of plastic deformation and fracture in polysynthetically twinned (PST) crystals of TiAl. *Computational Materials Science* 19 97-107, 2000.

[8] M.R. Kabir, L. Chernova, M. Bartsch, Numerical investigation of room-temperature deformation behaviour of a duplex type γTiAl alloy using a multi-scale modelling approach. *Acta Materialia*, 58, 5834-5847, 2010.

[9] R. Lebensohn, H. Uhlenhut, C. Hartig, H. Mecking, Plastic flow of γ-TiAl-based polysynthetically twinned crystals: micromechanical modeling and experimental validation. *Acta Materialia,* 46,4701-4709, 1998.

[10] A.Cornec, M.R. Kabir, N. Huber, Numerical prediction of the stress–strain response of a lamellar γTiAl polycrystal using a two-scale modelling approach. *Materials Science & Engineering* A, 620, 273-285, 2014.

[11] K. Kishida, H. Inui, M. Yamaguchi, Deformation of lamellar structure in TiAl-Ti_3Al two-phase alloys. *Philosophical Magazine A*, 78, 1-28, 1998.

[12] H. Uhlenhut, *Ph.D. Thesis, Ursachen plastischer Anisotropie von γ-TiAl-Basislegierungen*. Technical University of Hamburg-Harburg, Shaker Verlag, 1999.

[13] H. Inui , K. Kishida , M. Misaki , M. Kobayashi , Y. Shirai & M. Yamaguchi Temperature dependence of yield stress, tensile elongation and deformation structures in polysynthetically twinned crystals of Ti-Al. *Philosophical Magazine A*, 72, 1609-1631, 1995.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

LAYOUT OPTIMIZATION OF PIEZOELECTRIC ELEMENTS WITH EXTERNAL ELECTRIC CIRCUITS IN SMART CONSTRUCTIONS BASED ON SOLUTION OF THE NATURAL VIBRATIONS PROBLEM

Nataliya A. Iurlova [1], Valerii P. Matveenko[2], Dmitriy A. Oshmarin[2], Nataliya V. Sevodina[2], Maksim A. Yurlov [2]

[1] Institute of Continuous Media Mechanics of the Ural Branch of RAS
Acad. Korolev str., 1, Perm, 613013 Russia
e-mail: yurlova@icmm.ru

[2] Institute of Continuous Media Mechanics of the Ural Branch of RAS
Acad. Korolev str., 1, Perm, 613013 Russia
e-mail: mvp@icmm.ru, oshmarin@icmm.ru, natsev@icmm.ru, yurlovm@icmm.ru

Keywords: Electroviscoelesticity, Piezoelectric element, Natural vibrations, Optimal placement, Electric potential

Abstract. *The presence of the elements, which work as sensors or actuators, is one of the main features of smart-constructions. These elements are commonly made of piezoelectric materials, the main feature of which is a direct and inverse piezoelectric effect, which allows its functioning as a sensor and actuator. In addition, one and the same element can combine both functions. Moreover electroding of surfaces of the piezoelectric elements provides an additional means of controlling the dynamic processes due to application of different electric circuits connected to piezoelectric elements. The location of the piezoelectric element, its size and geometry play a significant role in damping the vibration of structures containing elements made of piezoelectric materials. In this paper, a comparative analysis of different possible options of smart-function realization is made by comparing the patterns of strains and electric potentials generated on the piezoelectric element surface. The numerical results of determining the optimal piezoelectric element location, which will provide the generation of maximal electric potential at certain resonant frequencies, are presented. The numerical implementation is carried out by the finite element method using the commercial ANSYS software package.*

1 INTRODUCTION

In the last few years the method of damping the structure vibrations based on converting the energy of mechanical vibrations into the electrical energy via piezoelectric elements embedded in the structure and connected to the external electric circuits has been the area of extensive research in mechanics. The piezoeletric effect provides a transformation of part of the mechanical vibration energy into the electrical energy, which is then dissipated through the external electric shunt circuit operating as a mechanism of passive vibration damping [1]. Changing the parameters of the external circuit one can suppress the vibrations at the specified frequency.

Piezoelectric elements should provide the best execution of the ultimate purpose of their application in smart-structures, that is, damping of structure vibrations, controlling the shape of structures subject to external impacts of different nature, detecting the structural defects, etc. The efficiency of peizoelement performance depends not only on the control parameters (the corresponding schemes of external electric circuit connection) but also on the arrangement of piezoelectric elements serving as sensors and actuators [2].

Depending on the criterion, the optimal schemes for location of sensors (actuators) can differ from one another. The optimization criterion can use as a basis different factors: maximization of forces, moments or deformations transferred by a peizoelement playing the role of the actuator, slight variation of dynamic characteristics of the original structure, selection of desired dynamic characteristics for the original structure, minimization of control function, maximal damping of original structure vibrations, etc.

Traditionally, the passive vibration dampers are located in the regions of the structure, in which the strain energy is maximal [3]. As a rule, these zones are found based on the analysis of structure deformation patterns. Nevertheless, the problem of optimal arrangement of piezoelectric elements is in the center of attention of many researchers, whose studies in this area are devoted to the development of both the optimization algorithms and criteria for optimal arrangement of piezoelectric elements.

It should be noted that thus far no universal approach to the examined problem has been developed due to the necessity of taking into account a variety of factors: characteristics and geometry of piezoelements, their arrangement in the structure, the number of piezoelements, etc. In work [4], which is a survey of more than 100 papers and in work [5], which makes reference to about 50 papers devoted to the problem of piezoelement arrangement in structures, the authors advance an opinion that despite the abundance of methods and approaches to this problem, the appropriate location of a sensitive element has long been a matter of skill that relies on researcher's intuition. Such an approach contradicts the intention of design engineers to reduce the number of intuitive solutions. In view of this fact, work [4] presents and systemizes all previously developed criteria for optimal arrangement of peizoelements in from the viewpoint of minimization of intuitive approaches to this problem.

As mentioned in [4], an advanced approach to the problem of appropriate arrangement of sensors (actuators) in the structure is based on the six optimization criteria: maximization of modal forces or moment transferred through the actuators; maximization of strains in the original structure; minimization of control forces or maximization of dissipated energy; maximization of the degree of controllability; maximization of the degree of observability and minimization of side effects. The most general recommendations on the optimal location obtained on the basis of these criteria are systemized in the review paper [4] for beam and plate smart-structures.

The application of voltage to the piezoelectric element placed on the structure surface causes transverse deformations (deflection) of the structure. The transverse deformation of the initial structure is thus the function of the sensor (actuator) position. Hence, as it was shown in works [6-8], the transverse deformation of the initial structure can be used as a criterion for selection of the optimal location of sensors or actuators.

In work [9] the authors offer the strategy, which is based on the mathematical analysis of the deformation fields under structure vibrations and pursues the aim of finding the regions of maximal deformation which are equal in size to a piezoelement. This strategy was applied to a structure of complex geometry made of anisotropic material with locally varying properties and a small value of the internal material damping ratio.

In what follows, we restrict the discussion to the problem of piezoelement location only as a means of damping the vibration in the original structure.

The distribution of damping energy over the volume of the examined structure gives a rather clear idea of the best location for a passive damping element. However, it should be noted that the modal analysis of the electroviscoelastic bodies composed of structural materials with the embedded piezoelectric element allows us to obtain information on the relative electric potential generated on the electroded surfaces of piezoelements, which can be also used for identification of optimal variants of piezoelement location.

In works [10-12] it has been noted that the location of piezoelements operating as sensors in the zones, which are unsuitable for generation of considerable electric potential, can even impair the rate of vibration damping.

In paper [3], the optimal location of a piezoelement is determined by taking into account both the electromechanical coupling coefficient and the value of the electric potential generated on the piezoelement surface. However, the authors of this paper restrict their consideration to the case when the selected location of the piezoelement is optimal only for suppressing the first mode of beam vibration.

The objective of this paper is to construct an algorithm, which will allow us to determine the piezoelement location such that provides generation of the maximal relative electric potential in the structure experiencing deformation at the prescribed frequency. This is essential for the purpose of providing maximal damping of vibrations at this particular frequency using the external shunt circuits. A search for an optimal location of the piezoelement is carried out in the framework of the natural vibration problem, which is formulated for electrovicoelastic body with the external electric circuits.

2 MATHEMATICAL STATEMENT OF THE NATURAL VIBRATION PROBLEM

Consider a piecewise-homogeneous body of volume $V = V_1 + V_2$, where the volume V_1 consists of homogeneous elastic parts and the volume, the volume V_2 - consists of piezoelectric elements. The piezoelectric elements are connected through the electroded surfaces to RLC-circuit of arbitrary structure comprising a resistance, capacitance and inductance elements.

For the problem of natural vibration we seek a solution written as

$$u_i(x,t) = \overline{u}_i(\vec{x})e^{-i\omega t}, \varphi_i(x,t) = \overline{\varphi}_i(\vec{x})e^{-i\omega t} \tag{1}$$

where ω - is the eigenfrequency of vibrations, $\overline{u}_i(\vec{x}), \overline{\varphi}_i(\vec{x})$ - are the eigenmodes of displacements and electric potential for the elastic sub-regions V_1 and V_2.

A detailed mathematical formulation of the problem is given in works [13-14].

The numerical implementation of the stated variational problem was carried out by the finite element method using the commercial ANSYS software package.

In the finite element formulation the variational problem reduces to the algebraic eigenvalue problem

$$\left(\left[K\right]-\omega^2\left[M\right]\right)\{u\}=0 \tag{2}$$

The condition of the existence of nontrivial solution to equation (2) is

$$D(\omega)=\det\left(-\omega^2\left[M\right]+\left[K\right]\right)=0 \tag{3}$$

Here, we accepted the following notation: $\{u\}$ - is the vector of nodal displacements, ω - is the eigenfrequency, $\left[K\right]$– is the stiffness matrix, $\left[M\right]$ – is the mass matrix.

In the problem of electroviscoous elasticity each point of the body is associated with the vector of state $\{u\}=\{u_1,u_2,u_3,\varphi\}^T$, where u_i are the components of the displacement vector, φ is the electric potential.

Out of other factors, determining the effectiveness of piezoelement performance in smart structures, we used the value of the electric potential generated on the electroded surfaces of piezoelements during the dynamic processes.

The results obtained from the solution of the natural vibration problem formulated for electroelastic bodies allow us to estimate the value of the electric potential generated on the surface of the piezoelement undergoing deformation.

The relative value of the electric potential can be estimated using the relative value of the strain energy in the zone of piezoelement location. The optimal values of the relative strain energy can be also obtained from the modal analysis.

At small dimensions of piezoelements, the attachment of which to the structure does not cause essential changes in the character of natural vibrations, the relative values of the strain energy in the zone of piezoelement location can be estimated using the eigenmodes of the structure including no piezoelements. i.e., based on the results of solving the problem of natural vibrations of elastic or viscoelastic bodies.

The value of the relative strain energy J in the zone of piezoelement location is defined by the strain distribution ε_r over the structure surface under the piezoelement

$$J=\int_{S_n}\varepsilon_r ds \tag{4}$$

where S_n is the area of the piezoelement, ε_r is the deformation in the direction perpendicular to the structure surface.

The application of voltage to a piezoelement attached to the surface of the structure induces a transverse deformation (deflection) of the structure. Note that the arising transverse deformation of the original structure depends on the location of the piezoelement. Hence, as was stated in review paper[4], the transverse deformation of the original structure can be used as a criterion for the optimal location of the piezoelement.

In work [14], it was stressed that from the viewpoint of computational costs the problem of natural vibrations is most efficient for determining the relative values of the electric potential generated on the surface of the piezoelement. The described approach was applied to plane objects in the form of plates. It has been shown that in this case, to provide a proper work of the piezoelement, it is necessary to locate it asymmetrically about the structure axis (if there is any).

From the analysis of literature it follows that the efficiency of peiezoelement performance in the smart-structures is often estimated in terms of the electromechanical coupling coeffi-

cient, which reflects the efficiency of converting the mechanical energy of vibrations into the electrical energy [1, 3, 15].

The coefficient of electromechanical coupling can be determined based on the eigenfrequencies of the structure with piezoelement in the open circuit regime (termed as o/c regime) and in the short-circuit regime (s/c regime). The o/c regime is realized in the case when the lower surface of the piezoelement is grounded (the lower surface potential is equal to zero) and the upper surface is free of load. In the s/c regime the potential prescribed on the electroded surfaces is equal to zero. As a result, the electromechanical coupling coefficient is evaluated by formula (5) [1]

$$K = \sqrt{\frac{\omega_{o/c}^2 - \omega_{s/c}^2}{\omega_{s/c}^2}} \tag{5}$$

The larger is the coefficient, the more effective is the energy conversion and thus the larger is the quantity of energy that can be dissipated by the elements of the external circuit.

As it was shown in works [1, 3, 15], the location of the piezoelement can be considered optimal in the case when the electromechanical coupling coefficient determined by formula (5) takes a maximum value.

At the same time, the larger is the value of the relative electric potential generated on the surface of the piezoelement undergoing deformation, the more effective is its performance in the structure. Hence, the location of the piezoelement on the surface of the structure, providing the generation of the maximal electric potential, is considered to be optimal.

In turn, as the value of the relative electric potential is specified by the degree of the piezoelement deformation, it should correlate with the value of the relative strain energy in the zone of piezoelement location. This observation has been supported in work [14] for plane objects in the form of plates.

In the present work, we have calculated the electromechanical coupling coefficient and determined the value of the relative electric potential and the value of the integral J, which characterizes the damping energy in the zone of peizoelement location defined by the distribution of the strain ε_r over the surface of the structure under the piezoelement.

The coefficient of the electromechanical coupling is the absolute quantity in contrast to the electric potential and strain energy, which are the relative quantities. This allows us to make a qualitative comparison of the obtained results.

3 NUMERICAL RESULTS

In this study we investigate a thin-walled shell in the form of a half cylinder having the following dimensions: $R = 76mm, L = 300mm, h = 0.25mm$ (fig.1). The shell was made of the elastic isotropic material displaying the following physic-mechanical characteristics: $E = 1.96 \cdot 10^{11} Pa; \nu = 0.3; \rho = 7700 kg/m^3$. The piezoelement was made of PZT-4 ceramics in the form of a ring segment, the physic-mechanical characteristics of which are given below in the cylindrical coordinates

$$C_{11} = 13.5 \cdot 10^{10}\ N/m^2 \qquad C_{12} = C_{13} = 7.43 \cdot 10^{10}\ N/m^2 \qquad C_{23} = 7.78 \cdot 10^{10}\ N/m^2$$
$$C_{22} = C_{33} = 13.9 \cdot 10^{10}\ N/m^2 \qquad C_{44} = C_{66} = 2.56 \cdot 10^{10}\ N/m^2 \qquad C_{55} = 3.06 \cdot 10^{10}\ N/m^2$$
$$\beta_{21} = \beta_{31} = -5.2\ C/m^2 \qquad \beta_{11} = 15.2\ C/m^2 \qquad \beta_{42} = \beta_{63} = 12.7\ C/m^2$$
$$э_{22} = э_{33} = 6.45 \cdot 10^{-9}\ F/m \qquad э_{11} = 5.62 \cdot 10^{-9}\ F/m \qquad \rho = 7500\ Kg/m^3$$

and has dimensions: $r = 76.25$ mm, $\varphi = 15.08^0$, $h = 1.2$ mm.

The upper and lower surfaces of the piezoelement are electroded and the polarization axis is aligned with the r -axis. The shell is rigidly clamped at the ends and freely supported at the generating line. (fig.1).

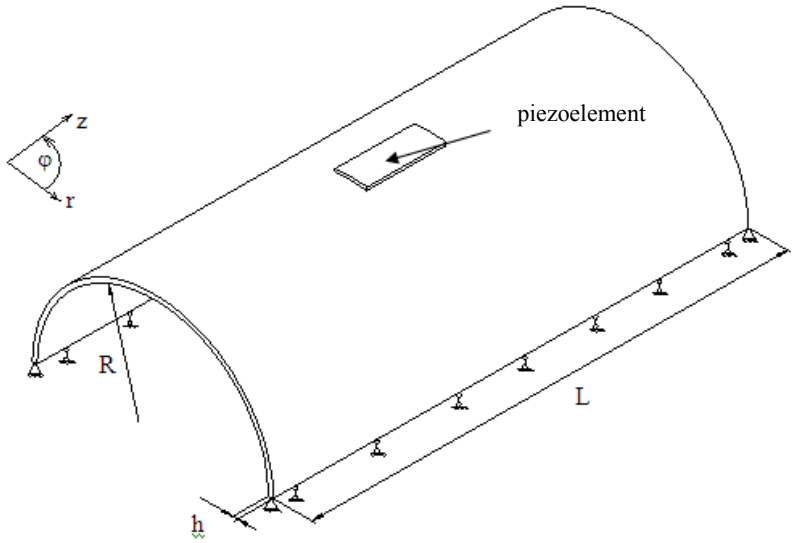


Figure 1: Schematics of the connected piezoelement

Let us estimate the efficiency of piezoelement performance using only two parameters: the value of the relative electric potential generated on the surface of the piezoelement under deformation and the value of the electromechanical coupling coefficient.

The optimal location of the piezoelement was determined by the scanning method using the previously proposed approach [13-14], which is based on solving the natural vibration problem.

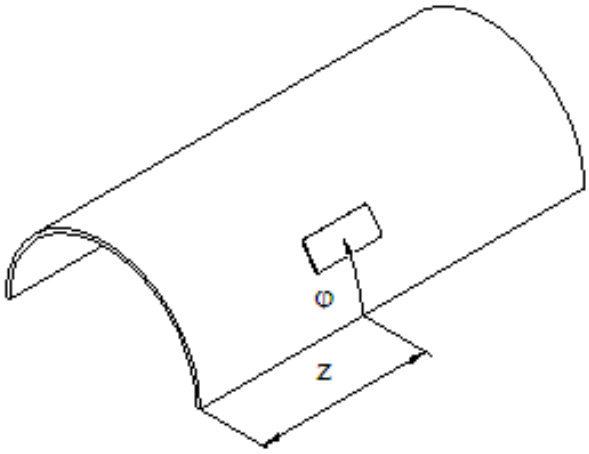
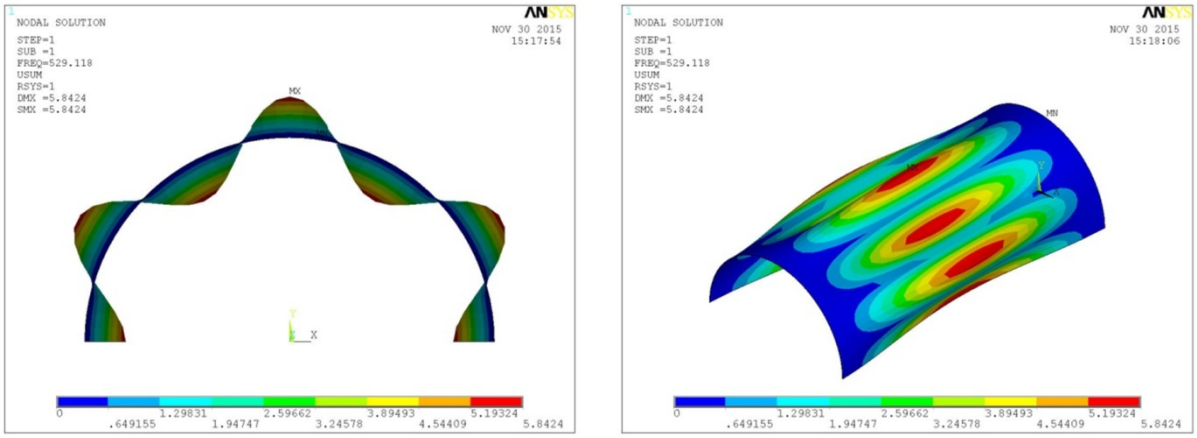


Figure 2: The directions of angular (φ) and length-wise (z.) variations in the coordinates of the piezoelement mass center.

To this end, a series of computations was performed for different variants of piezoelement location, which was determined by the angular and transverse coordinates of its mass center (fig.2). The range of angular variation of the coordinate is $\varphi \in [10°,170°]$ with a step of

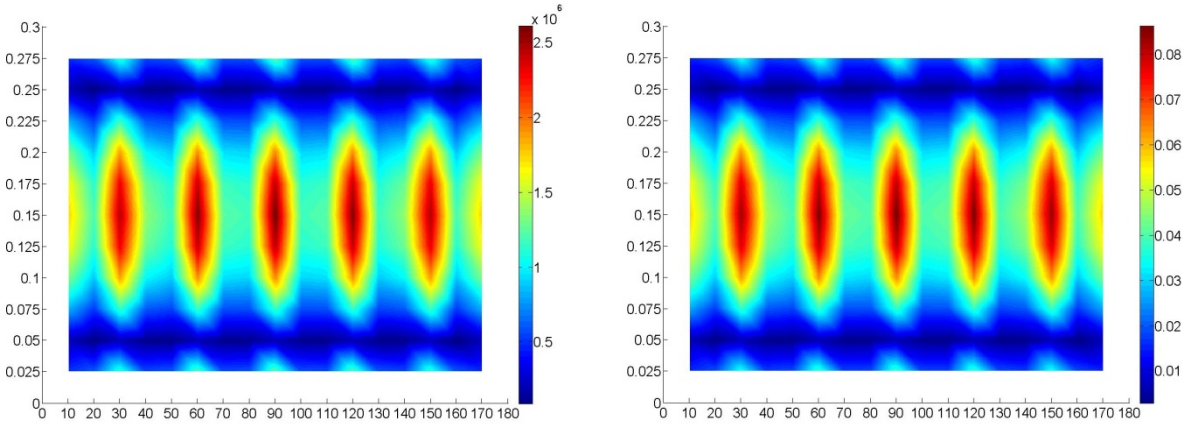
10˚.The range of length-wise variation of the coordinate is $z \in [25, 275]$ mm with a step of 25mm.

Figure 3 presents the shape of the first eigenmode of the system.



(a) (b)

Figure 3: The first vibrational eigenmode.



(a) (b)

Figure 4: Distribution of the values of the relative electric potential values (a) and electromechanical coupling coefficient (b) for the first vibrational mode of the shell.

Figure 4 presents images of the distributions of the computed relative electric potential (fig.4, a) and electromechanical coupling coefficient (fig.4,b) for the first vibration mode, which are used as the criteria of finding regions for the optimal location of the piezoelement on the shell surface.

To demonstrate this more clearly, the surface of the shell is unrolled onto a plane- the angular coordinate φ is laid off along the x-axis, and longitudinal coordinate z is laid off along the y-axis.

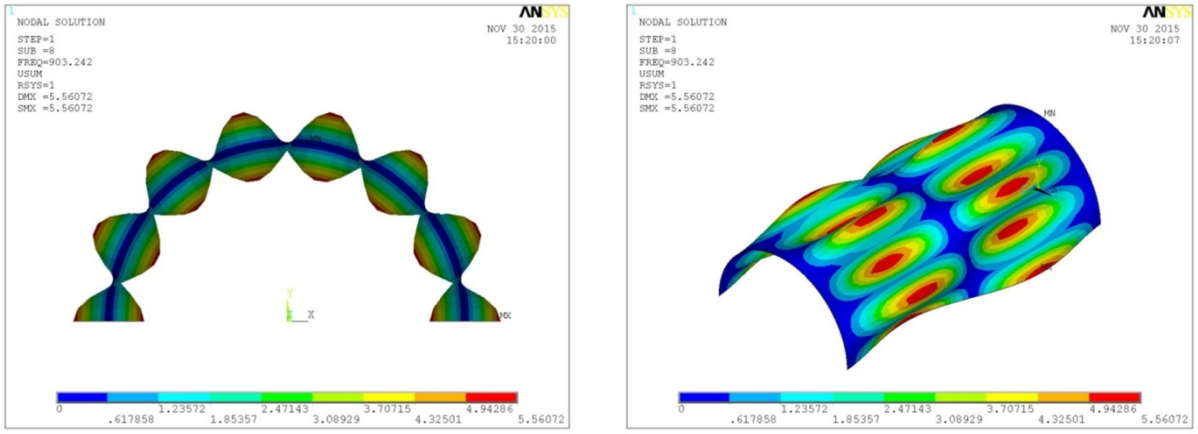
The analysis of the relative electric potential distribution (fig.4,a) allows us to conclude that the piezoelement performs at its best in the case when its mass center is located at the point with coordinates $\varphi = 90^{\circ}$, $z = 150$ *mm*.

Figures 4(a) and 4(b) demonstrate a complete qualitative agreement between the results of calculation with respect to two different parameters - electric potential and electromechanical coupling coefficient, which allows us to draw a conclusion about the equivalency of both pa-

rameters for deciding the question of optimal location of the piezoelement on the structure surface.

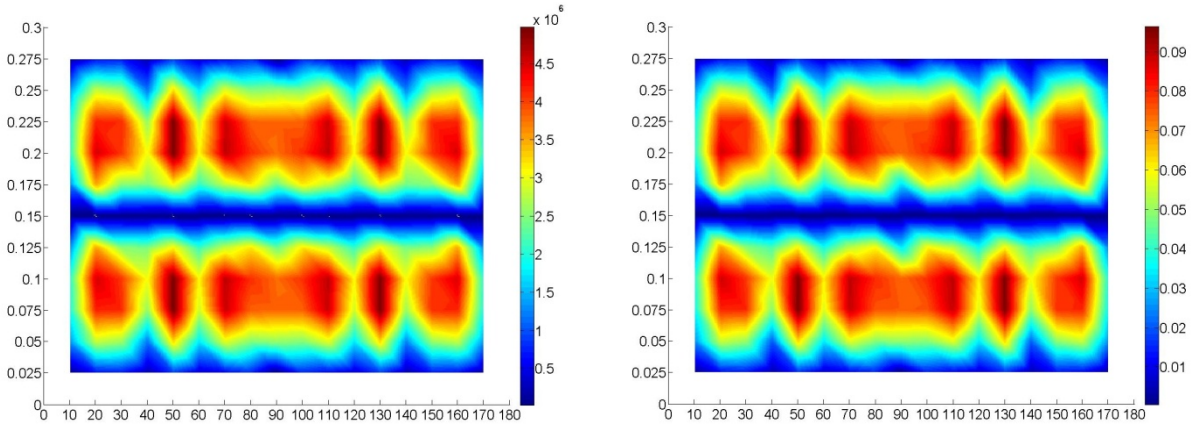
It should also be emphasized that the images of the distribution of the above parameters (the relative electric potential and electromechanical coupling coefficient) qualitatively correlate with the vibration mode: the points, at which the relative electric potential has the highest values coincide with the convexity points (points of maximal displacement amplitudes) of the vibration mode.

Note that for the shell frequencies from the second to the fourth order one can observe the same picture of distributions of the relative electric potential and electromechanical coupling coefficient, which determine the regions of optimal location of the piezoelement on the shell surface.



(a) (b)

Figure 5: The eighth vibrational eigenmode of the shell.



(a) (b)

Figure 6: The distribution of the values of the relative electric potential (a) and the electromechanical coupling coefficient (b) for the eighth vibrational eigenmode of the shell.

As another constitutive parameter we used the eighth mode of the shell, which is shown in fig.5. It also presents the results of computation of the relative electric potential (fig.6,a) and the electromechanical coupling coefficient (fig.6,b) for the mode of vibration which specifies the regions of optimal location of the piezoelement. As is evident from the figure, the correla-

tion between the distributions of the above parameters and vibrational mode of the shell is rather good.

According to the results displayed in fig. 6 (a, b) for the eighth vibrational mode, there are 4 optimal zones for the location of the piezoelement. For these zones the coordinates of the piezoelement mass center are as follows: *φ=50˚, z=87.5mm; φ=50˚, z=212.5 mm; φ=130˚, z=87.5mm; φ=130˚, z=212.5 mm*. At the same time the situation, similar to the one in fig. 6, is observed for the vibration frequencies of the shell from the fifth to the eighth one.

From the comparison of the results of computation presented in fig. 4and 6 it can be inferred that the location of the piezoelement such that its subsequent connection to the external electric circuit can provide effective damping of the vibrations at the first frequency, proves to be improper for damping vibrations at the eighth frequency and vice versa, the piezoelement location being optimal for damping of the vibration at the eighth frequency will fail to provide good performance of the piezoelement at the first frequency.

4 CONCLUSIONS

- In this paper, we have considered and verified the possibility of applying the natural vibration problem for electroviscoelastic bodies for optimization of smart structures based on the identification of appropriate zones for piesoelement location such that ensures the generation of the maximal relative electric potential at the corresponding vibrational mode.
- It has been shown that the results of determining the best location of the piezoelement on the structure surface providing a maximum increase in the efficiency of its performance based on the values of the relative electric potential generated at the electroded surfaces of the piezoelement and the electromechnical coupling coefficient are equally applicable for this particular optimization problem. The advantage of using the electromechanical coupling coefficient is that the computation yields its true values compared to the electric potential, the values of which are evaluated up to a factor. However, the latter criterion has one essential disadvantage - the necessity of performing two times as many calculations because of the need to determine the eigenfrequencies for two regimes - open circuit and short circuit regimes. In contrast, the value of the electric potential for each location can be obtained from a single computation run, namely in the open circuit regime.

The work supported by RFBR (projects № 15-01-03976, 14-01-96003 *r-ural*_a,) and UB RAS complex program (project 15-10-1-18).

REFERENCES

[1] N.W. Hagood, A. von Flotow, Damping of structural vibrations with piezoelectric materials and passive electrical networks. *Journal of Sound and Vibration* **146**(2), 243–268, (1991).

[2] Crawley E.F., de Luis J. Use of Piezoelectric Actuators as Elements of Intelligent Structures// *AIAA Journal*, **25**(10), 1373–1385, 1985.

[3] S. Venna, Y.-J. Lin. An Effective Approach for Optimal PZT Vibration Absorber Placement on Composite Structures, *Modern Mechanical Engineering* **3**, 21-26, 2013.

[4] V. Gupta, M. Sharma, N. Thakur. Optimization Criteria for Optimal Placement of Piezoelectric Sensors and Actuators on a Smart Structure: A Technical Review, *Journal of Intelligent Material Systems and Structures*, **21**(8), 1227-1243, 2010.

[5] F. Bachmann, A.E. Bergamini, P. Ermanni. Optimal piezoelectric positioning a strain-energy based finite element approach, *Journal of Intelligent Material Systems and Structures*, **23**(14), 1575–1591, 2012.

[6] J.C. Bruch Jr., J.M. Sloss, S. Adali and I.S. Sadek, Optimal piezo-actuator locations/lengths and applied voltage for shape control of beams. *Smart Mater. Struct*. **9**, 205-211, 2000.

[7] V.M.F. Correia, M.A.A Gomes, A. Suleman, C.M.M. Soares, C.A.M. Soares, Modelling and design of adaptive composite structures. *Computer Methods in Applied Mechanics and Engineering,* 185(2-4), 325-346, 2000.

[8] V.M.F. Correia, C.M.M. Soares, C.A.M. Soares, Refined models for the optimal design of adaptive structures using simulated annealing. *Composite Structures*, 54(2-3), 161-167, 2000.

[9] P. Kostka, K. Holeczek, W. Hufenbach, Structure-integrated Active Damping System: Integral Strain-based Design Strategy for the Optimal Placement of Functional Elements. *International Journal of Composite Materials*. **3**(6B), 53-58, 2013.

[10] R.E. Holman, S.M. Spencer, E.M. Austin, C.D. Johnson, Passive damping technology demonstration, *SPIE*, **3045**(1), 51-59, 1997.

[11] J.J. Hollkamp, R.W. Gordon, An experimental comparison of piezoelectric and constrained layer damping, *Smart Mater. Struct*., **5**, 715–722, 1996.

[12] S. Wu, T.L. Turner, S.A. Rizzi, Piezoelectric shunt vibration damping of F-iS panel under high acoustic excitation, *SPIE*, **3989**(2), 276-287, 2010.

[13] V.P. Matveenko, E.P.Kligman, N.A. Yurlova, M.A. Yurlov, Simulation and optimization of dynamic characteristics of piezoelectric smart structures. *Physical Mesomechanics,* **15**(3–4), 190-199, 2012.

[14] N.V. Sevodina, N.A. Yurlova, D.A. Oshmarin, The optimal placement of the piezoelectric element in a structure based on the solution of the problem of natural vibrations. *Solid State Phenomena*, **243**, 67-74, 2015.

[15] M. A. Trinidade, A. Benjeddou, Effective electromechanical coupling coefficients of piezoelectric adaptive structures: critical evaluation and optimization. Mechanics of Advanced Materials and Structures, **16**(3), 210-223, 2009.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DETERMINATION OF OPTIMAL PARAMETERS FOR A PASSIVE RL-CIRCUIT BY SOLVING THE PROBLEM ON NATURAL VIBRATIONS OF ELECTROELASTIC BODIES

Maksim A. Yurlov [1], Valerii P. Matveenko[2], Dmitriy A. Oshmarin[2], Nataliya V. Sevodina[2], Nataliya A. Iurlova[2]

[1] Institute of Continuous Media Mechanics of the Ural Branch of RAS
Acad. Korolev str., 1, Perm, 613013 Russia
e-mail: yurlovm@icmm.ru

[2] Institute of Continuous Media Mechanics of the Ural Branch of RAS
Acad. Korolev str., 1, Perm, 613013 Russia
e-mail: mvp@icmm.ru, oshmarin@icmm.ru, natsev@icmm.ru, yurlova@icmm.ru

Keywords: Electoelasticity, External electrical circuit, Piezoelement, Natural vibrations, Damping, Numerical modeling.

Abstract. *The use of elements made of piezoelectric materials and connected to external passive electric RLC-circuits is one of the effective ways of damping the vibrations of structure elements. The key problem for this technique is identification of optimal parameters for the elements of the external circuits, which can provide maximal damping of a single mode or simultaneous damping of several modes. In this study, based on the mathematical formulation of the natural vibration problem for electroelastic bodies with external electric circuits we have developed an approach for evaluating the parameters of the external RL-circuit which can provide maximal damping of the particular vibration modes. A criterion for determining the external circuit parameters has been proposed based on the analysis of the obtained vibrational complex eigenfrequencies. For implementation of the proposed approach a finite-element algorithm has been constructed using the commercial ANSYS software package. A series of numerical experiments have been carried out for e few objects to test the performance of the developed technique.*

.

1 INTRODUCTION

Damping capacity of materials, namely their ability to absorb vibrations, plays an important role in the dynamic behavior of structures [1]. A qualitative estimate of dissipative properties of structures generally relies on the results of solving two problems. The first problem deals with consideration of natural vibrations. Here the system dissipation manifests itself as vibration damping and the rate of damping qualitatively evaluates the dissipative properties of the system. The second problem is concerned with the forced steady-state vibrations. In this case, the dissipative properties of the system reveal themselves in the restriction of resonance amplitudes.

A search for optimal structures with maximal damping capacity using the numerical simulation methods requires heavy computation. On the one hand, it is necessary to investigate the influence of parameters on the damping properties of the system in the examined range of vibrations. On the other hand, one needs to analyze the behavior of the structure in a certain range of dynamic actions for each combination of these parameters. In the case of natural vibration this implies finding solution to the dynamic problem under different initial conditions, whereas in the forced vibration problem it is necessary to find solution to the examined problem for different loading regimes in the specified frequency range of vibrations.

An alternative to the above approaches, including the method of solving the optimization problem is the problem of natural vibrations allowing us to estimate damping properties of the system, irrespective of external force impacts, kinematic and other actions. In this case the eigenfrequencies are the complex quantities. Their real parts are the frequency and the imaginary parts are the damping ratio (the damping rate) of natural vibrations.

The object of investigations is a piece-wise homogeneous body composed of elastic and viscoelastic elements. The dissipative properties and resonance frequencies of the system can be controlled by the mechanical properties of the material, geometrical parameters of the system and boundary conditions. The employment of new materials in structure, for example functionally graded materials (smart-structures) makes it possible to extend the list of control parameters.

In the present paper, we consider complex structures, the elements of which are made of piezoelectric material. A distinctive feature of the structure is that the electroded surfaces of piezoelectric elements are located at certain points of the structure and connected via external electric circuits, comprising a resistor, (resistance), inductor (inductance) and capacitor (capacitance), to zero potential point.

The piezoelectric elements coupled with a shunt circuit and connected to a mechanical structure serve as the devices that provide energy dissipation provoking thus additional damping of vibrations [2].

The available literature gives a comprehensive description of the applications of shunt technique for damping vibrations of space structures and helicopter blades, suppressing the nonlinear supersonic panel flutter, controlling the mode shapes, damping the telescope vibrations, turbomachinery blade vibrations, reducing the level of noise in vehicles, suppressing the echo reflected from submarines and vibrations of computer disc drives, etc.

The use of shunt circuits offers some advantages over the traditional methods of damping the mechanical vibrations. Among these is the ease of arranging the external electric circuits generated by a set of standard electronic elements. Sometimes installation of additional electronic elements or power sources is not required at all and, in a few particular cases there is no need in feedback sensors. Furthermore, from the viewpoint of stability and reliability the proposed approach compares favorably with other techniques because it does not require the

incorporation of complicated digital processors and in many situations the development of parametric models for designing new vibration damping systems is not necessary either [1].

The simplest shunt circuit is the connection of a resistive element (resistor) via the electrodes. Although the first reference to this method was made in work [3], the idea of vibration damping with the aid of shunting by external electric circuits was attributed to Hagood and von Flotow [4], who showed that a series -connected RL-circuit can essentially weaken a single vibration mode [5].

From a mechanical standpoint it can be said that the RLC-circuits play the part of external elastic and viscoelastic elements. In the context of such formulation the calculated complex eigenvalues determine the frequencies of natural vibrations and damping ratios. The complex frequencies and eigenmodes depend not only on the structural and engineering characteristics of shell systems but also on the parameters of piezoelectric elements, their arrangement schemes and characteristics of the external circuits comprising a resistors, inductor and capacitors. Targeted variation of these parameters allows one to optimize the resonance frequencies and damping ratios

2 MATHEMATICAL STATEMENT OF THE PROBLEM

A complete mathematical statement of the problem for electroviscous piecewise-homogeneous bodies is given in work [6].

The variational equation of motion for a body consisting of elastic and piezoelectric elements is formulated using the relations of the linear elasticity theory and quasi-static Maxwell equations [1, 7-8]:

$$\int_{V_1}\left(\sigma_{ij}\delta\varepsilon_{ij}+\rho\ddot{u}_i\delta u_i\right)dV+\int_{V_2}\left(\sigma_{ij}\delta\varepsilon_{ij}-D_i\delta E_i+\rho\ddot{u}_i\delta u_i\right)dV-\int_{\Omega_\sigma}\delta u_i P_i d\Omega-\int_{\Omega_{el}}q_e\delta\varphi d\Omega=0 \quad (1)$$

Here, $\int_{\Omega_{el}} q_e\delta\varphi d\Omega$ - is the work done by the external electric circuit, D, E – are the vectors of the electric inductance and electric field strength; σ^{ij} – the components of a symmetric Cauchy stress tensor, $\varepsilon_{ij}=\frac{1}{2}\left(u_{i;j}+u_{j;i}\right)$ – are the components of the tensor of linear deformations, u_i – are the components of the displacement vector, P_i - are the components of the load vector, q_e and φ are the surface charge density and electric potential, V_1 is the volume of the elastic and viscoelastic parts of the body, Ω_{el} is the surface that bounds the body of volume V_2. Equation (1) is written in terms of the rectangular Cartesian coordinates.

For the electric field the condition of equipotentiality is fulfilled: $\varphi_{,i}=-E_i$.

For isothermal processes in linear electroelastic media the following physical relations hold true:

$$\sigma_{ij}=C_{ijkl}\varepsilon_{kl} \quad \text{- for elastic part of volume } V_1 \quad (2)$$

$$\left.\begin{aligned}\sigma_{ij}&=C_{ijkl}\varepsilon_{kl}-\beta_{ijk}E_k\\ D_k&=\beta_{ijk}\varepsilon_{ij}+e_{ki}E_i\end{aligned}\right\} \text{ - for piezoelement of volume } V_2 \quad (3)$$

Here C_{ijkl} is the tensor of elastic constants, β_{ijk} and e_{ki} are the tensors of piezoelectric and dielectric coefficients. For viscoelastic materials the tensor of elastic constants C_{ijkl} in relation (2) can be replaced by the corresponding viscoelastic operator [9-11].

In the electroelastic problems the boundary conditions can be divided into two groups: mechanical and electrical conditions. In the natural vibration problem the mechanical boundary conditions are written as

$$S_\sigma: \quad \sigma_{ij} n_j = 0, \; S_n: \quad u_i = 0, \tag{4}$$

where $S = S_\sigma + S_n$ is the surface that bounds the volume V of the examined body.

The formulation of the physically realizable electrical boundary conditions depends on the way of electrical energy supply to a piezoelectic element. Both, the supply and release of energy from the deformed piezoelectric element is carried out via the electroded coatings applied to some parts of the body surface. They are assumed to be ideal conductors having negligible mass. Coating of the body surface Ω_{el} with a thin current conducting layer (electroding) ensures its equipotentiality.

$$\int_{\Omega_{el}} \delta\varphi \, q_e \, d\Omega = \delta\varphi \int_{\Omega_{el}} q_e \, d\Omega = \delta\varphi Q^{el} \tag{5}$$

Here Q^{el} is the total electrode charge.

The numerical implementation of the stated variational problem is carried out by the finite element method using the commercial ANSYS software package.

3 NUMERICAL RESULTS

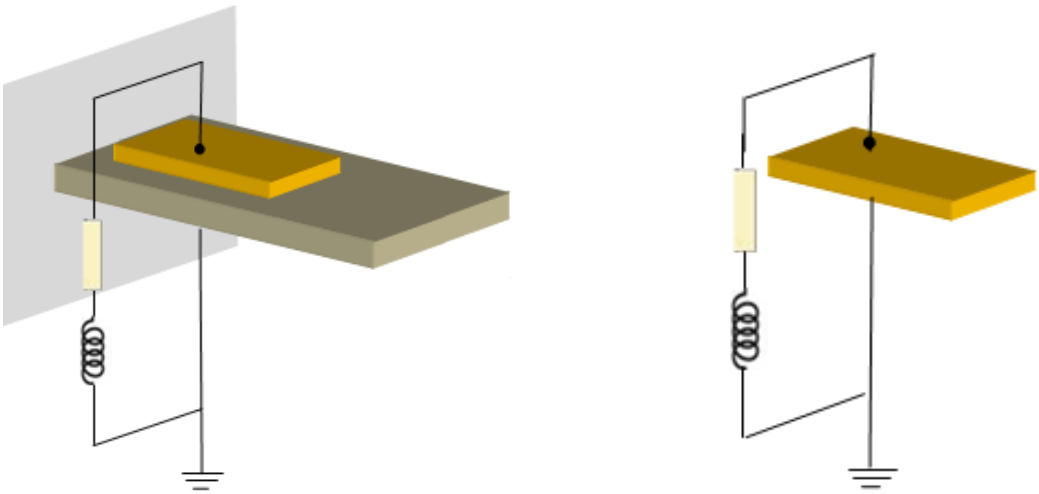
The object of our investigation is a cantilevered plate, the surface of which is coupled with a piezoelement and external series-connected *RL*-circuit shunting its electroded surfaces (fig.1). The plate measures 210x26x0.6 mm and the piezoelectric element measures 50x20x0.36 mm. The piezoelement is located 12mm away from the edge of the plate and symmetrically about its axis of symmetry.

The plate is made of viscoelastic material with the following mechanical characteristics: bulk modulus – B =167.7 *GPa,* instantaneous shear modulus – G =76.923 *GPa,* loss tangent for the shear modulus G - $\delta = 0.01$, specific density - $\rho = 7800$ kG/m^3. The piezoelement is made from piezoceramics PZT-4-z [1] and polarized along the z-axis. The plate and piezoelement are perfectly bonded.

The numerical analysis was carried out for the following structures: the plate taken alone, the plate with the piezoelement connected to its surface, the plate with the piezoelement and external series-connected *RL*-circuit (fig.1,a), the pezoelement by itself with the external electric circuit connected to its electroded surfaces (fig.1,b).

Table 1 presents the values of the first eight eigenfrequencies of the plate with the piezoelement ω_R and the values of the damping ratio ω_I. Note that that column 2 includes the values of eigenfrequencies obtained in the open circuit (o/c) regime and the column 3 includes the values obtained in the short circuit ((s/c) regime.

The developed software package for computing the vibrational eigenfrequencies of electroviscoelastic structures connected to the external electric circuit was used to calculate eigenfrequencies of the examined objects at different values of the resistance R and inductance L of the external electric circuit.



a b

Figure1: Computational scheme; (a) the plate with piezoelement and external series-connected *RL*- circuit (system), (b) the peizoelement with external electric circuit connected to its electroded surfaces (electric circuit).

№	Eigenfrequencies of vibrations			
	Plate $\omega = \omega_R + i\omega_I$	Plate with piezoelement in o/c regime $\omega = \omega_R + i\omega_I$	Plate with piezoelement in s/c regime $\omega = \omega_R + i\omega_I$	Plate with piezoelement and external electric circuit (*R=100* ohm, *L=1000* (H) $\omega = \omega_R + i\omega_I$
	1	2	3	4
1	11.239 - 0.0473i	13.860 - 0.0490i	13.576 - 0.0467i	13.548 - 0.0464i
				47.347 - 0.04089i
2	70.396 - 0.297i	72.334 - 0. 290i	72.149 - 0.2895i	72.472 - 0.2889i
3	177.707 - 0.879i	198.325 - 0. 795i	197.899 - 0.7906i	198.351 - 0.7951i
4	197.271 - 0.830i	204.451 - 0. 874i	204.451 - 0. 8737i	204.451 - 0. 8737i
5	387.268 - 1.622i	404.159 - 1.542i	399.795 - 1.5119i	404.220 - 1.5421i
6	477.558 - 2.064i	489.406 - 2.021i	489.406 - 2.0215i	489.406 - 2.0214i
7	537.424 - 2.649i	584.108 - 2.472i	584.108 - 2.4717i	584.108 - 2.4717i
8	641.704 - 2.672i	674.968 - 2.509i	665.456 - 2.4730i	675.018 - 2.5088 i

Table 1: Spectrum of the system eigenfrequencies

The piezoelectric element possessing the capacitance properties form together with the external electric circuit the electric oscillatory circuit (fig. 1,b).

In the case when the external electric circuit is connected to the plate with piezoelement, this leads to the appearance of additional frequency in the spectrum of eigenfrequencies. The value of this frequency is determined by the values of *R* and *L*.

Since the piezoelectric elements are the reversible mechanical- to- electric transducers, i.e., the devices capable of converting the mechanical energy into the electrical energy and vice versa, a coincidence of the structure vibration frequency with the frequency of the electric oscillatory circuit causes a maximum amount of the mechanical vibration energy to convert

into the electrical energy and dissipate in the external electric circuit. Therefore, to provide effective damping of the structure vibrations it is necessary to determine such values of the parameters of the external *RL*–circuit, at which the eigenfreqiecies of the structure and electric circuit coincide.

For our investigation we used the scanning method, which implies variation of the sought-for optimization parameters (the values of the inductance *L* and resistance *R*) with the aim to find the values, which could provide the maximum damping ratio of the plate with piezoelectric element and external electric circuit. This method allowed us to investigate the dependence of the normal eigenfrequencies of the examined structure (fig.1,a) first of al on the value of inductance of the external electric circuit *L*, and then on the value of the resistance of the external electric circuit *R* for the evaluated optimal (providing close coincidence of eigenfrequencies of the plate with piezoelement and electric oscillatory circuit) value of the inductance *L*.

Figure 2 shows the dependence of eigenfrequency of the plate with embedded piezoelement and external electric circuit in the vicinity of the first eigenfrequency of the plate with piezoelement on the value of the inductance *L* at a small constant value of the resistance, for example at *R=10* Ohms. In the figure, ω_{R1}, ω_{Re} denote the eigenfrequencies of the plate with the piezoelement and electric circuit, respectively, and ω_{I1}, ω_{Ie} denote the corresponding damping ratios.

Figure 2 demonstrates a highly pronounced effect of convergence of frequencies ω_{R1}, ω_{Re} in the range of the inductance L=10000 ÷ 20000 *H*. At some value of the inductance $L=L_r$ (in our calculations $L_r \sim 12000$ *H*) one can observe a maximal convergence of frequencies and a sharp change (in modulo) of the damping ratios: decrease of ω_{I1} and increase of ω_{Ie}.

The obtained results show that the inductance *L* has an essential effect on the value of eigenfrequency of the electric circuit ω_{Re} in the eigenfrequency spectrum of the system consisting of the plate with piezoelement and external electric circuit. It provides a shift and convergence of the eigenfrequency of the plate with piezoelement with the eigenfrequecy of the electric circuit.

Figure 3 shows how the resistance *R* affects the eigenfreqiencies of the system comprising the plate with piezoelement and external electric circuit in the vicinity of the first vibrational eigenfrequency of the plate with piezoelement at the evaluated constant optimal value of the inductance $L_r = 12000$ *H*.

As it is seen from fig.3, the frequencies of the electric circuit and the plate with piezoelement remain unchanged up to a certain value of the resistance *R*.

A convergence of the vibration frequencies ω_{R1}, ω_{Re} and a sharp increase (in modulo) of damping ratios ω_{I1}, ω_{Ie} take place in some range of resistance (*R*=300 ÷ 800 *kOhms*). In our calculations the closest coincidence of frequencies ω_{R1}, ω_{Re} occurs when the value of the optimal resistance is $R_r \sim 400$ *kOhm*.

Further, with increasing resistance the frequency of the electric circuit reduces up to zero, which suggests the absence of circuit vibrations and the eigenfrequecy of the plate with piezoelement changes to the value of frequency in the o/c regime, which implies the absence of shorting the electroded surfaces of the piezoelement.

Figure 4 illustrates the behavior of the imaginary parts of the first eigenfrequency of the plate with piezoelement in the vicinity of the values of parameters *L* and *R*, at which damping of the vivbrations occurs at the maximum rate. The values of the external circuit parameters *L* and *R*, at which the imaginary part of the eigenfrequencies of the plate with piezoelement

reaches its maximum value are just the values, which are optimal for damping the vibrations of the system at these frequencies.

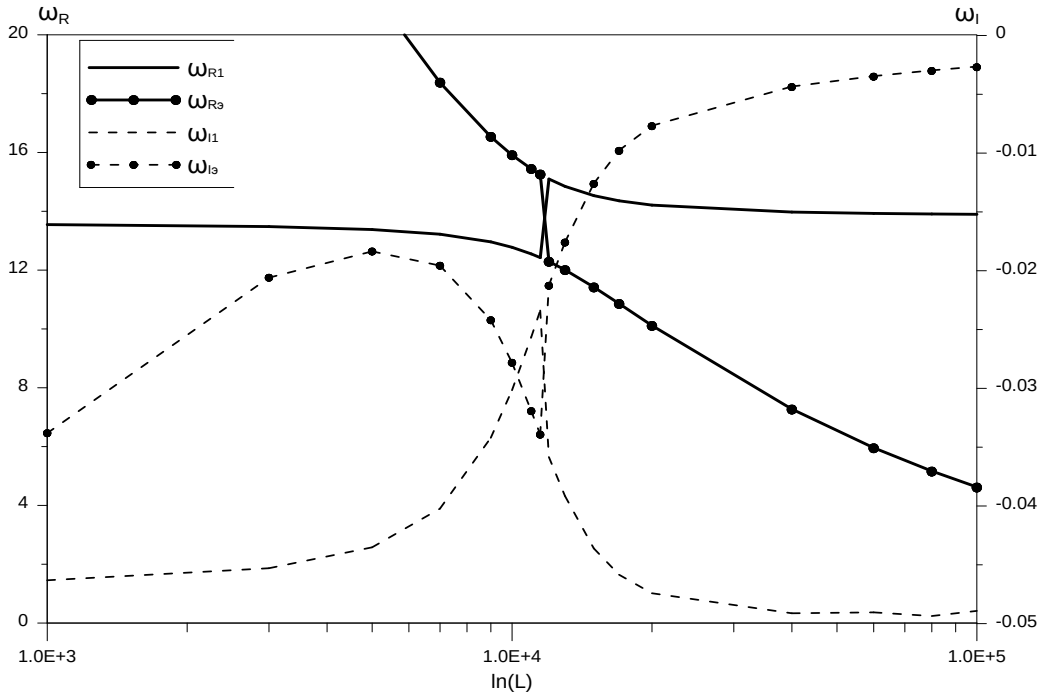


Figure 2: The influence of the value of inductance *L* (*R*=10 *Ohm*) on the first eigenfreqiency of the system consisting of the plate with piezoelement and external series-connected *RL*-circuit.

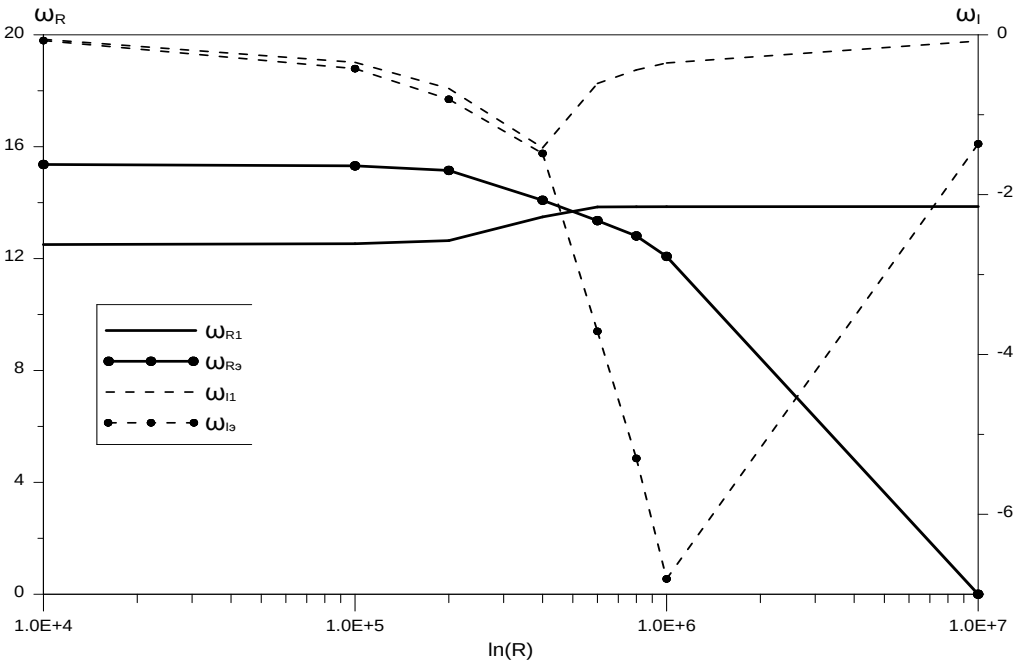


Figure 3: The influence of the value of resistance *R* (*L*=12000 *H*) on the first eigenfreqiency of the system consisting of the plate with piezoelement and external series-connected *RL*-circuit.

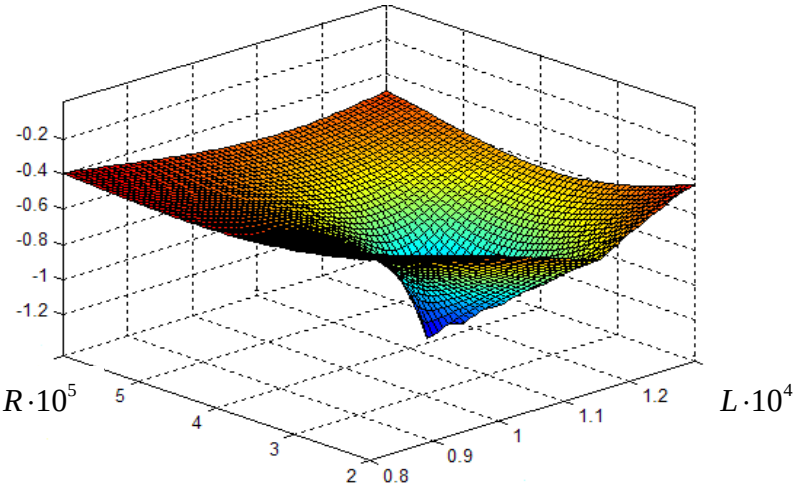


Figure 4: The behavior of the imaginary part of the first eigenfrequency of the plate with piezoelement

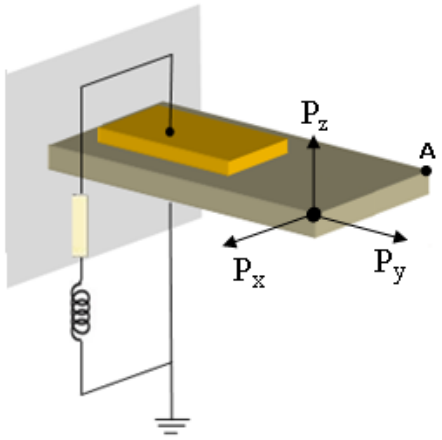


Figure 5: Computational scheme for determining the ampltude-frequency characteristics of the system.

Based on the optimal values of the parameters for the external electric circuit (L=11198 *H*, R=402493 *Ohm*) calculated for the first resonance frequency of vibrations we constructed the amplitude-frequency characteristics of the displacements U_s ($U_s = \sqrt{U_x^2 + U_y^2 + U_z^2}$) of point A on the cantilever part of the plate (fig.6) in the range of the first four resonance frequencies produced by the action of external axial harmonic forces $P_X = P_Y = P_Z$ =0.01 N applied to the angular point on the simply supported end of the plate (as shown in fig.5). Here the dashed line shows the AF characteristics of the plate with piezoelement and without the external electric circuit. The difference in the AF characteristics is observed only in the vicinity of the first resonance.

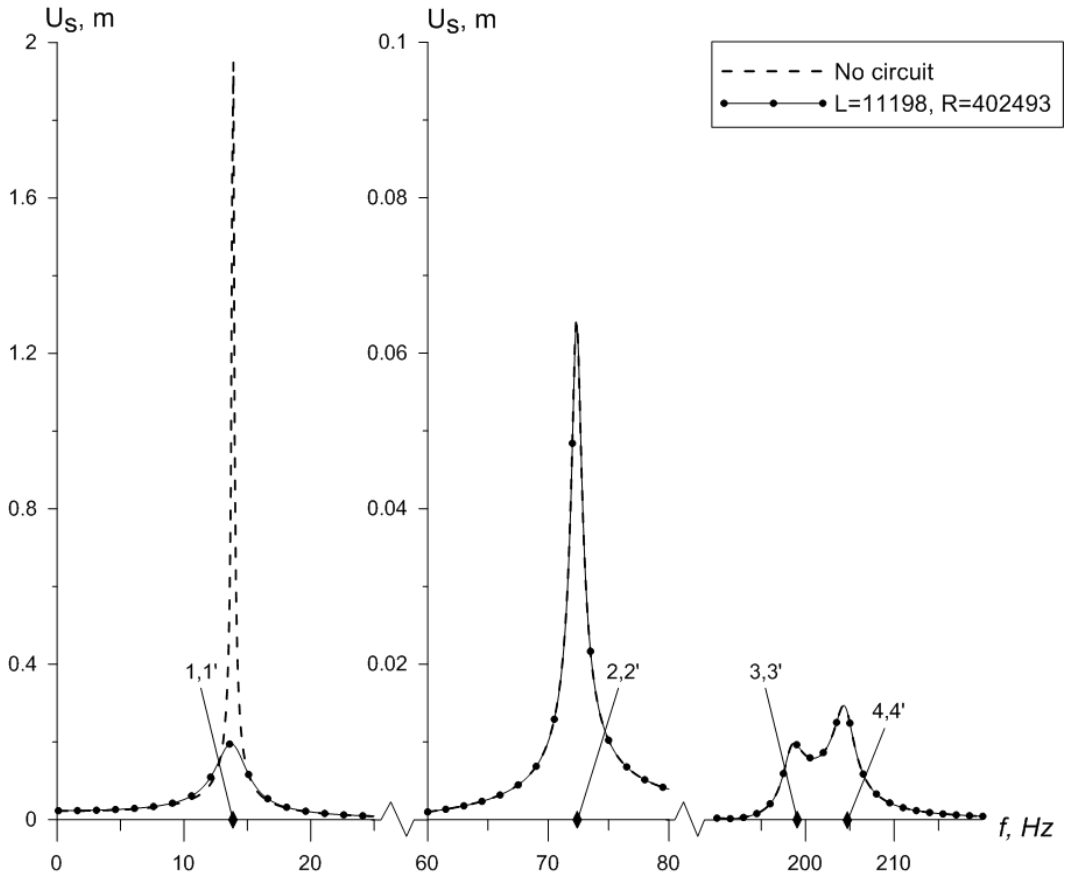


Figure 6: The amplitude-frequency characteristics of the plate with piezoelement and external electric circuit (fig. 5) in the range of the first 8 resonance frequencies in the case of tuning the electric circuit to the first resonance frequency.

As it is seen from the fig.6, a reduction of the displacement amplitude U_s at the free angular point A on the cantilever part of the plate at the first resonance frequency is - by a factor of 11.47. Note that in this case there is no change in the displacement amplitude U_s at other resonance frequencies (coincidence of the solid and dash lines in fig.6).

4 CONCLUSIONS

- The paper considers the problem of modeling and optimizing the dynamic characteristics, namely the resonance frequencies and the parameters characterizing the damping properties of the systems embedding piezoelectric elements in the presence of external electric circuits including resistance and inductance elements.
- A method for determining the optimal parameters of the external electric circuit, which will provide the maximal suppression of particular vibration frequencies has been developed based on the solution of the natural vibration problem for electro-viscoelastic bodies with external electric circuits.
- Numerical examples have been discussed to demonstrate the efficiency of using the problem of natural vibrations of electro-viscoelastic bodies with external electric circuits for optimization of damping properties of systems made on the basis of pizoelectric materials.

- It has been shown that the natural vibration problem is more effective than the problem of forced steady-state vibrations, since the obtained complex values of eigenfrequencies are the system characteristics, which do not depend on the type and method of external load application, which essentially reduces the computation time.
- It has been found that the imaginary part of the complex eigenfrequencies is the parameter, which characterizes the vibration damping ratio at the specified parameters of the external electric circuit.

The work was supported by the Russian Foundation for Basic Research (in the framework of projects № 16-31-00094-young_a, 15-01-03976-a, 14-01-96003 r_ural_a).

REFERENCES

[1] A. Nashif, D. Jones, J. Henderson, *Damping of vibrations* / Translated into Russian by L.G. Korneychuk, edited by E.I. Grigoliuk. - M.: Mir, 1988.

[2] H.-J. Lee, D. Saravanos, Layerwise finite elements for smart piezoceramic composite plates in thermal environments. *NASA TM-106990 AIAA-96-1277*, 48, 1996.

[3] R.L. Forward, Electronic damping of vibrations in optical structures. *Journal of Applied Optics*, **18**(5), 690–697, 1979.

[4] N.W. Hagood, A.Von Flotow, Damping of structural vibrations with piezoelectric materials and passive electrical networks. *Journal of Sound and Vibration*, **146**(2), 243–268, 1991.

[5] S.O.R. Moheimani, A.J. Fleming, *Piezoelectric transducers for vibration control and damping*. Springer, 2006.

[6] V.P. Matveenko, E.P. Kligman, M.A. Yurlov, N.A. Yurlova. Modeling and optimization of dynamic characteristics of smart-structures with piezoelectric elements. *Physical mesomechanics*, **15**(1), 75-85, 2012.

[7] V.Z. Parton, B.A. Kudryavtsev, *Electromagnetic elasticity of piezoelectric and electrically conducting bodies*. M: Nauka, 1988

[8] V.G. Karnaukhov, I.F. Kirichok, *Электrothermoviscous elasticity.* Kiev: Nauk dumka, 1988.

[9] V.P. Matveyenko, E.P. Kligman, Natural Vibration Problem of Viscoelastic Solids as Applied to Optimization of Dissipative Properties of Constructions. *Journal of Vibration and Control*, **3**, 87-102, 1997.

[10] E.P. Kligman, V.P. Matveenko, N.A. Yurlova. Dynamic characteristics of thin-walled electroelastic systems. *Bulletin of RAS, SM*. **2**, 179-187, 2005.

[11] V.P. Matveenko, E.P. Kligman, N.A. Yurlova, M.A. Yurlov, Optimization of the dynamic characteristics of electroviscoelastic systems by means of electric circuits. *Advanced Dynamics and Model Based Control of Structures and Machines* (ed. H. Irschik, M. Krommer, A.K. Belyaev), Springer-Verlag/Wien, 151-158, 2011.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

COMPUTATIONAL MODELLING OF MECHANICAL BEHAVIOR OF BIOLOGICAL TISSUES FOR BIOMEDICAL APPLICATIONS

Simin Li[1], Yang Liu[1], Juan Du[1], Begum Zeybek[1], Lorenzo Zani[1], Mark P. Lewis[2] and Vadim V. Silberschmidt[1]*

[1]Wolfson School of Mechanical and Manufacturing Engineering

[2]School of Sport, Exercise and Health Sciences

Loughborough University, UK

E-mail: V.Silberschmidt@lboro.ac.uk

Keywords: Computational Modelling, Tissues, Biomedical Applications

Abstract. *Computational modelling of biomedical applications has gained significant momentum in recent years, in part to meet demands related to recent technical advancements in manufacturing of personalized biomedical equipment. However, our understanding of mechanics of biological tissues, their properties and performance as well as their interaction with biomedical equipment still remains limited. This is a result of multiple factors, most important being a hierarchical and heterogeneous nature of biological tissues, non-trivial loading and environmental conditions to which they are exposed as well as multi-disciplinary nature of the systems involved.*

This paper presents an overview of the latest research activities and achievements in the area of mechanics of biomaterials and tissues at Loughborough University, UK. It covers various types of biological materials and tissues – both hard (bones) and soft (muscles, etc.) – that have been studied in previous studies [1-4] at various spatial and temporal domains. These studies laid a foundation for development and implementation of advanced computational modelling of mechanics of these biological tissues at different stages (healthy, diseased and traumatic conditions) and for several areas of biomedical applications (injury prevention, wound care and rehabilitation). Performed numerical simulations, on the one hand, elucidate processes of deformation of biological tissues and, on the other hand, provide solutions for design and optimization of medical and rehabilitation procedures and devices. This work underpins a unique partnership between engineers, clinics and rehabilitation centres in UK aiming to transfer the latest scientific and technological advancements into personalized biomedical applications using computational schemes and tools.

1 INTRODUCTION

Every year approximately 6,000 major limb amputations are performed in England according to the National Health System [5]. The clinical practice after an amputation involves a long period of physiotherapy and rehabilitation that requires the use of prosthesis. The main function of prosthesis for amputee is to substitute a missing part of the body with an artificial device that restores structure and functionality to the maimed musculoskeletal system caused by a limb loss.

During the initial rehabilitation stage, routine wound-healing techniques are usually incapable of achieving complete wound closure. Therefore, advanced wound-care methods are applied. The care of complex wounds is one of the most significant challenges for healthcare systems today. In an ideal scenario, residual extremity is covered with well-vascularized muscles, fascia and skin. However, in the case of traumatic amputation, the remaining skin structure is not always sufficient to fully cover the operation area, and it is therefore difficult to dress the wounded area [6-9]. To overcome this problem, a negative-pressure wound therapy (NPWT) is applied to (a) protect the open-wound area from infection instead of using conventional dressing and (b) promote the healing process [10].

NPWT has become a significant part of modern wound care and is used routinely in hospitals throughout the world. It involves controlled application of sub-atmospheric pressure to a wound bed through a wound filler (foam or gauze) placed in the wounded area [11]. The wound is then sealed with an adhesive drape that allows pressure to be applied and helps to provide a moist environment supporting the wound healing. The wound filler and drape protect the wound bed from bacteria and other contaminants and reduce a risk of friction or shear as well as other bodily fluids, enhancing the body's ability to heal [12]. Sub-atmospheric pressure removes an exudate and promotes healing in various ways.

After a successful wound healing, the next rehabilitation step for the amputee is to wear prosthesis. A typical lower-limb prosthesis is composed by four principal components: a socket, a suspension mechanism, a pylon providing the anchor point for a prosthetic foot. The socket is the most important factor for a successful prosthesis design because it acts as an interface between the residual and the artificial limb. The main function of the socket is to transmit forces during the ambulation, assure prosthesis stability, hold and protect the residual limb. Currently, the socket design and manufacture still relies on an artisanal process, which is non-standard and non–repeatable resulting in non-predictable performance [13, 14]. From a patient's point of view, the most important functionality of the socket is its usability, where a socket should provide the wearer at least the stability and comfort during day-to-day wearing.

Instead of using a socked-based prosthesis approach, another method to replace the limb loss could be the use of an osseointegrated implant [15]. A successful use of such an implant relies on structural integrity of tissues surrounding the bone, which is subjected to a continuous remodelling process [16]. The investigation of this process is important to predict and improve bone adaptation after amputation. Human bone, as one of the most functional tissues in the body, exhibits the ability of restructure itself in responding to mechanical as well as metabolic stimuli, adjusting its internal structure and external shape. Although numerous computational models have focused on a fracture healing process, few studies have investigated the bone-remodelling process after lower limb amputation. [16].

In the past decade computational modelling has become increasingly popular in the field of biomechanics thanks to exponentially growing computer power [17]. In order to reach the goal of an individualised model, a set of tools including medical imaging, image acquisition and processing, mesh generation, material modelling and finite-element simulation becomes a necessity. The field of biomechanics suffers from one very severe restriction: in general, it is

not possible for ethical reasons to measure forces and pressure within the human body. Thus, a typical measurement technology in biomechanics works on the interface between a body and its environment. The only possible way to address this challenge is to develop realistic biomechanical models of the human body, which can adequately predict mechanical behaviours of the tissues [17]. For this purpose, the developed computational models are provided as a framework allowing an insight into the analysis of mechanical interaction between tissues, wounded area and biomedical devices.

The aim of this paper is to give a better understanding of healing, recovery and remodelling processes at various stages of rehabilitation period after a lower limb amputation.

2 MATERIALS AND METHOD

Three finite-element models were developed in order to observe (i) the effects of negative pressure healing therapy after trans-femoral amputation, (ii) pressure distribution acting on transtibial residuum during daily activities, and (iii) prediction of a bone-remodelling process that occurs after the transtibial amputation. For the purpose of creating more realistic geometry in the three models, CT-scan-based image data for human legs containing bones and soft issues (including skin, fat, muscle, tendon and ligament) were obtained. The acquired geometries were imported into ABAQUS 6.14 software for individual finite-element analysis of each model.

2.1 Model of negative pressure wound therapy

NPWT involves a controlled application of sub-atmospheric pressure to a wound bed through a wound filler (foam or gauze), placed in the wound. The developed model contains three main domains: a rigid bone structure, a soft tissue with a wounded area (12 mm in depth) located at the distal region of the residual limb and wound filler modelled as polyurethane foam. The soft tissue including muscles, fat and skin was modelled as a single bulk elastic material.

In a human anatomic structure, all layers appear more or less tied to their neighbours by network of infiltrating collagen fibres (although some can be identified more easily than others). Hence, epidermis, dermis, hypodermis, muscles, tendons and bones are all embedded in a same continuous collagenous network. Therefore, a tie constraint was defined between a lateral surface of the bone and muscles. However, a lower part of the bone was not tied to the muscle to present a result of the amputation process. Meanwhile, the filler and the soft tissue were tied to each other in order to obtain integrity between them during the negative pressure application. A drape on top of the filler is neglected to simplify the geometry.

A choice of the element type for finite-element meshing procedure depends on a type of analysis to be performed and an expected level of precision [18].Nodes of the model were arranged according with the most stressed regions expected. A 10-node quadratic tetrahedron (C3D10) element type was defined, as shown in Figure 1.

The model aims to simulate application of the negative pressure wound therapy in a trans-femoral amputee. At this stage, a fluid removal process was neglected in order to create a more simple approach. Different negative pressure levels of 70 mmHg, 110 mmHg, 125 mmHg and 150 mmHg were applied within the wounded tissue area and the top of the filler.

2.2 Model of pressure distribution in residuum

The finite element model was composed by four parts: resected bones (tibia and fibula), soft tissue, liner and prosthetic socket. While the residuum geometry came from segmentation process, the prosthetic components were designed individually. The liner was modelled with a

digital wrapping technique resulting in a constant wall thickness equal to 6 mm to simulate a market available product [Iceross Comfort - Össur's Reykjavik Iceland].

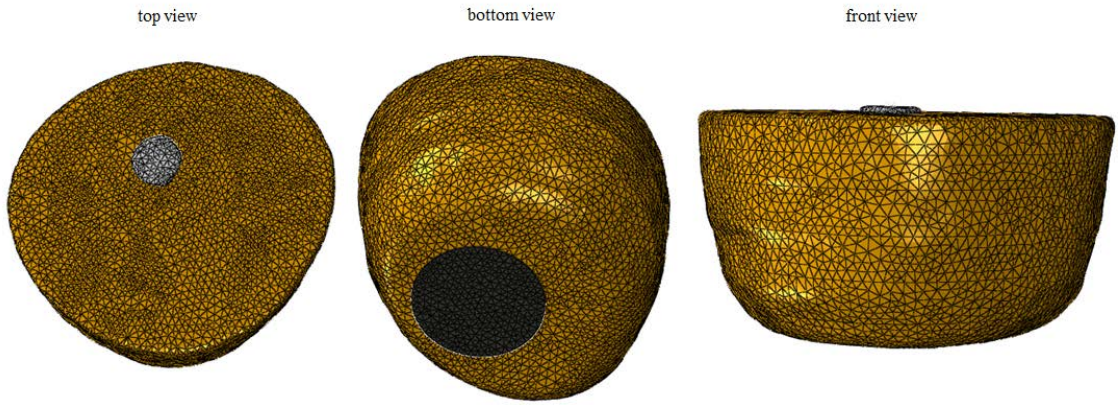


Figure 1: Top, bottom and front views of meshed geometry used in NPWT model

The same approach was followed in prosthetic socket modelling. The bottom end of the socket was combined with a disk having 50 mm height and 60 mm diameter aligned with tibial axis for socket/prosthetic limb adapter. Bones and socket parts were assumed to be linear-elastic, homogeneous and isotropic having Young's moduli of 15GPa and 1.3GPa respectively and Poisson ratio of 0.3. Residuum soft tissues and liner materials used in the analysis were assumed to be hyper-elastic, homogeneous and isotropic. An Ogden's strain-energy-function model was applied to the soft tissues and the liner in order to model the nonlinear behaviour of isotropic rubber-like material behaviour [19].

Coefficients of the Ogden's model describing hyper-elastic mechanical behaviour of the soft tissues were evaluated by fitting the experimental stress-strain curve of passive muscle behaviour investigated by Calvo *et al*. [20]. The prosthetic liner was assumed made of silicon rubber according with the manufacturer's description. The Ogden's model was then applied fitting the stress-strain curve based on experimental data shown in Sanders *et al*. [21].

All the domains were free-meshed with 10-node quadratic tetrahedral elements type (C3D10). A pin encastre boundary condition (U1=U2=U3=0) was applied on the tibia plate, while a single vertical force equal to a half body weight (370 N) was applied on the lower surface of the socket/prosthetic limb adapter simulating a stance scenario. The main features of the model are graphically summarised in Figure 2.

2.3 Model of bone-adaptation process

A 2D finite-element model of a transtibial amputation was developed based on the geometry obtained from the CT scan. The bone-remodelling process was simulated caused by an osseointegrated prosthetic implant. There are three main components in the model: a cortical-bone region modelled as elastic materials, an osseointegrated implant and a bone-remodelling region, which permits the adaptation of the materials as a result of mechanical loading (Figure 3). The shape and dimensions of the model were based on the cross-sectional area obtained from the previous model (see Section 2.2). An osseointegrated implant was fitted in a tibia's medullary cavity and modelled as analytical rigid body. The material properties of cortical bone such as its Young's modulus and Poisson's ratio were taken from the literature [22]. Both cortical- and virtual-bone domains were meshed with a CPS4R element type.

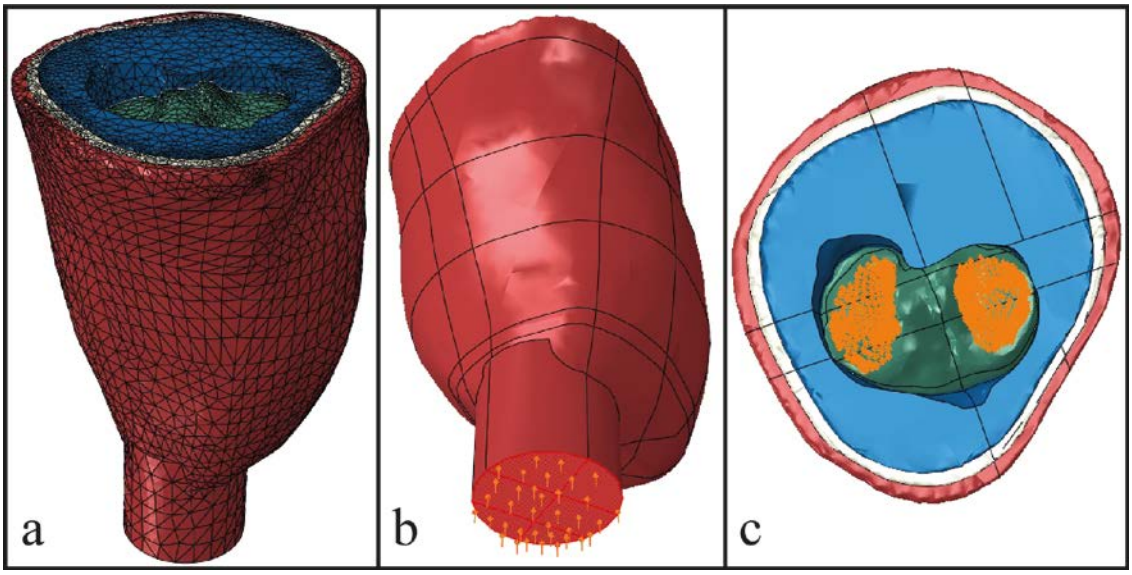


Figure 2: Finite-element model of transtibial limb and prosthesis:(a) model mesh; (b) loading condition; (c) boundary condition

In order to investigate the effect of different loading conditions on the process of bone remodelling related to osseointegration, three loading conditions were simulated: (a) a normal walking condition, with the load applied in both transverse and vertical directions; (b) a transverse loading condition, and (c) a transverse loading condition with 40% load reduction. The data were obtain from the work of Kutzner *et al.* [23]. 1/8 of the peak load was applied in the current model, in order to convert the force obtained from 3D to 2D model. The forces were applied on the bottom of the prosthesis. The tibia's top surface was fixed.

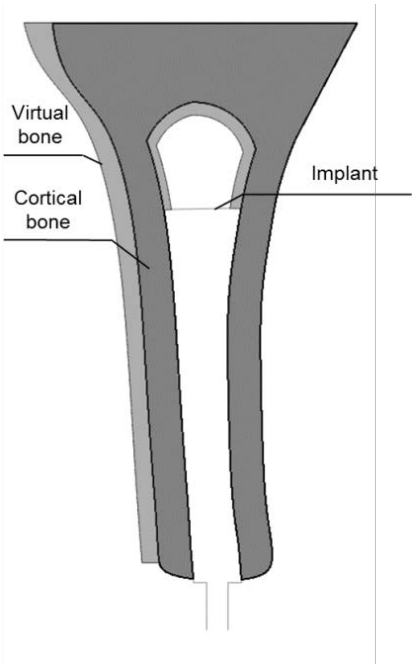


Figure 3: Two-dimensional FE model of transtibial osseointegration with implant and virtual bone-growth domain

In this study, a cubic relationship between the Young's modulus and bone density was in-

troduced for the bone-remodelling region according to the theory of Carter and Hayes [24]. Furthermore, based on the remodelling theory of Weinans *et al*. [25], the density change in the bone tissue was in accordance with the strain-energy density. An ABAQUS user subroutine USDFLD was developed to implement the remodelling algorithm. After completing the bone-adaptation algorithm, elements with zero density were removed from the region.

3 RESULTS AND DISCUSSION

In this section the results of three different models are presented and discussed separately.

3.1 NPWT model

The main goal of this model was to advance our understanding of the mechanical effects of NPWT on healing and investigate a soft-tissue response to these effects propagating thorough the wounded tissue during the therapeutic application. The obtained results indicated that an increasing magnitude of negative pressure in the therapy increased levels of stresses and deformations inside the muscle tissue. Figure 4 shows the computed stress distributions inside the tissue for a selected sagittal plane of the model. To facilitate visualization of the soft tissue area, the bone and the filler were removed from the graph. Maximum stresses were localised at the interface between the filler and the foam. Unsurprisingly, increasing the level of negative pressure had an increasing effect on the interface between the bottom part of the amputee's bone and the soft tissue. Maximum deformation occurred around the interface between the filler and the tissue (Fig. 5) and increasing negative-pressure levels caused higher deformation around the tissue-filler interface.

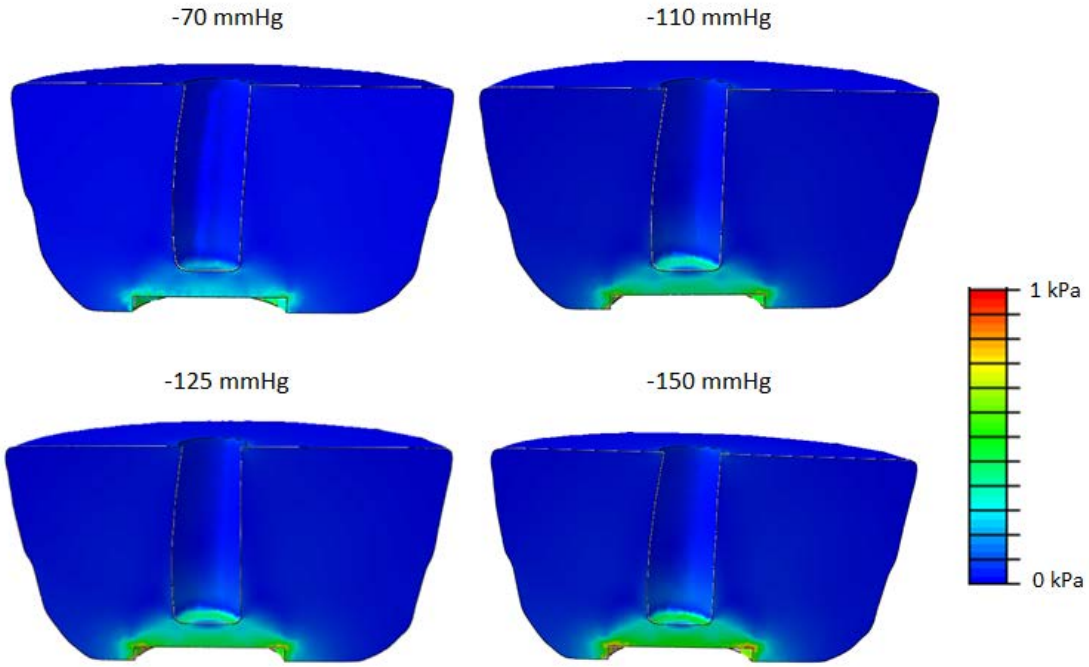


Figure 4: Comparison of distributions of von-Mises stress in sagittal plane of soft tissue for different levels of negative pressure

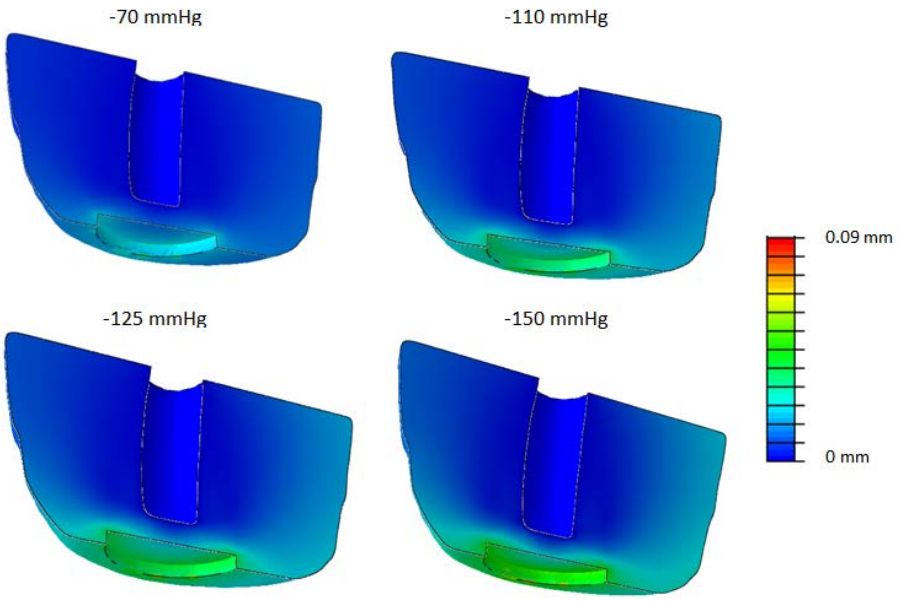


Figure 5: Comparison of deformations in sagittal plane of soft tissue for different levels of negative pressure

3.2 Model of pressure distribution in residuum

Numerical analysis of distribution of pressure (normal stress) was performed for the stump's internal soft tissue and at the interface between the residuum and the liner. A peak of pressure of 65 kPa was found deep in the soft tissues surrounding the inferior-anterior part of the tibia bone (Figure 6a). On the interface between the residuum and the liner 50 kPa maximum pressure was observed at the inferior-anterior part of the soft tissue (Figure 6b) as a result of the same loading condition. The results obtained are in agreement with a previous study [26] showing that stresses were concentrated in the flap tissue under the tibia's end and, therefore, a further study is needed on the clinical relevance in identifying risk factors of developing a deep-tissue injury (DTI) and other complications.

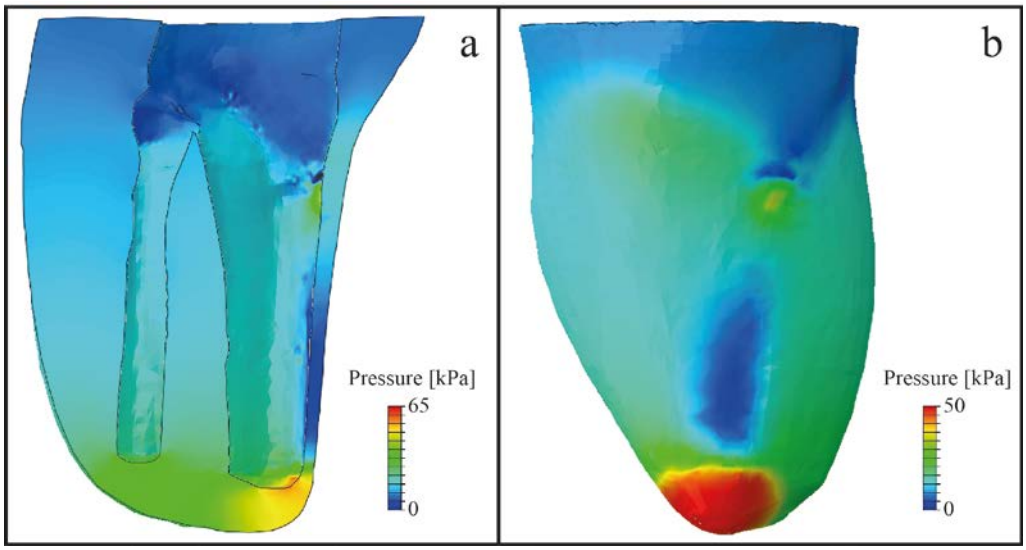


Figure 6: Pressure distribution in residuum: (a) cross-section view of residuum deep tissue; (b) anterior view of residuum-prosthesis interface.

3.3 Model of bone-adaptation process

Three different shapes of bone-growth patterns were predicted for different loading conditions: normal walking, transverse loading and reduced transverse loading (by 40%) (Figure 7). For the first conditions, the bone growth was predicted mostly for the top area of the left edge and the inner surface of the medullary cavity of the tibia. For the transverse loading condition, the bone growth on the left edge decreased significantly compared with the normal walking condition, with the shape of growth on the upper surface of the medullary cavity shifted as well. In case of reduced transverse loading, after the bone adaptation process, bone formed only at the top surface of the medullary cavity.

Strain-energy distributions for the normal walking condition before and after the remodelling process were also investigated in this study (Figure 8). The peak value of strain energy was 0.23 mJ, located near the osseointegrated implant before the bone-remodelling process. After the process, the peak strain energy reduced to 0.11 mJ. Comparison of strain energy before and after the remodelling process at the top corner of the medullary cavity demonstrted a reduction of strain energy from 0.187 mJ to 0.017 mJ.

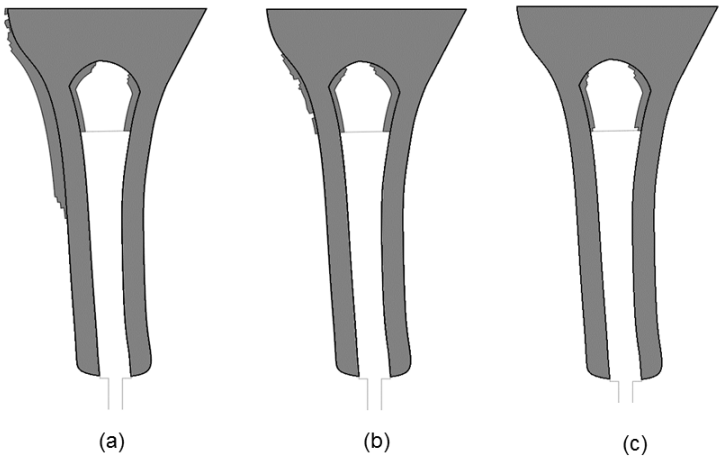


Figure 7: Predicted bone formation patterns for different loading conditions: (a) normal walking (load in both axial and transverse directions); (b) transverse loading; (c) reduced transverse loading

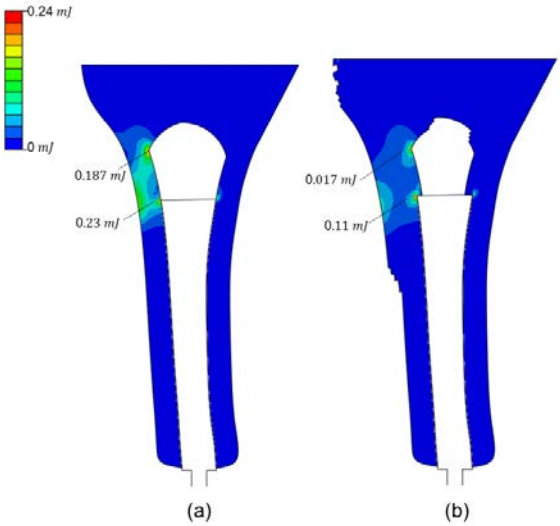


Figure 8: Strain-energy distribution for normal walking condition before (a) and after (b) remodelling process

4 CONCLUSIONS

Three developed models provided initial understanding of biomechanical processes within the residual limb that is often not possible to achieve otherwise. The major advantage of this study is that advanced computational models provided intuitive interpretation of detailed mechanical behaviour of the biological tissues at various stages (healthy, diseased and traumatic conditions) and for several areas of biomedical applications (injury prevention, wound care and rehabilitation). With a repeated use of the model, various potential scenarios can be analysed, making it much easier to optimize the studied processes.

REFERENCES

[1] S. Li, E. Demirci, and V.V. Silberschmidt, “Variability and anisotropy of mechanical behaviour of cortical bone in tension and compression.” J. Mech. Behav. Biomed., vol. 21, pp.109-120, 2013.

[2] A.A. Abdel-Wahab, K Alam, VV Silberschmidt, “Analysis of anisotropic viscoelastoplastic properties of cortical bone tissues.” J. Mech. Behav. Biomed., vol. 4 no. 5, pp. 807-820, 2011.

[3] S. Li, A.A. Abdel-Wahab, V.V. Silberschmidt, “Analysis of fracture processes in cortical bone tissue.” Eng. Frac. Mech., vol. 110, pp. 448-458. 2012.

[4] S. Li, A.A. Abdel-Wahab, E. Demirci, V.V. Silberschmidt, “Penetration of cutting tool into cortical bone: Experimental and numerical investigation of anisotropic mechanical behaviour.” J. Biomech., vol. 47, no. 5, pp. 1117–1126, 2014.

[5] NHS. 2015. Amputation. [ONLINE] Available at: http://www.nhs.uk/conditions/ amputation/pages/introduction.aspx. [Accessed 15th March 2016]

[6] D. G. Armstrong and L. A. Lavery, “Negative pressure wound therapy after partial diabetic foot amputation: a multi-centre randomised controlled trial.” Lancet (London, England), vol. 366, no. 9498, pp. 1704–10, 2005.

[7] D. Hinck, A. Franke, and F. Gatzka, “Use of vacuum-assisted closure negative pressure wound therapy in combat-related injuries--literature review.” Mil. Med., vol. 175, no. 3, pp. 173–181, 2010.

[8] K. Richter and B. Knudson, “Vacuum-Assisted Closure Therapy for a Complicated, Open, Above-the-Knee Amputation Wound.” J. Am. Osteopath. Assoc., vol. 113, no. 2, pp. 174–176, Feb. 2013.

[9] K. Couch and A. Stojadinovic, “Negative-pressure wound therapy in the military: Lessons learned.” Plast. Reconstr. Surg., vol. 127, pp. 117–130, 2011.

[10] V. Wong and S. Akaishi, “Pushing back: wound mechanotransduction in repair and regeneration.” J. Invest. Dermatol. vol. 131, pp. 2186–2196, 2011.

[11] E. Shweiki and K. E. Gallagher, “Negative pressure wound therapy in acute, contaminated wounds: documenting its safety and efficacy to support current global practice.” Int. Wound J., vol. 10, no. 1, pp. 13–43, 2013.

[12] S. Mendez-Eastman, “Guidelines for using negative pressure wound therapy.” Adv. Skin Wound Care, vol. 14, no. 6, pp. 314–323, 2001.

[13] J. E. Sanders, E. L. Rogers, E. A. Sorenson, G. S. Lee, D. C. Abrahamson. “CAD/CAM trans-tibial prosthetic sockets from central fabrication facilities: how accurate are they?” J. Rehabil. Res. Dev. vol. 44, no. 3, pp. 395–405, 2007.

[14] J. E. Sanders, M. R. Severance, K. J. Allyn, “Computer-socket manufacturing error: how much before it is clinically apparent?” J. Rehabil. Res. Dev., vol. 49, no. 4,

pp. 567–582, 2012.
[15] R. Brånemark, P. Brånemark, B. Rydevik, and R. R. Myers, “Osseointegration in skeletal reconstruction and rehabilitation: A review.” Journal of Rehabilitation Research and Development, vol. 38, no. 2, pp. 175–181, 2001.
[16] D. Taylor, “Fracture and repair of bone: A multiscale problem.” J. Mater. Sci., vol. 42, no. 21, pp. 8911–8918, 2007.
[17] K. Engel, R. Herpers, and U. Hartmann, “Biomechanical Computer Models.” Elsevier, 2011.
[18] W. Maurel, D. Thalmann, Y. Wu, and N. M. Thalmann, Biomechanical “Models for Soft Tissue Simulation.” Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.
[19] R. W. Ogden, “Large deformation isotropic elasticity – On the correlation of theory and experiment for incompressible rubberlike solids.” Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, vol. 326, no. 1567, pp. 565–584, 1972.
[20] B. Calvo, E. Peña, M.A. Martínez, M. Doblaré, “An uncoupled directional damage model for fibered biological soft tissues. Formulation and computational aspects.” Int. J. Numer. Meth. Eng., vol. 69, pp. 2036–2057, 2007.
[21] J.E. Sanders, J.M. Greve, S.B. Mitchell, S.G. Zachariah, “Material properties of commonly-used interface materials and their static coefficient of friction with skin and socks.” J. Rehab. Res. Dev., vol. 35, pp. 161–176, 1998.
[22] D. T. Reilly and A. H. Burstein, “The mechanical properties of cortical bone.” J. Bone Jt. Surg., vol. 56, pp. 1001–1022, 2011.
[23] Kutzner, B. Heinlein, F. Graichen, A. Bender, A. Rohlmann, A. Halder, A. Beier, and G. Bergmann, “Loading of the knee joint during activities of daily living measured in vivo in five subjects.” J. Biomech., vol. 43, no. 11, pp. 2164–2173, 2010.
[24] D. R. Carter and W. C. Hayes, “Bone compressive strength: the influence of density and strain rate.” Science, vol. 194, no. 4270, pp. 1174–1176, 1976.
[25] H. Weinans, R. Huiskes, H.J. Grootenboer, “The behavior of adaptive bone-remodeling simulation models.” J. Biomech., vol. 25, no. 12, pp. 1425–1441, 1992.
[26] S. Portnoy, Z. Yizhar, N. Shabshin, Y. Itzchak, A. Kristal, Y. Dotan-Marom, I. Siev-Ner, A. Gefen, “Internal mechanical conditions in the soft tissues of residual limb of a trans-tibial amputee.” J. Biomech., vol. 41, pp. 1897-1909, 2008.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ANALYSIS OF FIBRE BRAGG GRATINGS SENSORS OPTIMAL PLACEMENT FOR MONITORING OF DAMAGE PROPAGATION IN LAMINATE COMPOSITES

Mikhail A. Tashkinov[1]

[1] Perm National Research Polytechnic University
Komsomolsky Ave., 29, Perm 614990, Russia
e-mail: m.tashkinov@pstu.ru

Keywords: Numerical models, laminate composites, damage modes, delamination, FBG sensors.

Abstract. *Composite materials are widely used in a broad field of applications due to their unique properties. However, development and propagation of inherent microstructural defects occurred in composites during their manufacturing or exploitation can be critical for overall performance of structural elements. Failure of composite structures can be predicted by creation of advanced monitoring systems utilizing sensor elements. One of the promising technologies is based on the optical fibre sensors with Bragg gratings, which are being embedded between the plies and can measure the changes of internal mechanical state as well as some physical properties. In order to be able to capture the growth of the defects, it is necessary to evaluate the distance between a defect and the sensor at which the latter can measure the respective state changes. This can be done by preliminary numerical modelling of defects behaviour. The scope of this work is numerical studies of the delamination growth in GFRP composite plate subjected to compression load using finite element analysis. Virtual crack closure technique was used as a computational instrument for delamination modelling. Evaluation of the damage effect on the internal strain fields was performed. Recommendation for FBG sensors positions were developed basing on specifics of strain field changes in the plies.*

1 INTRODUCTION

Over the past decades, the aircraft industry has significantly expanded the scope of composite materials, including polymer-based composites. The advantages of composites over traditional materials are high strength to weight ratio, high stability against fatigue loads and corrosion, possibility of manufacturing of complex-shaped parts with tailored properties. On the other hand, compared to traditional materials (aluminium and titanium alloys) applied in aircrafts, one of the factors restricting the application of composites in this area is a rather high sensitivity of these materials to microstructure damage including delamination, fibre rupture, fluid absorption, impact damage, matrix fracture, decrease in strength and stiffness at high temperatures, etc., produced by operational factors. The necessity of early detection of such defects leads to the requirement of the efficient control of the current state of structures and further maintenance for extension of their operational life.

The prospective approach for monitoring of composite structures during exploitation is connected with creation of smart-materials and smart-systems with the sensor elements, which can significantly improve the safety and reduce the maintenance costs [1]. The solution was found with the development of the technologies based on the optical fibre sensors with Bragg gratings (fibre Bragg gratings, or FBG sensors) [2-4]. Such sensors can be embedded between the layers of the material during its manufacturing and are able to measure changes of magnetic, strain and temperature fields. The sensors allow to obtain data, which, subject to the further analysis, can be used for diagnostics of damage initiation as well as for prediction of damage progression and residual life of structures.

One of the main questions is how the FBG sensors should be placed in composite structures in order to be able to register changes in mechanical and physical parameters in critical zones. Within this scope, the important task is to develop configurations of the FBG sensors network for specific composite structures under operating loading conditions.

The aim of this work is to develop recommendations concerning optimal placement of FBG sensors basing on numerical modelling of mechanical behaviour of composite laminates.

2 MONITORING OF COMPOSITES USING FBG SENSORS

2.1 Optical fibres with Bragg gratings sensors

One of the important challenges in non-destructive testing of composite structures nowadays is development of instruments which would allow continuous monitoring of structures' conditions during exploitation for timely detection of occurred damage and prevention of its propagation. The popular solution is to construct a structural health monitoring (SHM) system based on optical fibres with Bragg gratings. Fibre Bragg grating is formed by inducing a periodic modulation of the refractive index in the core of a single mode optical fibre [5, 6]. Thus, FBG filters a broadband light, which is transmitted into the fibre core, reflecting light at a single wavelength [7, 8]. The form of FBG is changing subject to external micro-mechanical loads and so does the grating spectral response. The direct application of such feature is strain measurement. The relations for transformation of the light spectrum data into mechanical characteristics are presented, for instance, in [8].

As a rule, optical fibre carries several Bragg gratings. This gives an opportunity to arrange a combination of sensing points with a single fibre, what is major advantage for creation of monitoring systems, integrated into materials.

By now, many works were devoted to investigation of possibilities of employing FBG sensors for detection of strain and temperature changes in laminate composite parts and structures [9]. Technological solutions allow the optical fibres to be embedded between the plies or

to be attached on the surface. The questions of operability of such embedded fibres were also discussed – it has been proved, that fibres can stand different regimes of composite manufacturing, including autoclave forming [10].

Recent computational and experimental studies showed that embedded fibres effect on initial mechanical properties of laminate composites can be considered negligible [10, 11], what is important conclusion for the further development of SHM systems for composite structures based on FBG sensors.

2.2 Detecting damage using FBG sensors

The important and high-promising field of FBG sensors application is prediction of damage and failure. One of the major failure mechanisms in laminated composite structures is delamination. There are many factors that can stimulate the delamination growth, both induced during manufacturing process and appeared during exploitation, such as impact damage. Delamination is often difficult to reveal since it occurs inside the structure. Thus, the promising solution is to detect such kind of damage using the internal sensor systems.

There is a large number of research works were devoted to studying possibility of delamination and other defects detection using embedded FBG sensors [12-16]. However, due to limited sensibility of sensors, the results strongly depend on the distance between the defect and the sensor. It is necessary to understand where exactly the optical fibres as well as Bragg gratings should be positioned in the structure in order to be able to capture the critical changes of mechanical state, caused by propagating damage. This is also important for development of cost-reducing practical solutions for implementation of the sensors scheme, which gives the accurate measurement results with less number of gratings.

Preliminary numerical modelling can be performed in order to assess the non-uniform strain field near defects as well as their growth. The purpose is to study the delamination growth between the skin layers to see if the modelled defect will cause any changes in the strain field, which could be directly measured by FBG sensors.

3 DELAMINATION MODELLING

3.1 Geometrical and material properties

The subject investigated in this work was laminate composite plate made of GFRP, which represents skins of sandwich panels. It had dimensions of 50x50mm and consisted of 14 plies with 0.3mm thickness, each oriented 0 degree. The normalized mechanical properties of plies are presented in Table 1.

Initial delamination was modelled as a centred circular section with 6mm radius. An additional layer with 2mm thickness and significantly reduced mechanical properties was introduced between 5^{th} and 6^{th} ply of this section.

Ex/Ez,,, GPa	Ey/Ez, GPa	Ez/Ez, GPa	νxy	νyz	νxz	Gxy/Ez, GPa	Gyz/Ez, GPa	Gxz/Ez, GPa
4.1	3.1	1	0.15	0.18	0.42	0.6	0.5	0.5

Table 1: Normalized mechanical properties of glass fibre/epoxy material.

The critical strain energy release rates were taken as presented in Table 2. The overall dimensions of the studied plate are schematically presented on Fig. 1. Compression load, equal to 45kN has been applied to the top edge of the panel.

G_{IC}, N/mm	G_{IIC}, N/mm	G_{IIIC}, N/mm
0.07	0.45	0.45

Table 2: Critical strain energy release rates.

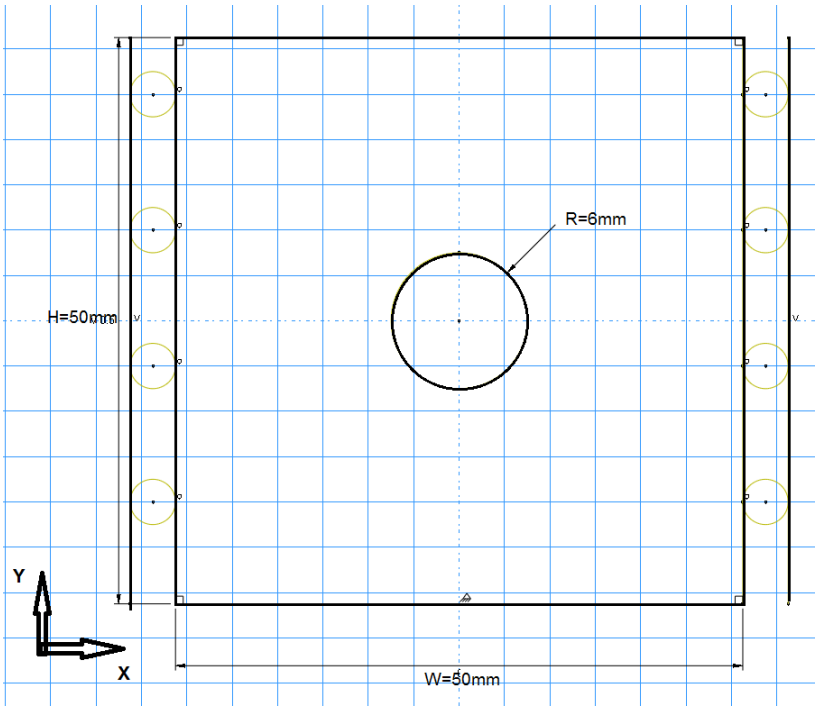


Figure 1: Graphical dimensions of the laminate panel.

3.2 Virtual crack closure technique

Delamination behaviour of laminated composites can be modelled using different concepts. One of the most commonly used approaches is the virtual crack closure technique (VCCT) [17], which assumes that the energy released when a crack is extended by a certain amount is the same as the energy required to close the crack. According to VCCT, the strain energy release rate can be calculated using the following relation [17]:

$$G = \frac{1}{2\Delta l}\int_0^{\Delta l} u(r)\,\sigma(r-\Delta l)\,dr\,, \tag{1}$$

where Δl is length at which crack is extended, r is a distance from the crack tip, $u(r)$ is corresponding displacements, σ is the stress. In general case, the crack will open when $G/G_C \geq 1$, where G is calculated equivalent strain energy release rate, G_C is critical equivalent strain energy release rate based on strength properties of the interface. The mixed-mode criteria can

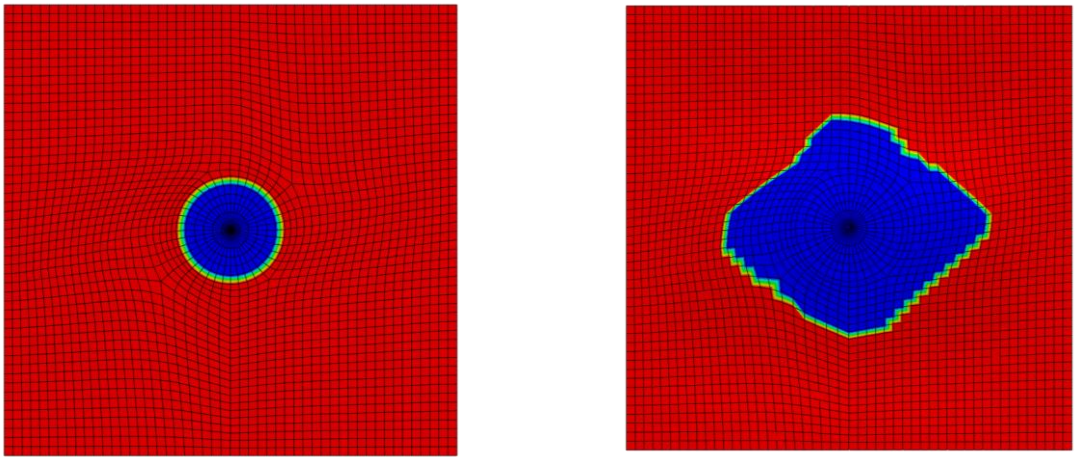
be used for calculation of critical equivalent release rate G_C. The Benzeggagh-Kenane model was used, which involves values for modes I, II, and III:

$$G_C = G_{IC} + \left(G_{IIC} - G_{IC}\right)\left(\frac{G_{II} + G_{III}}{G_I + G_{II} + G_{III}}\right)^n, \quad (2)$$

where $G_{IC,IIC,IIIC}$ and $G_{I,II,III}$ are, respectively, calculated and critical equivalent strain energy release rate for modes I, II and III.

4 NUMERICAL RESULTS AND DISCUSSION

The 3D finite element analysis was performed using shell elements. Compression loading leads to buckling of the laminate, which expedite the delamination process. The growth of the delamination between 5th and 6th layer is shown on Fig. 2.



a *b*

Figure 2: Growth of delamination under compression.

Each ply was analysed in order to assess the delamination effect on the strain fields. The modelled components ε_Y of strain fields on the top of the plies are presented on Fig. 3 for the plies, closest to delamination.

The analysis has shown that induced delamination affect only strain fields of the nearest plies. With distancing from the delamination, the correlation between delamination and strain is becoming much less obvious due to contribution of interaction of plies between each other. This means that to detect the growing defect using direct strain field measurement, the FBG sensors need to be positioned in the vicinity of defect, not farther than 2 or 3 plies distance from the delamination.

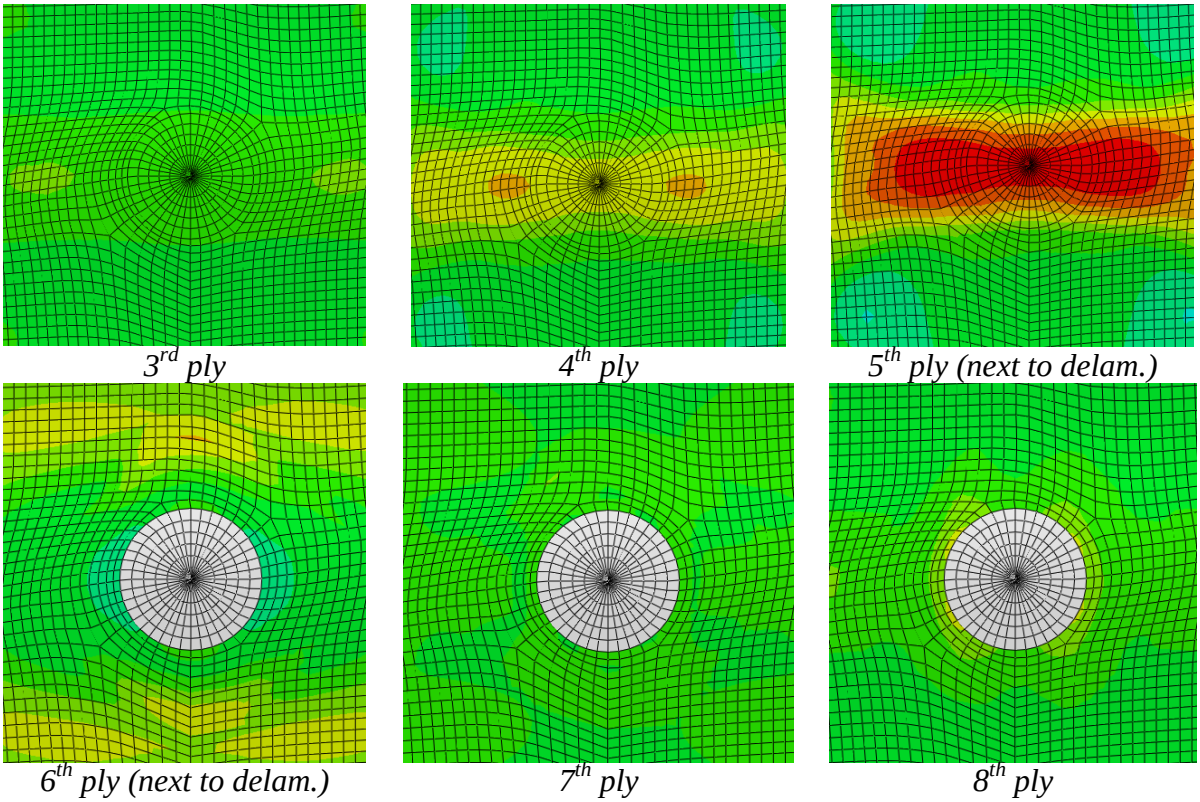


Figure 3: Strain fields in plies on the different distance of delamination. The maximum ε_Y logarithmic strain value is 19×10^{-4}, the minimum is -15×10^{-4}.

The changing of strain field in the nearest to delamination ply in dependence of growing compression load is shown on Fig. 4. For the studied case, the strain field homogeneity was influenced on the distance of approximately three radius of the initial delamination. At the same time, the delamination brings inhomogeneity to the strain field of the nearest ply at loading values between 20 and 30 kN, which means that correctly placed FBG sensors can detect the early stage of its growth. It should be noted that Bragg gratings can output the incorrect data if they will be affected by debonding process.

In order to avoid dense sensor network, monitoring of delamination in complex composite structures using mechanical characteristics, interpreted from FBG sensors data, can be reliable only for preliminarily determined critical zones, where possibility of defect emergence is predicted. Further studies are required for developing of effective FBG sensors-based monitoring system, which would be able to response on accidentally caused damage or manufacturing-induced defects.

There are also several other approaches, connected with damage detection irrelatively of direct displacements and strain measurement. For instance, Takeda et al. [18], proposed approach that is aimed to find a correlation between the obtained form of spectrum, reflected from Bragg gratings, as well as the intensity ratio of spectrum peaks to the delamination length. In practical implementation, the difficulty of such method is distortion of the spectrum caused by inhomogeneity of the strain field near the defect. This could be tackled by using multiplexed short FBG sensors, which allow to restore quasi-continuous strain profiles along the sensor [18, 19].

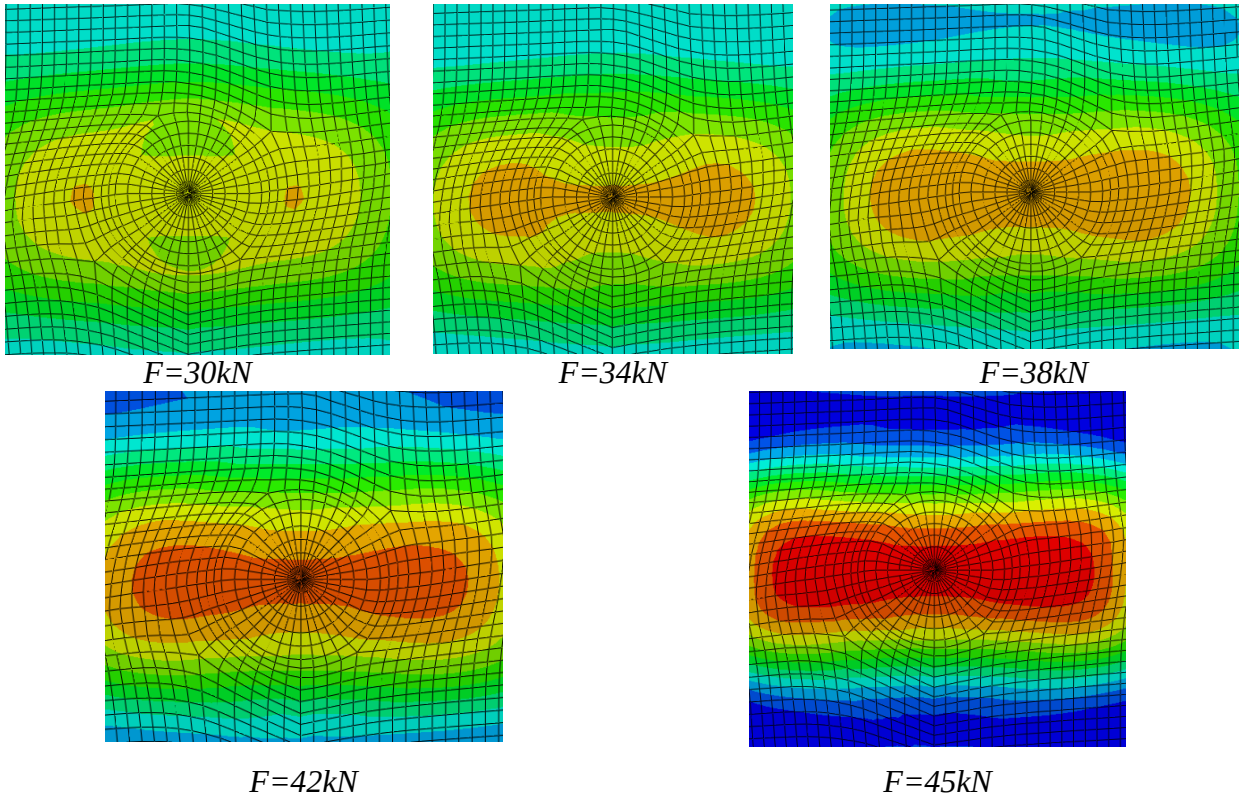


Figure 4: Strain field changing in the 6th ply, in dependence on applied compression force F. The maximum ε_Y logarithmic strain value is 19×10^{-4} , the minimum is -15×10^{-4} .

5 CONCLUSIONS

- The problem of the delamination detection in composite laminate structures using the embedded optical fibres sensors with Bragg gratings was examined.
- Modelling of delamination in composite laminate panel was performed using VCCT in order to analyse the effect of defect growth on the strain field in the internal plies.
- Some conclusions were made regarding the optimal placement of the fibres which would allow to detect the damage propagation using the mechanical data interpreted from FBG sensors.
- The obtained numerical results can be used further to control the adequacy of experimental data collected from embedded FBG sensors.

ACKNOWLEDGEMENTS

The research was performed at the Perm National Research Polytechnic University, with the support of the Russian Science Foundation (project №15-19-00243). Computations were performed at Institute of Continuous Media Mechanics of Russian Academy of Science.

REFERENCES

[1] Daniel F.O. Braga, S.M.O. Tavares, Lucas F.M. da Silva, P.M.G.P. Moreira, Paulo M.S.T. de Castro, Advanced design for lightweight structures: Review and prospects,

Progress in Aerospace Sciences, **69**, 29-39, 2014, http://dx.doi.org/10.1016/j.paerosci.2014.03.003.

[2] K.I. Tserpes, V. Karachalios, I. Giannopoulos, V. Prentzias, R. Ruzek, Strain and damage monitoring in CFRP fuselage panels using fiber Bragg grating sensors. Part I: Design, manufacturing and impact testing, *Composite Structures*, **107**, 726-736, 2014, http://dx.doi.org/10.1016/j.compstruct.2013.09.053.

[3] Jeannot Frieden, Joël Cugnoni, John Botsis, Thomas Gmür, Dragan Ćorić, High-speed internal strain measurements in composite structures under dynamic load using embedded FBG sensors, *Composite Structures*, **92** (8), 1905-1912, 2010, http://dx.doi.org/10.1016/j.compstruct.2010.01.007.

[4] Fabiano Colpo, Laurent Humbert, John Botsis. Characterisation of residual stresses in a single fibre composite with FBG sensor, *Composites Science and Technology*, **67**, 1830–1841, 2007.

[5] Jian Cai, Lei Qiu, Shenfang Yuan, Lihua Shi, PeiPei Liu and Dong Liang. *Structural Health Monitoring for Composite Materials*, *Composites and Their Applications*, Prof. Ning Hu (Ed.), InTech, 2012.

[6] Loayssa A. Optical Fiber Sensors for Structural Health Monitoring. In: Mukhopadhyay S. (ed.) *New Developments in Sensing Technology for Structural Health Monitoring*. Heidelberg: Springer-Verlag; 335-358, 2011.

[7] Ramzyzan Ramly, Wahyu Kuntjoro, Mohd. Kamil Abd. Rahman, Using Embedded Fiber Bragg Grating (FBG) Sensors in Smart Aircraft Structure Materials, *Procedia Engineering*, **41**, 600-606, 2012, http://dx.doi.org/10.1016/j.proeng.2012.07.218.

[8] L. Sorensen, *The response of embedded FBG sensors to nonuniform strains in CFRP composites during processing and delamination*, EPFL, 2007.

[9] KSC Kuang, WJ Cantwell, Use of conventional optical fibers and fiber Bragg gratings for damage detection in advanced composite structures: A review, *Appl Mech Rev*, **56**, (5), 493-514, 2003.

[10] Guemes, J.A., Mcnendez, J.M., Frovel, M., Fernandez, L, Pintado, J.M.. Experimental analysis of buckling in aircraft skin panels by fibre optic sensors. *Smart Materials & Structures*, **10**(3), 490-496, 2001.

[11] Carrion. G., Frovel, M., Pintado. J.M., Cabrerizo, F.. Extensive characterisation of the intrusiveness of fibre Bragg grating sensors embedded in a carbon fibre reinforced polymer. In: *4th European Workshop on Structural Health Monitoring*, 2008.

[12] E. Farmand-Ashtiani, J. Cugnoni, J. Botsis, Monitoring and characterization of the interfacial fracture in sandwich composites with embedded multiplexed optical sensors, *Composite Structures*, **96**, 476-483, 2013, http://dx.doi.org/10.1016/j.compstruct.2012.08.059.

[13] G. Pereira, C. Frias, H. Faria, O. Frazão, A.T. Marques, On the improvement of strain measurements with FBG sensors embedded in unidirectional composites, *Polymer Testing*, **32**(1), 99-105, 2013, http://dx.doi.org/10.1016/j.polymertesting.2012.09.010.

[14] Xuefeng Zhao, Jihua Gou, Gangbing Song, Jinping Ou, Strain monitoring in glass fiber reinforced composites embedded with carbon nanopaper sheet using Fiber Bragg Grating (FBG) sensors, *Composites Part B: Engineering*, **40**(2), 134-140, 2009, http://dx.doi.org/10.1016/j.compositesb.2008.10.007.

[15] Christopher K.Y. Leung, Zhenglin Yang, Ying Xu, Pin Tong, Stephen K.L. Lee, Delamination detection in laminate composites with an embedded fiber optical interferometric sensor, *Sensors and Actuators A: Physical*, **119**(2), 336-344, 2005, http://dx.doi.org/10.1016/j.sna.2004.10.007.

[16] Prabhugoud. M.. Peters, K.. Efficient simulation of Bragg grating sensors for implementation to damage identification in composites. *Smart Materials & Structures*, **12**(6), 914-924, 2003.

[17] Rybicki, E.F., Kanninen, M.F., A Finite Element Calculation of Stress Intensity Factors by a Modified Crack Closure Integral, *Engineering Fracture Mechanics*, **9**, 1977.

[18] S. Takeda. S. Minakuchi, Y. Okabe, and N. Takeda. Delamination monitoring of laminated composites subjected to low-velocity impact using small-diameter FBG sensors. *Composites Part A-Applied Science and Manufacturing*, **36**, 903-908, 2005.

[19] D. S. Canovas, *Advances in fibre Bragg grating sensors for damage detection in composite laminates: application in quasi-static and fatigue delamination tests*, Universitat de Girona, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SEVEN DIFFERENT WAYS TO MODEL VISCOELASTICITY IN A GEOMETRICALLY EXACT SETTING

Alexey V. Shutov[1] [2]

[1]Novosibirsk State University
Pirogova 2, 630090, Novosibirsk, Russia
e-mail: alexey.v.shutov@gmail.com

[2] Lavrentyev Institute of Hydrodynamics
pr. Lavrentyeva 15, 630090, Novosibirsk, Russia
e-mail: alexey.v.shutov@gmail.com

Keywords: Finite strains, viscoelasticity, Maxwell fluid, weak invariance, volumetric-isochoric split.

Abstract. *Seven different approaches to the phenomenological finite-strain viscoelasticity are analyzed. The focus of the analysis is put on the Maxwell fluid, due to its paramount importance for phenomenological material modeling. The first modeling approach to the Maxwell fluid is based on the multiplicative decomposition of the deformation gradient, combined with hyper-elastic relations between stresses and elastic strains. The second approach employs the additive decomposition of the strain rate tensor in combination with hypoelastic relations. The third approach makes use of the additive decomposition of the material Hencky strain. The fourth and fifth models are the covariant and contravaraint models of the Maxwell fluid, proposed by Haupt and Lion (2002). Finally, the sixth and the seventh approaches are the modifications of these two models, introduced here to enforce a pure volumetric-isochoric split. The approaches are analyzed with respect to their objectivity, thermodynamic consistency, fading memory, weak invariance under isochoric change of the reference configuration, and a possibility of a pure volumetric-iscohoric split.*

1 INTRODUCTION

Various phenomenological approaches to large strain viscoelasticity are analyzed in this talk. We put the main focus on the constitutive relations of the Maxwell body, since these relations act as the backbone for many advanced models of technical materials. Apart from viscoelastic applications, the generalization of the discussed phenomenological ideas to viscoplasticity is in many cases obvious and straightforward. Nowadays, the small-strain formulation of the Maxwell body which is based on the classical Prandtl-Reuss decomposition of the infinitesimal strain tensor is widely accepted by the modeling community [1]. Within this model, the infinitesimal strain tensor $\boldsymbol{\varepsilon}$ is decomposed into the inelastic part $\boldsymbol{\varepsilon}_{\mathrm{i}}$ and the elastic part $\boldsymbol{\varepsilon}_{\mathrm{e}}$

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_{\mathrm{i}} + \boldsymbol{\varepsilon}_{\mathrm{e}}. \tag{1}$$

The infinitesimal stress tensor $\boldsymbol{\sigma}$ is computed using the isotropic Hooke law

$$\boldsymbol{\sigma} = k\ \mathrm{tr}(\boldsymbol{\varepsilon}_{\mathrm{e}})\ \mathbf{1} + 2\ \mu(\boldsymbol{\varepsilon}_{\mathrm{e}})^{\mathrm{D}}, \tag{2}$$

where $k \geq 0$ and $\mu > 0$ are the bulk and shear moduli of the material, respectively; $\mathbf{1}$ stands for the identity tensor; $(\cdot)^{\mathrm{D}}$ is the deviatoric part of a tensor.. The evolution of the inelastic strain is given by the flow rule with an initial condition at $t = t_0$

$$\dot{\boldsymbol{\varepsilon}}_{\mathrm{i}} = \frac{1}{\eta}\,\boldsymbol{\sigma}^{\mathrm{D}}, \quad \boldsymbol{\varepsilon}_{\mathrm{i}}|_{t=t_0} = \boldsymbol{\varepsilon}_{\mathrm{i}}^0, \quad \mathrm{tr}(\boldsymbol{\varepsilon}_{\mathrm{i}}^0) = 0. \tag{3}$$

Here, $\mathrm{tr}(\cdot)$ denotes the trace of a tensor, $\eta > 0$ is the viscosity of the material.

At the same time, there is a big variety of different generalizations of this model to the geometrically exact setting of large deformations. Seven different generalizations of these equations are considered in the paper. The generalizations differ in the way how the total deformation is decomposed into the elastic and inelastic parts (multiplicative decomposition of the deformation gradient vs. additive decomposition of a generalized strain or strain rate), how the elastic properties are modeled (hyperelasticity vs. hypoelasticity), which stress tensors determine the evolution of the inelastic deformations (Mandel stress vs. weighted Cauchy stress or others). Other generalizations of the Maxwell body employ material or objective time derivatives.

Interestingly, a number of approaches was discovered and then rediscovered by other researchers. The situation is even more complicated since for several large strain approaches there is a number of equivalent formulations, and different authors were not aware of that equivalence. Thus, it is very important to analyze the links between these formulations, since for some of the formulations, simple, robust, and efficient numerical algorithms are already available. After the links are established, these algorithms can be easily adjusted to cover the equivalent formulations (as was carried out in [2] for the multiplicative inelasticity). In this paper, the seven different approaches are classified with respect to the following criteria: objectivity, thermodynamic consistency, w-invariance [3], pure decomposition of the stress response in the volumetric and the isochoric parts. New modifications of covariant and contravariant Maxwell models are proposed in order to enforce the desired volumetric-isochoric split.

2 SOME GENERAL CLASSIFICATION CRITERIA

Apart from the well-known properties like objectivity and thermodynamic consistency, some other classification criteria will be considered in this paper.

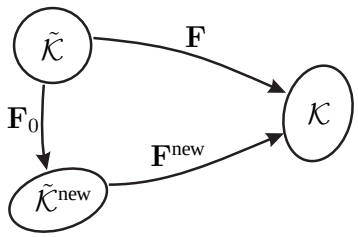


Figure 1: Commutative diagram, illustrating a change of the local reference configuration: $\tilde{\mathcal{K}}$ is replaced by a new reference $\tilde{\mathcal{K}}^{\text{new}}$.

2.1 Weak invariance under isochoric reference change

Let us consider a material model in the following form: The current value of the Kirchhoff stress tensor $\mathbf{S}$ depends on the local history of the deformation gradient $\mathbf{F}$ and a set of initial conditions $\mathcal{Z}_0$

$$\mathbf{S}(t) = \underset{t_0 \le t' \le t}{\mathbf{S}} \big(\mathbf{F}(t'), \mathcal{Z}_0\big). \tag{4}$$

Here, a general response functional of a simple material appears on the right-hand side; t_0 is the initial time instance.

Next, we consider a second-rank tensor $\mathbf{F}_0$, such that $\det(\mathbf{F}_0) = 1$. Let $\widetilde{\mathcal{K}}$ be the original reference configuration and $\widetilde{\mathcal{K}}^{\text{new}} := \mathbf{F}_0\widetilde{\mathcal{K}}$ be a new reference (cf. Figure 1). The corresponding new deformation gradient (also known as the relative deformation gradient) is obtained by the following push-forward

$$\mathbf{F}^{\text{new}}(t) := \mathbf{F}(t)\ \mathbf{F}_0^{-1}. \tag{5}$$

We say that the material model (4) is weakly invariant under the transformation (5) if there is a new set of initial data

$$\mathcal{Z}_0^{\text{new}} = \mathcal{Z}_0^{\text{new}}(\mathcal{Z}_0, \mathbf{F}_0) \tag{6}$$

such that the same stress response is predicted:

$$\underset{t_0 \le t' \le t}{\mathbf{S}} \big(\mathbf{F}(t'), \mathcal{Z}_0\big) = \underset{t_0 \le t' \le t}{\mathbf{S}} \big(\mathbf{F}^{\text{new}}(t'), \mathcal{Z}_0^{\text{new}}\big). \tag{7}$$

If a certain model is invariant under arbitrary isochoric changes of the reference configuration, we say that the model is *weakly invariant* or, shortly, *w-invariant*. From the mathematical standpoint, this w-invariance constitutes a certain (generalized) symmetry of the governing equations. As any other symmetry property, w-invariance provides additional insights into the structure of the constitutive equation.

2.2 Pure volumetric-iscohoric split

Suppose that (4) holds true. Let $\overline{\mathbf{F}}(t) := \det(\mathbf{F}(t))^{-1/3}\ \mathbf{F}(t)$ be the isochoric part of the deformation history $\mathbf{F}(t)$. We say that the material model (4) exhibits a pure volumetric-isochoric split (v-i split) with elastic volume changes, if the following three conditions are satisfied:

i:
$$\text{tr}\Big(\underset{t_0 \le t' \le t}{\mathbf{S}} \Big(\overline{\mathbf{F}}(t'), \mathcal{Z}_0\Big)\Big) \equiv 0, \ \text{whenever}\ \ \text{tr}\ \mathbf{S}|_{t=t_0} = 0;$$

ii: there is $\mathcal{Z}_0^{dev} = \mathcal{Z}_0^{dev}(\mathcal{Z}_0)$ such that

$$\Big(\underset{t_0 \le t' \le t}{\mathbf{S}} \big(\mathbf{F}(t'), \mathcal{Z}_0\big)\Big)^{\text{D}} \equiv \underset{t_0 \le t' \le t}{\mathbf{S}} \big(\overline{\mathbf{F}}(t'), \mathcal{Z}_0^{dev}\big);$$

iii: $\mathrm{tr}\big(\mathbf{S}(t)\big)$ is a function of the instant value $\det\big(\mathbf{F}(t)\big)$.

Remark 1: The essential part in obtaining a pure volumetric-iscohoric split is the condition i. Indeed, consider, for example, a model where the initial conditions are formulated with respect to the Kirchhoff stresses: $\mathcal{Z}_0 = \{\mathbf{S}|_{t=t_0}\}$. If the property i is satisfied for a certain model, then the properties ii and iii can be enforced by putting

$$\underset{t_0\le t'\le t}{\mathbf{S}}\big(\mathbf{F}(t'),\mathbf{S}|_{t=t_0}\big) := \underset{t_0\le t'\le t}{\mathbf{S}}\big(\overline{\mathbf{F}}(t'),(\mathbf{S}|_{t=t_0})^{\mathrm{D}}\big) + p\big(\det(\mathbf{F}(t))\big)\mathbf{1}, \tag{8}$$

where $p = \mathrm{tr}\,\mathbf{S}$ is a suitable function of the current $\det\mathbf{F}$.

Remark 2: Condition ii takes the simple form thanks to the appropriate choice of the stress measure. If the Kirchhoff stress $\mathbf{S}$ is replaced by the Cauchy (true) stress $\mathbf{T}$, condition ii can be recast in the equivalent form:

$$\Big(\underset{t_0\le t'\le t}{\mathbf{T}}\big(\mathbf{F}(t'),\mathcal{Z}_0\big)\Big)^{\mathrm{D}} \equiv \frac{1}{\det(\mathbf{F}(t))}\ \underset{t_0\le t'\le t}{\mathbf{T}}\big(\overline{\mathbf{F}}(t'),\mathcal{Z}_0^{dev}\big). \tag{9}$$

3 MULTIPLICATIVE DECOMPOSITION OF THE DEFORMATION GRADIENT COMBINED WITH HYPERELASTICITY

In this section we discuss the Maxwell model which is covered by general constitutive equations of Simo and Miehe [4]. The model is based on the multiplicative decomposition of the deformation gradient $\mathbf{F}$ into the viscous (dissipative) part $\mathbf{F}_{\mathrm{i}}$ and the elastic (conservative or energetic) part $\mathbf{F}_{\mathrm{e}}$

$$\mathbf{F} = \mathbf{F}_{\mathrm{e}}\,\mathbf{F}_{\mathrm{i}}. \tag{10}$$

This constitutive assumption generalizes (1) to the case of finite strains. Another important ingredient of the model is that the elastic properties are modeled using hyperelastic relations between stresses and elastic strains.

3.1 Formulation on the reference configuration

Let us summarize the constitutive equations of this model, formulated on the reference configuration (here we follow the presentation of Lion [5]). The total strain can be described with the right Cauchy-Green tensor $\mathbf{C} := \mathbf{F}^{\mathrm{T}}\mathbf{F}$. Its inelastic part is captured using the inelastic right Cauchy-Green tensor

$$\mathbf{C}_{\mathrm{i}} := \mathbf{F}_{\mathrm{i}}^{\mathrm{T}}\mathbf{F}_{\mathrm{i}}. \tag{11}$$

Let $\tilde{\mathbf{T}}$ be the second Piola-Kirchhoff stress tensor; ρ_{R} stands for the mass density in the reference configuration; $\psi(\mathbf{C}\,\mathbf{C}_{\mathrm{i}}^{-1})$ stands for the free energy per unit mass. The function $\psi(\cdot)$ is assumed to be an isotropic function of its argument. Moreover, we assume the following volumetric-isochoric split

$$\psi(\mathbf{C}\,\mathbf{C}_{\mathrm{i}}) = \psi_{\mathrm{iso}}(\overline{\mathbf{C}\,\mathbf{C}_{\mathrm{i}}^{-1}}) + \psi_{\mathrm{vol}}(\det(\mathbf{C}\,\mathbf{C}_{\mathrm{i}}^{-1})). \tag{12}$$

The closed system of constitutive equations is given by

$$\tilde{\mathbf{T}} = 2\rho_{\mathrm{R}}\frac{\partial\psi(\mathbf{C}\mathbf{C}_{\mathrm{i}}^{-1})}{\partial\mathbf{C}}\Big|_{\mathbf{C}_{\mathrm{i}}=\mathrm{const}},\quad \dot{\mathbf{C}}_{\mathrm{i}} = \frac{2}{\eta}\big(\mathbf{C}\tilde{\mathbf{T}}\big)^{\mathrm{D}}\mathbf{C}_{\mathrm{i}},\quad \mathbf{C}_{\mathrm{i}}|_{t=t_0} = \mathbf{C}_{\mathrm{i}}^0. \tag{13}$$

Here, the superimposed dot denotes the material time derivative.

3.2 Formulation on the current configuration

The elastic strains can be characterized by the elastic left Cauchy-Green tensor

$$\mathbf{B}_{\mathrm{e}} := \mathbf{F}_{\mathrm{e}}\,\mathbf{F}_{\mathrm{e}}^{\mathrm{T}}. \tag{14}$$

The velocity gradient tensor is defined as

$$\mathbf{L} := \dot{\mathbf{F}}\,\mathbf{F}^{-1}. \tag{15}$$

Using this operator, the following objective derivative, known as Lee derivative or contravariant Oldroyd rate, is introduced

$$\mathscr{L}_v(\mathbf{A}) := \mathfrak{O}_{contravar}(\mathbf{A}) := \mathbf{F}\frac{d}{dt}(\mathbf{F}^{-1}\mathbf{A}\mathbf{F}^{-\mathrm{T}})\mathbf{F}^{\mathrm{T}} = \dot{\mathbf{A}} - \mathbf{L}\mathbf{A} - \mathbf{A}\mathbf{L}^{\mathrm{T}}. \tag{16}$$

The system of constitutive equations on the current configuration takes the following form

$$\mathbf{S} = 2\rho_{\mathrm{R}}\frac{\partial\psi(\mathbf{B}_{\mathrm{e}})}{\partial\mathbf{B}_{\mathrm{e}}}\mathbf{B}_{\mathrm{e}}, \quad -\mathscr{L}_v(\mathbf{B}_{\mathrm{e}})\mathbf{B}_{\mathrm{e}}^{-1} = \frac{2}{\eta}\mathbf{S}^{\mathrm{D}}, \quad \mathbf{B}_{\mathrm{e}}|_{t=0} = \mathbf{B}_{\mathrm{e}}^{0}. \tag{17}$$

3.3 Properties of the model

Constitutive equations (13) and (17) are equivalent (cf. [2]). The material model is objective and thermodynamically consistent [5]. The inelastic flow is incompressible: $\det(\mathbf{C}_{\mathrm{i}}) \equiv 1$ and $\det(\mathbf{B}_{\mathrm{e}}) \equiv (\det(\mathbf{F}))^2$. The model exhibits a pure volumetric-iscohoric split in the sense of properties i, ii, iii. Moreover, as shown in [6], this model is w-invariant under arbitrary isochoric changes of the reference configuration. In particular, the w-invariance implies that the form of the constitutive equations $(13)_1$ and $(13)_2$ does not change if a new reference is considered; only the initial conditions $(13)_3$ have to be adjusted to match the new reference. In case of a very fast loading (as viscosity $\eta \to \infty$) the inelastic flow becomes frozen and the material response converges to pure hyperelasticity. Another important property of the model is the fading memory. More precisely: the model exhibits the so-called exponential stability of the solution with respect to small perturbations of the initial values [6]. The model is free from spurious oscillations of shear stresses under simple shear. Concerning the numerical implementation of the model, this exponential stability allows one to suppress the accumulation of the integration error [6], both for implicit and explicit time-stepping procedures. A simple explicit update formula for implicit time integration is available for neo-Hookean potentials [2]. Its generalization for Yeoh potential is presented in [7]. Some generalizations of this multiplicative Maxwell model are possible to capture a nonlinear kinematic hardening [8, 9, 10] and even a nonlinear distortional hardening [11, 12].

4 ADDITIVE DECOMPOSITION OF THE STRAIN RATE WITH HYPOELASTICITY

4.1 Constitutive equations

Another class of Maxwell models is based on the additive decomposition of the strain rate tensor $\mathbf{D} := \frac{1}{2}(\mathbf{L} + \mathbf{L}^{\mathrm{T}})$ into the inelastic part $\mathbf{D}_{\mathrm{i}}$ and the elastic part $\mathbf{D}_{\mathrm{e}}$ (cf. [13, 14]).

$$\mathbf{D} = \mathbf{D}_{\mathrm{i}} + \mathbf{D}_{\mathrm{e}}. \tag{18}$$

The elastic strain rate $\mathbf{D}_\mathrm{e}$ is related to a certain objective rate of the stress tensor $\mathbf{S}$ via grade-zero hypoelasticity

$$\mathbf{D}_\mathrm{e} = \mathbb{H} : \overset{\circ}{\mathbf{S}}, \tag{19}$$

where $\mathbb{H}$ is a constant fourth rank tensor (compliance tensor). Here we suppose that elastic properties are isotropic. Therefore, for the inverse $\mathbb{H}^{-1}$ (stiffness tensor) we have

$$\mathbb{H}^{-1} : \mathbf{X} = \frac{E}{3(1-2\nu)}\mathrm{tr}(\mathbf{X})\ \mathbf{1} + \frac{E}{1+\nu}\mathbf{X}^\mathrm{D}, \text{ for all } \mathbf{X} \in \mathrm{Sym}, \tag{20}$$

where E and ν are elastic constants. Note that (20) naturally implies the volumetric-isochoric split. In this section we consider corotational rates only:

$$\overset{\circ}{\mathbf{S}} = \dot{\mathbf{S}} + \mathbf{S}\boldsymbol{\Omega} - \boldsymbol{\Omega}\mathbf{S}, \quad \boldsymbol{\Omega} \in \mathrm{Skew}, \tag{21}$$

where Skew is the set of skew-symmetric second-rank tensors, the tensor $\boldsymbol{\Omega}$ is the so-called spin tensor, which is a kinematic quantity (it depends on the deformation process only). The inelastic flow rule takes the form

$$\mathbf{D}_\mathrm{i} = \frac{1}{\eta}\,\mathbf{S}^\mathrm{D}. \tag{22}$$

Combining (19) with (22) and introducing appropriate initial condition for the Kirchhoff stress, we obtain the following local initial value problem

$$\overset{\circ}{\mathbf{S}} = \mathbb{H}^{-1} : (\mathbf{D} - \frac{1}{\eta}\,\mathbf{S}^\mathrm{D}), \quad \mathbf{S}|_{t=t_0} = \mathbf{S}_0. \tag{23}$$

To close the system of constitutive equations, it remains to define the spin tensor $\boldsymbol{\Omega}$. In case of the classical Zaremba-Jaumann rate, we have

$$\boldsymbol{\Omega}^\mathrm{ZJ} := \mathbf{W} := \frac{1}{2}(\mathbf{L} - \mathbf{L}^\mathrm{T}). \tag{24}$$

Another popular choice is the Green-Naghdi rate, also known as Green-Naghdi-Dienes rate, Green-McInnis rate or polar rate, defined by the spin

$$\boldsymbol{\Omega}^\mathrm{GN} := \dot{\mathbf{R}}\mathbf{R}^\mathrm{T}, \quad \mathbf{R} := (\mathbf{F}\mathbf{F}^\mathrm{T})^{-1/2}\,\mathbf{F} = \mathbf{F}(\mathbf{F}^\mathrm{T}\mathbf{F})^{-1/2}. \tag{25}$$

A common drawback of these corotational rates is that the constitutive equations are not integrable even for a frozen inelastic flow (which corresponds to $\eta = \infty$): Even for $\mathbf{D} \equiv \mathbf{D}_\mathrm{e}$ the stress state depends on the loading path, which contradicts the notion of path-independent elasticity.

Considereing the models of type (23) with $\mathbb{H} = const$, $\boldsymbol{\Omega} \in$ Skew and $\eta = \infty$, only the so-called logarithmic spin $\boldsymbol{\Omega}^\mathrm{log}$ allows one to obtain path-independent stress-strain relations (cf. [15]). According to (21), the corotational logarithmic rate is given by

$$\overset{\circ}{\mathbf{S}}{}^\mathrm{log} := \dot{\mathbf{S}} + \mathbf{S}\boldsymbol{\Omega}^\mathrm{log} - \boldsymbol{\Omega}^\mathrm{log}\mathbf{S}, \tag{26}$$

where $\boldsymbol{\Omega}^\mathrm{log}$ stands for the so-called log-spin [16]. The logarithmic spin is defined implicitly through the following property

$$\mathbf{D} = (\ln \overset{\circ}{\mathbf{V}})^\mathrm{log}, \quad \text{where} \quad \mathbf{V} := (\mathbf{F}\mathbf{F}^\mathrm{T})^{1/2}. \tag{27}$$

An explicit expression for $\boldsymbol{\Omega}^\mathrm{log}$ can be found in [16].

4.2 Properties of the models

Let us summarize the basic properties of the models described by (23). The models are objective and the inelastic flow is incompressible. The material response exhibits a pure volumetric-iscohoric split in the sense of conditions i, ii, iii. In particular, $\mathrm{tr}(\mathbf{S}) = k\ln(\det(\mathbf{F}))$. Among the considered spins $\boldsymbol{\Omega}^{\mathrm{ZJ}}$, $\boldsymbol{\Omega}^{\mathrm{GN}}$, and $\boldsymbol{\Omega}^{\log}$, only the logarithmic spin $\boldsymbol{\Omega}^{\log}$ allows one to obtain a thermodynamically consistent stress response. For this spin, in case of a very fast loading (as $\eta \to \infty$), the stress response converges to a certain hyperelastic response (with a quadratic logarithmic strain energy function). The model (23) is w-invariant iff the corresponding spin is independent of the choice of the reference configuration [3]. The Zaremba-Jaumann spin $\boldsymbol{\Omega}^{\mathrm{ZJ}}$ yields a weakly invariant model, whereby $\boldsymbol{\Omega}^{\mathrm{GN}}$ and $\boldsymbol{\Omega}^{\log}$ depend on the choice of the reference [3]. Thus, dealing with relations of type (23), it is impossible to combine thermodynamic consistency and the w-invariance in a single model. Finally, it is well known that the spin $\boldsymbol{\Omega}^{\mathrm{ZJ}}$ yields unphysical oscillatory stress response under simple shear.

5 MODEL BASED ON THE ADDITIVE DECOMPOSITION OF THE LOGARITHMIC STRAIN

5.1 Constitutive equations

Following [17, 18, 19], we consider an approach which employs the structure of the the small-strain relations (1) – (3). In order to use these relations in the large strain case, the linearized strain tensor $\boldsymbol{\varepsilon}(t)$ is replaced by the referential logarithmic strain (Hencky strain)

$$\mathbf{H}(t) := \frac{1}{2}\ln(\mathbf{C}(t)). \tag{28}$$

The stress response $\boldsymbol{\sigma}$, resulting from (1) – (3) with $\boldsymbol{\varepsilon}(t) = \mathbf{H}(t)$, is a stress measure operating on the reference configuration which has to be power conjugate to $\mathbf{H}$

$$\boldsymbol{\sigma} : \dot{\mathbf{H}} = \widetilde{\mathbf{T}} : \Big(\frac{1}{2}\dot{\mathbf{C}}\Big) \text{ for all } \dot{\mathbf{C}} \in Sym. \tag{29}$$

In order to obtain the relation between $\boldsymbol{\sigma}$ and the second Piola-Kirchhoff stress, we recust (29) in the form

$$\boldsymbol{\sigma} : \Big(\frac{1}{2}\frac{\partial \ln(\mathbf{C})}{\partial \mathbf{C}} : \dot{\mathbf{C}}\Big) = \widetilde{\mathbf{T}} : \Big(\frac{1}{2}\dot{\mathbf{C}}\Big) \text{ for all } \dot{\mathbf{C}} \in Sym. \tag{30}$$

Thus, we arrive at

$$\Big(\frac{\partial \ln(\mathbf{C})}{\partial \mathbf{C}} : \boldsymbol{\sigma}\Big) : \dot{\mathbf{C}} = \widetilde{\mathbf{T}} : \dot{\mathbf{C}} \text{ for all } \dot{\mathbf{C}} \in Sym. \tag{31}$$

Therefore,

$$\widetilde{\mathbf{T}} = \frac{\partial \ln(\mathbf{C})}{\partial \mathbf{C}} : \boldsymbol{\sigma}, \quad \mathbf{T} = \frac{1}{\det \mathbf{F}}\mathbf{F}\Big(\frac{\partial \ln(\mathbf{C})}{\partial \mathbf{C}} : \boldsymbol{\sigma}\Big)\mathbf{F}^{\mathrm{T}}. \tag{32}$$

The entire model is subdivided into three modules (or computation steps) [18]:

- equation (28) as a geometric preprocessor ($\mathbf{C} \mapsto \mathbf{H}$)
- equations (1) – (3) as a constitutive model ($\mathbf{H} \equiv \boldsymbol{\varepsilon} \mapsto \boldsymbol{\sigma}$)
- equation $(32)_1$ as a geometric postprocessor ($\boldsymbol{\sigma} \mapsto \widetilde{\mathbf{T}}$)

5.2 Properties of the model

The considered material model is objective and thermodynamically consistent. For a given deformation process, the model exhibits exponential stability of the solution with respect to small perturbations of initial data. The inelastic flow is incompressible. Through some computations, it can be shown that the model exhibits a pure volumetric-iscohoric split in the sense of restrictions i, ii, iii. In particular, $\mathrm{tr}(\mathbf{C}\widetilde{\mathbf{T}}) = \mathrm{tr}\ \mathbf{S} = \mathrm{tr}\ \boldsymbol{\sigma} = k\ln(\det(\mathbf{F}))$. The model is *not w-invariant*: Even for $\eta = \infty$, the elastic response depends on the choice of the reference configuration [3]. The model is free from spurious oscillations of shear stresses under simple shear. Some well-established numerical procedures of small-strain theory can be applied for the numerical integration of the evolution equations.

6 COVARIANT MAXWELL MODEL

6.1 Constitutive equations

Let us denote the covaraint Oldroyd derivative by $\mathfrak{O}_{covar}$:

$$\mathfrak{O}_{covar}(\mathbf{S}) := \dot{\mathbf{S}} + \mathbf{L}^{\mathrm{T}}\,\mathbf{S} + \mathbf{S}\,\mathbf{L}. \tag{33}$$

The so-called covariant Maxwell fluid, as proposed by Haupt and Lion to model incompressible materials [20], is given by the following local initial value problem

$$\mathfrak{O}_{covar}(\mathbf{S}) + \frac{2\mu}{\eta}\mathbf{S} = 2\mu\mathbf{D}, \quad \mathbf{S}|_{t=t_0} = \mathbf{S}_0. \tag{34}$$

Contravariant pull-back of (34) to the reference configuration yields its equivalent formulation in terms of the convected stress tensor $\widetilde{\mathbf{t}} := \mathbf{F}^{\mathrm{T}}\,\mathbf{S}\,\mathbf{F}$ and Green's strain tensor $\mathbf{E} := \frac{1}{2}(\mathbf{C} - \mathbf{1})$

$$\dot{\widetilde{\mathbf{t}}} + \frac{2\mu}{\eta}\widetilde{\mathbf{t}} = 2\mu\dot{\mathbf{E}}, \quad \widetilde{\mathbf{t}}|_{t=t_0} = \widetilde{\mathbf{t}}_0. \tag{35}$$

Since the model (33) was originally developed for incompressible materials, the bulk modulus k does not enter this constitutive formulation.

In order to incorporate the volumetric response in case of a general compressible material, one may consider the following generalization: The stresses predicted by (34) are now interpreted as auxiliary stresses $\mathbf{S}_{\text{aux}}$, used to determine the deviatoric part of the overall stress response

$$\mathbf{S}(t) = (\mathbf{S}_{\text{aux}}(t))^{\mathrm{D}} + p(\det(\mathbf{F}(t)))\mathbf{1}, \tag{36}$$

$$\dot{\mathbf{S}}_{\text{aux}} + (\mathbf{L}^{\mathrm{D}})^{\mathrm{T}}\,\mathbf{S}_{\text{aux}} + \mathbf{S}_{\text{aux}}\,\mathbf{L}^{\mathrm{D}} + \frac{2\mu}{\eta}\mathbf{S}_{\text{aux}} = 2\mu\mathbf{D}^{\mathrm{D}}, \quad \mathbf{S}_{\text{aux}}|_{t=t_0} = \mathbf{S}_0. \tag{37}$$

Here, $p(\det(\mathbf{F}(t)))$ is a suitable function. In analogy to the previous models, discussed in Sections 4 and 5, one may put $p(\det(\mathbf{F}(t))) = k\ln(\det(\mathbf{F}(t)))$.

6.2 Properties of the model

The model (34) is objective and thermodynamically consistent (cf. [20]). It follows immediately from (35) that for a given (local) deformation process, the solution exhibits exponential stability with respect to small perturbations of the initial data. Equations (34) are w-invariant,

since all quantities which appear in (34) are invariant under the change of the reference configuration. For that reason, equations (35) are w-invariant as well. In case of a very fast loading (as $\eta \to \infty$), the stress response converges to a pure hyperelasticity. Thanks to the very simple structure of equations (35), efficient and robust time-stepping algorithms are available. Unfortunately, the covariant model (34) *does not exhibit* a pure volumetric-iscohoric split: the condition i is violated. Indeed, for arbitrary incompressible process (with $\det(\mathbf{F}) \equiv 1$), taking the trace of (33), we arrive at

$$\frac{d}{dt}(\mathrm{tr}\mathbf{S}) + 2\,\mathbf{S} : \mathbf{D} + \frac{2\mu}{\eta}\mathrm{tr}\mathbf{S} = 0. \tag{38}$$

In general, the stress power is not equal to zero: $\mathbf{S} : \mathbf{D} \neq 0$. Thus, according to (38), $\mathrm{tr}\mathbf{S} \neq 0$ even if $\mathrm{tr}\mathbf{S}|_{t=t_0} = 0$. According to (38), even for incompressible processes, the deviatoric part of the stress influences the evolution of the volumetric part.

7 MODIFIED COVARIANT MODEL WITH A PURE V-I SPLIT

7.1 Constitutive equations

In this paper we modify the model (34) in order to enforce the property i: For arbitrary incompressible deformations we consider

$$\mathfrak{O}_{covar}(\mathbf{S}) + \frac{2\mu}{\eta}\mathbf{S} = 2\mu\mathbf{D} + \frac{2}{3}(\mathbf{S} : \mathbf{D})\mathbf{1}, \quad \mathbf{S}|_{t=t_0} = \mathbf{S}_0. \tag{39}$$

The covariant pull-back of these relations yields

$$\dot{\widetilde{\mathbf{t}}} + \frac{2\mu}{\eta}\widetilde{\mathbf{t}} = 2\mu\dot{\mathbf{E}} - \frac{1}{3}\big(\widetilde{\mathbf{t}} : \frac{d}{dt}(\mathbf{C}^{-1})\big)\,\mathbf{C}, \quad \widetilde{\mathbf{t}}|_{t=t_0} = \widetilde{\mathbf{t}}_0. \tag{40}$$

In order to model the stress response of a compressible Maxwell fluid, in analogy to (36)-(37), we may consider the following generalization

$$\mathbf{S}(t) = \mathbf{S}_{\text{aux}}(t) + p(\det(\mathbf{F}(t)))\mathbf{1}, \tag{41}$$

$$\dot{\mathbf{S}}_{\text{aux}} + (\mathbf{L}^{\mathrm{D}})^{\mathrm{T}}\,\mathbf{S}_{\text{aux}} + \mathbf{S}_{\text{aux}}\,\mathbf{L}^{\mathrm{D}} + \frac{2\mu}{\eta}\mathbf{S}_{\text{aux}} = 2\mu\mathbf{D}^{\mathrm{D}} + \frac{2}{3}(\mathbf{S}_{\text{aux}} : \mathbf{D}^{\mathrm{D}})\mathbf{1}, \quad \mathbf{S}_{\text{aux}}|_{t=t_0} = \mathbf{S}_0. \tag{42}$$

7.2 Properties of the model

The suggested model is objective and w-invariant. The condition i is satisfied. Indeed, let $\det(\mathbf{F}) \equiv 1$. Then, taking the trace of (39), we obtain

$$\frac{d}{dt}(\mathrm{tr}\mathbf{S}) + \frac{2\mu}{\eta}\mathrm{tr}\mathbf{S} = 0, \quad \mathrm{tr}\mathbf{S}(t) = \exp\Big(-\frac{2\mu}{\eta}(t - t_0)\Big)\,\mathrm{tr}\mathbf{S}|_{t=t_0}. \tag{43}$$

If $\mathrm{tr}\mathbf{S}|_{t=t_0} = 0$, then $\mathrm{tr}\mathbf{S} \equiv 0$, which is exactly the required property i. Next, for the model (41)-(42), the properties i, ii, and iii are satisfied.

8 CONTRAVARIANT MAXWELL MODEL

8.1 Constitutive equations

Recall the contravaraint Oldroyd derivative $\mathfrak{O}_{contravar}$:

$$\mathfrak{O}_{contravar}(\mathbf{S}) := \dot{\mathbf{S}} - \mathbf{L}\mathbf{S} - \mathbf{S}\mathbf{L}^{\mathrm{T}}. \tag{44}$$

The so-called contravariant Maxwell fluid was proposed by Haupt and Lion in the incompressible case [20]:

$$\mathfrak{O}_{contravar}(\mathbf{S}) + \frac{2\mu}{\eta}\mathbf{S} = 2\mu\mathbf{D}, \quad \mathbf{S}|_{t=t_0} = \mathbf{S}_0. \tag{45}$$

Contravariant pull-back of (45) to the reference configuration yields the following equivalent formulation in terms of the second Piola-Kirchhoff stress

$$\dot{\widetilde{\mathbf{T}}} + \frac{2\mu}{\eta}\widetilde{\mathbf{T}} = -2\mu\frac{d}{dt}\mathbf{C}^{-1}, \quad \widetilde{\mathbf{T}}|_{t=t_0} = \widetilde{\mathbf{T}}_0. \tag{46}$$

Analogously to the previous section, we generalize the model to cover arbitrary deformations: The stresses predicted by (45) are, again, auxiliary stresses $\mathbf{S}_{\text{aux}}$, used to determine the deviatoric part of the overall stress response

$$\mathbf{S}(t) = (\mathbf{S}_{\text{aux}}(t))^{\text{D}} + p(\det(\mathbf{F}(t)))\mathbf{1}, \tag{47}$$

$$\dot{\mathbf{S}}_{\text{aux}} - \mathbf{L}^{\text{D}}\,\mathbf{S}_{\text{aux}} - \mathbf{S}_{\text{aux}}\,(\mathbf{L}^{\text{D}})^{\text{T}} + \frac{2\mu}{\eta}\mathbf{S}_{\text{aux}} = 2\mu\mathbf{D}^{\text{D}}, \quad \mathbf{S}_{\text{aux}}|_{t=t_0} = \mathbf{S}_0. \tag{48}$$

8.2 Properties of the model

Just as the covaraint model, the model (45) is objective and thermodynamically consistent (cf. [20]). It follows from (46) that the solution is exponentially stable with respect to small perturbations of the initial conditions. Equations (45) are w-invariant. In case of a very fast loading (as $\eta \to \infty$), the stress response converges to a pure hyperelasticity.[1] Due to the simple linear structure of equations (46), efficient and robust numerical algorithms are available. Just as the covaraint model, its contravariant counterpart (45) *does not exhibit* a pure volumetric-iscohoric split, since the condition i is again violated. To demonstrate this, consider an incompressible deformation process with $\det(\mathbf{F}) \equiv 1$. Taking the trace of (45), we have

$$\frac{d}{dt}(\text{tr}\mathbf{S}) + 2\,\mathbf{S} : \mathbf{D} + \frac{2\mu}{\eta}\text{tr}\mathbf{S} = 0. \tag{49}$$

Again, generally, the stress power $\mathbf{S} : \mathbf{D}$ is not equal to zero. Thus, according to (49), $\text{tr}\mathbf{S} \neq 0$ even if $\text{tr}\mathbf{S}|_{t=t_0} = 0$.

9 MODIFIED CONTRAVARIANT MODEL WITH A PURE V-I SPLIT

9.1 Constitutive equations

In order to enforce the property i, we modify the equations of the contravariant Maxwell fluid. For incompressible material we consider

$$\mathfrak{O}_{contravar}(\mathbf{S}) + \frac{2\mu}{\eta}\mathbf{S} = 2\mu\mathbf{D} - \frac{2}{3}(\mathbf{S} : \mathbf{D})\mathbf{1}, \quad \mathbf{S}|_{t=t_0} = \mathbf{S}_0. \tag{50}$$

An equivalent formulation is obtained by the contravariant pull-back to the reference configuration

$$\dot{\widetilde{\mathbf{T}}} + \frac{2\mu}{\eta}\widetilde{\mathbf{T}} = -2\mu\frac{d}{dt}\mathbf{C}^{-1} - \frac{1}{3}(\widetilde{\mathbf{T}} : \dot{\mathbf{C}})\,\mathbf{C}^{-1}, \quad \widetilde{\mathbf{T}}|_{t=t_0} = \widetilde{\mathbf{T}}_0. \tag{51}$$

[1] A linear combination of the covariant and contravariant models allows one to obtain a model, which generalizes Monney-Rivlin hyperelasticity to finite-strain viscoelasticity.

In a general compressible case we put

$$\mathbf{S}(t) = \mathbf{S}_{\text{aux}}(t) + p(\det(\mathbf{F}(t)))\mathbf{1}, \tag{52}$$

$$\dot{\mathbf{S}}_{\text{aux}} - \mathbf{L}^{\text{D}}\,\mathbf{S}_{\text{aux}} - \mathbf{S}_{\text{aux}}\,(\mathbf{L}^{\text{D}})^{\text{T}} + \frac{2\mu}{\eta}\mathbf{S}_{\text{aux}} = 2\mu\mathbf{D}^{\text{D}} - \frac{2}{3}(\mathbf{S}_{\text{aux}} : \mathbf{D}^{\text{D}})\mathbf{1}, \quad \mathbf{S}_{\text{aux}}|_{t=t_0} = \mathbf{S}_0. \tag{53}$$

9.2 Properties of the model

This modified contravarinat model is objective and w-invariant. The condition i is identically satisfied. To check this property, we consider an incompressible deformation with $\det(\mathbf{F}) \equiv 1$. Then, taking the trace of (50), we arrive at

$$\frac{d}{dt}(\text{tr}\mathbf{S}) + \frac{2\mu}{\eta}\text{tr}\mathbf{S} = 0, \quad \text{tr}\mathbf{S}(t) = \exp\Big(-\frac{2\mu}{\eta}(t-t_0)\Big)\,\text{tr}\mathbf{S}|_{t=t_0}. \tag{54}$$

Thus, for $\text{tr}\mathbf{S}|_{t=t_0} = 0$, we have $\text{tr}\mathbf{S} \equiv 0$ and the property i holds true. Finally, for the generalized model (52)-(53), the properties i, ii, and iii are satisfied.

10 CONCLUSIONS

Seven different approaches to finite strain Maxwell fluid are considered in this paper. The approaches are compared qualitatively, using a number of criteria. In particular, the models are analyzed with respect to the w-invariance under the isochoric change of the reference configuration and the possibility of a pure volumetric-iscohoric split. *Two new models are proposed here*, basing on the covariant and contravariant Maxwell models, previously proposed by Haupt and Lion. In contrast to the original formulations, the modified models exhibit the pure i-v split. In the follow-up paper, some other commonly used approaches to the phenomenological elasto-plasticity will be analyzed and the links between the similar frameworks will be established.

Acknowledgement: This research was supported by the Russian Science Foundation (project number 15-11-20013).

REFERENCES

[1] M. Reiner, *Deformation, Strain and Flow. An Elementary Introduction to Rheology, 2nd edition*. H. K. Lewis, London, 1960.

[2] A.V. Shutov, R. Landgraf, J. Ihelmann, An explicit solution for implicit time stepping in multiplicative finite strain viscoelasticity. *Computer Methods in Applied Mechanics and Engineering*, **256**, 213–225, 2013.

[3] A.V. Shutov, J. Ihlemann, Analysis of some basic approaches to finite strain elasto-plasticity in view of reference change. *International Journal of Plasticity*, **63**, 183–197, 2014.

[4] J.C. Simo, C. Miehe, Associative coupled thermoplasticity at finite strains: formulation, numerical analysis and implementation. *Computer Methods in Applied Mechanics and Engineering*, **98**, 41–104, 1992.

[5] A. Lion, A physically based method to represent the thermo-mechanical behaviour of elastomers, *Acta Mechanica*, **123**, 1–25, 1997.

[6] A.V. Shutov, R. Kreißig, Geometric integrators for multiplicative viscoplasticity: analysis of error accumulation. *Computer Methods in Applied Mechanics and Engineering*, **199**, 700–711, 2010.

[7] R. Landgraf, A.V. Shutov, J. Ihlemann, Efficient time integration in multiplicative inelasticity. *Proceedings in Applied Mathematics and Mechanics*, **15**, 325–326, 2015.

[8] A. Lion, Constitutive modelling in finite thermoviscoplasticity: a physical approach based on nonlinear rheological elements, *International Journal of Plasticity*, **16**, 469–494, 2000.

[9] A.V. Shutov, R. Kreißig, Finite strain viscoplasticity with nonlinear kinematic hardening: Phenomenological modeling and time integration. *Computer Methods in Applied Mechanics and Engineering*, **197**, 2015–2029, 2008.

[10] I. Vladimirov, M. Pietryga, S. Reese, On the modelling of non-linear kinematic hardening at finite strains with application to springback Comparison of time integration algorithms. *International Journal for Numerical Methods in Engineering*, **75**, 1–28, 2008.

[11] A.V. Shutov, S. Panhans, R. Kreißig, A phenomenological model of finite strain viscoplasticity with distortionalhardening. *ZAMM*, **91**8, 653–680, 2011.

[12] M. Freund, A. V. Shutov, J. Ihlemann, Simulation of distortional hardening by generalizing a uniaxial model of finite strain viscoplasticity. *International Journal of Plasticity*, **36**, 113–129, 2012.

[13] K.W. Neale, Phenomenological constitutive laws in finite plasticity. *Solid Mechanics Archives*, **6**, 79–128, 1981.

[14] S. Nemat-Nasser, On finite deformation elastoplasticity. *International Journal of Solids and Structues*, **18**, 857–872, 1982.

[15] O.T. Bruhns, H. Xiao, A. Meyers, Self-consistent Eulerian rate type elasto-plasticity models based upon the logarithmic stress rate. *International Journal of Plasticity*, **15**, 479–520, 1999.

[16] H. Xiao, O.T. Bruhns, A. Meyers, Logarithmic strain, logarithmic spin and logarithmic rate. *Acta Mechanica*, **124**, 89–105, 1997.

[17] P. Papadopoulus, J. Lu, A general framework for the numerical solution of problems in finite elasto-plasticity. *Computer Methods in Applied Mechanics and Engineering*, **159**, 1–18, 1998.

[18] C. Miehe, N. Apel, M. Lambrecht, Anisotropic additive plasticity in the logarithmic strain space: modular kinematic formulation and implementation based on incremental minimization principles for standard materials. *Computer Methods in Applied Mechanics and Engineering*, **191**, 5383–5425, 2002.

[19] Schröder, J., Gruttmann, F., Löblein, J., A simple orthotropic finite elasto-plasticity model based on generalized stress-strain measures. *Computational Mechanics*, **30**, 48–64, 2002.

[20] P. Haupt, A. Lion, On finite linear viscoelasticity of incompressible isotropic materials. *Acta Mechanica*, **159**, 87–124, 2002.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

SIMULATING ATOMIC FORCE MICROSCOPY FOR THE DETERMINATION OF THE ELASTIC PROPERTIES OF NANOPARTICLE REINFORCED EPOXY RESIN

Johannes Fankhänel*[1], Andreas Kempe[1], and Raimund Rolfes[1]

[1] Institute of Structural Analysis, Leibniz Universität Hannover
Appelstraße 9a, 30167 Hannover, Germany
*e-mail: j.fankhaenel@isd.uni-hannover.de

Keywords: Molecular Dynamic Finite Element Method (MDFEM), Computational Mechanics, Nanocomposite Simulation, Elastic Properties, Atomic Force Microscopy (AFM)

Abstract. *Nanoparticles show a great potential in improving especially the matrix-dominated mechanical properties of fiber reinforced plastics, like compressive strength or impact tolerance. The composition of the interphase between nanoparticles and the surrounding matrix is assumed to be of vital importance for the mechanical properties of the composite material. Characterizing nanoparticle-matrix interphases with experimental methods, e.g. using atomic force microscopy (AFM), is highly complex and time consuming. Therefore an AFM-simulation technique based on the Molecular Dynamic Finite Element Method (MDFEM) is introduced. The MDFEM provides a powerful method for simulating molecular dynamic problems within the finite element framework, perspectively allowing for the efficient simulation of multi-scale models and pure FE-models in order to reduce the numerical cost for bigger problems. With the presented method, the elastic properties of pure boehmite particles as well as pure epoxy resin have been determined and are in good agreement with experimentally obtained values. The influence of different particle-matrix interactions on the elastic properties of the interphase has been studied for unmodified boehmite particles. Supporting virtual tensile tests on cubic unit cells show outstanding accordance with experimental results. The presented method can contribute to the optimization of nanocomposite materials and at the same time reduce the need for experimental effort.*

1 INTRODUCTION

Carbon fiber composites are regarded as one of the most important lightweight construction materials for the future. By incorporating nanoparticles, it is possible to improve the properties of the weak constituent of the composite material, i.e. the matrix material. Arlt [1] shows that unmodified and taurine modified boehmite particles are able to improve matrix-dominated properties of carbon fiber/epoxy composites, like shear strength, shear modulus, compressive strength or compression after impact strength by 10 to 25 %. Other studies show similar results. Uddin & Sun [2] investigate silica particles in a glass/epoxy composite and report significant improvements regarding the compressive strength and modulus as well as the tensile strength perpendicular to the fiber direction. Subramaniyan & Sun [3] obtain similar results for a nanoclay/glass/epoxy composite.

In contrast, there are studies that report negative effects, especially for higher particle weight fractions. Shahid et al. [4] investigate boehmite particles with two surface modifications (lysine and para-hydroxybenzoate) in a carbon fiber/epoxy composite and report an increase of tensile and flexural properties for very small particle fractions. For weight fractions above approximately 5 %, the properties decrease compared to the unfilled composite. Siddiqui et al. [5] observe partly increased properties for nanoclay composites, like flexural modulus and fracture toughness and partly decreased properties, like flexural strength and impact toughness.

Comparing the results of these investigations, it can be concluded that the composition of the interphase between nanoparticles and matrix is of vital importance for the mechanical properties of the composite. This becomes especially obvious, when considering that Arlt and Shahid et al. investigate a similar boehmite/epoxy material combination, but with different particle surface modifications and achieved completely different results. Following the argumentation of Arlt, nanoparticles and their surface modifications highly influence the morphology of the surrounding matrix by impeding the cross-linking reaction and thus forming a soft interphase around the particles. Thereby two mechanisms may occur. Firstly, hydroxyl groups on the particle surface are assumed to be able to react with epoxy groups of the matrix during the curing reaction instead of hydroxyl groups of the matrix. This should lead to approximately the same amount of cross-links as compared to the undisturbed matrix. However, some of those cross-links are formed between particle and matrix instead of between prepolymer chains, resulting in an improved particle matrix interaction and in a softer matrix material. As a second possibility, the hydroxyl groups on the particle surface are assumed to be able to react with the prepolymer chains through etherification even before the hardener is added to the uncured matrix. Thus, many epoxy groups in the matrix near the particle surfaces are already occupied and no longer available for further reactions during the curing reaction, leading to a lower cross-linking density between prepolymer chains in this area. This should again lead to an improved particle matrix interaction at the cost of a soft matrix material around the particles. These phenomena have been substantiated in the work of Arlt by Atomic Force Microscopy experiments. It has been found that unmodified boehmite particles exhibit the aforementioned surrounding ring of soft material in the ambient matrix.

AFM is often used to validate molecular dynamics simulations but rather less literature on molecular dynamics AFM simulations of the indentation behavior is available. Many studies focus on characterizing the adhesion and separation behavior. Landman et al. [6] for example simulate adhesion and separation during the indentation of a nickel tip in a gold sample. Other studies, like Fang et al. [7], apply molecular dynamics AFM simulations to the investigation of the lithography process, in this case of crystalline copper with good agreement with experimental results. Still other studies investigate pull out processes, e.g. Heymann and

Grubmüller [8] perform single molecule atomic force microscopy simulations by pulling out a hapten molecule of a ANO2 binding pocket. In the present study, atomic force microscopy simulations in order to investigate the elastic properties of particle-matrix interphases are presented. The specific goal is to gain evidence for the discussed mechanisms and to validate the assumptions in the generation process of the models in a scale specific manner.

This paper is structured as follows: Starting with a short summary of the chosen simulation method (MDFEM), the generation process and the AFM simulation model is introduced in section 3. The numerical investigations presented in section 4 comprise two steps. The first step of the work presented includes AFM simulations, in which particle and matrix are analyzed separately. These preliminary investigations are used for the calibration and optimization of the AFM models. Different parameters like tip and sample size or boundary effects are analyzed. The second part deals with the actual characterization of the particle-matrix interphase. AFM simulations on samples consisting of both particles and matrix material as well as particle-matrix interphases are performed for a composite material comprising an anhydride-cured epoxy matrix and unmodified boehmite particles. The effect of three different particle-matrix interactions on the mechanical properties of the interphase is investigated and evaluated on the basis of intensive parametric studies. The results are discussed in the context of simulated tensile tests of the material combinations mentioned earlier.

2 THE MOLECULAR DYNAMIC FINITE ELEMENT METHOD

In this chapter, a brief overview of the Molecular Dynamic Finite Element Method (MDFEM) is given. For a more detailed presentation, the reader is referred to earlier work of the senior authors of the present paper [9, 10].

The MDFEM can be regarded as a framework to perform classical molecular dynamics (MD) simulations by means of the well-established finite element method (FEM). The main motivation for this approach is the simplification of multi-scale techniques, as for example the coupling of atomistic and continuum mechanics simulations can be performed in a very efficient way within one software package. Additionally, the need for special in-house developed software is reduced, since FEM software mostly comes with highly efficient solvers and pre- and post-processing tools.

When thinking of molecular structures as multi-body systems, where atoms are point masses and bonds are springs, the FEM seems to provide an appropriate framework for MD simulations, since structural elements such as trusses, beams and springs are an integral part of FE codes. A variety of such MD-FE models has been proposed for the simulation of graphene, see for example Wang [11], Zhang et al. [12] and Zhu et al. [13]. Nevertheless, there are some challenges or problems when applying standard truss or beam elements to general MD problems. The use of truss elements seems to be reasonable at the first glance, since nodes of truss elements similarly to atoms only have translational degrees of freedom. However, usual MD bending potentials require the introduction of either rotational springs and truss angles between two truss elements or normal springs between two connected outer nodes of the angle. Normal springs in turn, can only reproduce the bending potentials through very complex geometric relations for large deformations. Additionally, truss elements cannot reproduce torsion or inversion MD potentials, which makes truss elements only applicable for planar structures or special spatial structures, where torsion and inversion do not occur. Due to the lack of torsional stiffness, those models are also relatively instable.

An alternative approach is the approximation of bonds by means of standard beam elements. Many different models have been made proposed, like using e.g. rigidly connected beams or joints and constrained rotational degrees of freedom. Examples can be found in Li [14], Chwal et al. [15] and Flores et al. [16]. The main problem when using these models is

the determination of the beam constants. While normal stiffnesses can be calculated directly from the bond stretch potentials, bending stiffnesses can only be derived unambiguously for special cases, like planar structures, where torsion can be neglected. The reason for this behavior stems from the fact that torsion between two planes can also be expressed through bending of elements. Even for planar structures the beam element method can lead to wrong results, since beam rotations are transferred from beam to beam through the structure.

Even though it is possible to obtain the correct behavior with a certain superposition of rigidly and flexibly connected beam elements, as shown in [10], the MDFEM provides a more elegant method, using special 2-, 3- and 4-node elements for bond stretch and physical interactions, bending and torsion. By superposing those elements, various atomistic problems can be modeled. Advantages of the MDFEM, compared to the aforementioned approaches include automatic model relaxation due to intrinsic equilibrium parameters, the possibility of easily calculating large deformations and reproducing multiple equilibrium positions as well as a reduced numerical effort due to the elimination of the rotational degrees of freedom. Disadvantageous is the comparatively high effort to generate simulation models. Sample applications of the MDFEM include carbon nanotubes [9] as well as elastomers [17].

3 MATERIALS AND MODELS

The studies presented in the following chapters focus on boehmite nanoparticle-reinforced epoxy composites. The chemical structure of boehmite (γ–AlO(OH)) is assumed to be of the cmcm unit cell (see Figure 1 a)), as reported by Bokhimi et al. [18] and experimentally confirmed through X-ray diffraction carried out at the Institute for Particle Technology of the TU Braunschweig [19]. The primary particle size of the particles is in the range of 14 Å and the particles have an orthorhombic shape, as shown in Figure 1 b). The AFM tip material is silicon with an Fd3m (diamond) unit cell.

The matrix material is an epoxy resin based on Bisphenol-A-diglycidylether (DGEBPA) prepolymer chains and a Methyl-Tetrahydrophthalic Anhydride (MTHPA) hardener. Both chemical structures are shown in Figure 1 c) and d) respectively.

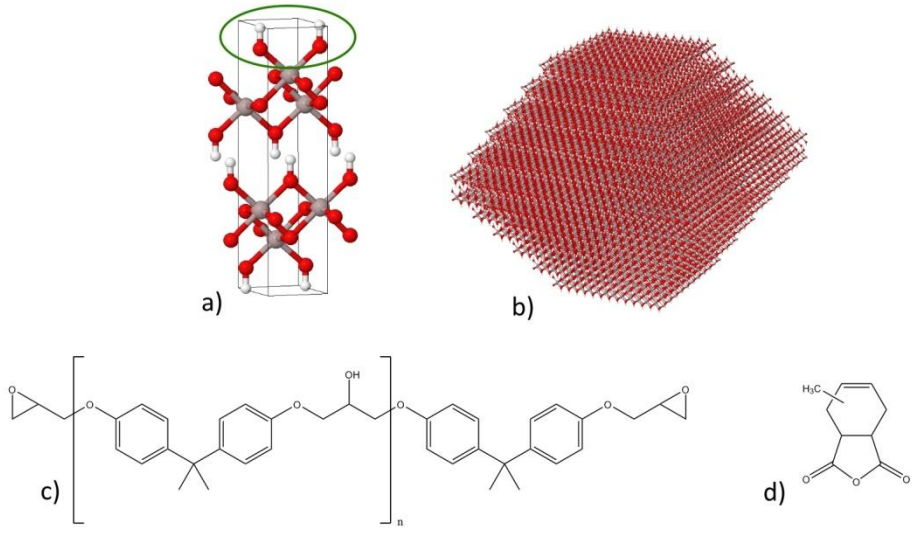


Figure 1: a) Crystallographic unit cell of cmcm boehmite; b) Orthorhombic shape of the boehmite particles; c) Chemical structure of DGEBPA prepolymer chains; d) Chemical structure of MTHPA hardener.

All models are generated with in-house software using Open Babel [20] and Packmol [21]. In a first step, the particle is created by duplicating the boehmite unit cell and cutting out the shape of the particle. Afterwards, prepolymer chains are randomly packed into the simulation box around the particle. Subsequently, the ester cross-links are created. The algorithm randomly iterates over all hydroxyl groups. If an epoxy group within a specified search radius is found, a hardener molecule is placed and connected. If multiple epoxy groups are found, the algorithm always uses the closest one. If all hardener molecules are placed or no more reaction partners are found, the same process is executed for the etherification reaction. Finally, the AFM-tip is created and placed, again by duplicating the silicon unit cell and cutting out the hemispherical shape.

Concerning the particle-matrix-interphase, three different variations are compared. All three variations are based on an unmodified boehmite particle. As a first variation, the particle is assumed to interact with the matrix only by physical interactions. In the second variation, the particle is bonded both physically and chemically. The surface hydroxyl groups of the boehmite (highlighted in Figure 1 a)) are enabled to develop covalent bonds to the epoxy groups of the matrix material by joining the cross-linking reaction. The third variation follows the argumentation of Arlt [1], that interphases around the particles can develop even before the epoxy is cured. To take this into account, a preliminary etherification step only for particle hydroxyl groups is inserted. Afterwards, the ester and ether connections for the remaining reactive sites are created. The three variations in the mentioned order are denoted as variation 1, 2 and 3 in the following.

All simulations presented in this study are performed using the MDFEM in combination with the Dreiding force field [22]. Even though it is possible to apply different cutoff radii for the physical interactions within the sample and between sample and tip, a cutoff of 15 Å has proven to be suitable for both cases by parametric studies. The tip itself is considered to be rigid, meaning that all chemical and physical interactions within the tip are eliminated, by removing the corresponding elements. Since real AFM tips are deformable, this has to be considered in the calculation of the elastic modulus of the sample (see chapter 4). The first step in the simulations is, as usual in molecular dynamics calculations, the determination of the equilibrium state of the system. Therefore, the whole model including the AFM tip is relaxed over a varying period of time ranging from 50 to 200 ps, depending on the model size. In the relaxation step, the equilibrium position of the tip (in the sense of the tip lying on the sample surface without cantilever forces) is obtained automatically. Subsequently, the load is applied in terms of a displacement of the tip by 0.5 to 1.5 nm, depending on the sample and tip size. A simulation time of 300 ps has proven to be suitable by parametric studies. All samples are constrained on all surfaces in a border region of 0.2 nm, except for the top surface, where the tip indents the material. A schematic representation of the used models can be seen in Figure 2.

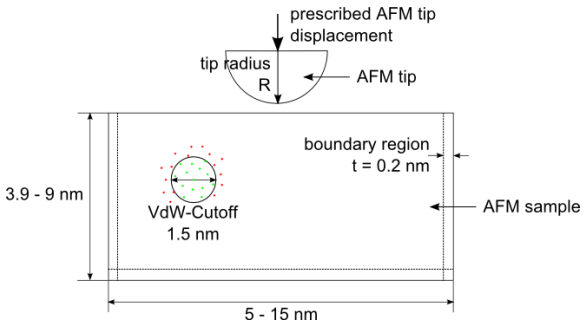


Figure 2: Schematic representation of the AFM models.

Due to the comparatively high numerical effort of MD simulations as well as MDFEM simulations, it is necessary to reduce the model size. The calculations for determining the interphase properties are performed on a 5 nm particle, which is considerably smaller than the primary particle size of approximately 14 nm observed in experiments. Minimum tip and sample sizes are obtained in preliminary simulations shown in the following section. To extract the interphase stiffnesses, the tip is shifted to the desired position in the global model after the relaxation step and subsequently the global model is reduced to a smaller submodel for each AFM-indent, as illustrated in Figure 3 a). After each shifting and reducing step, the 2 node elements representing the physical interactions are renewed and a short relaxation is performed. Then the actual load step is performed.

4 RESULTS AND DISCUSSION

As a result of the simulations, force-deformation curves are obtained. The Young's Modulus is then calculated through a Hertz fit similar to AFM experiments:

$$D = \left(\frac{F}{\sqrt{R} \cdot E_{tot}} \right)^{2/3} \tag{1}$$

with

$$\frac{1}{E_{tot}} = \frac{3}{4} \left(\frac{1 - {\nu_{tip}}^2}{E_{tip}} + \frac{1 - \nu^2}{E} \right), \tag{2}$$

where D is the deformation, F is the force acting on the cantilever beam (corresponding to $D_c \cdot k_c$ in the experiment, with the cantilever deflection D_c and the stiffness k_c of the cantilever beam), R is the tip radius and E_{tot} is the reduced elastic modulus of the series connection of tip and sample. If a rigid AFM tip is assumed, E_{tip} becomes infinite and the first term in the parenthesis in equation (2) vanishes. Combining Equation (1) and the reduced form of Equation (2) leads to

$$D = \left(\frac{3}{4} \cdot \frac{F(1 - \nu^2)}{\sqrt{R} E} \right)^{\frac{2}{3}}. \tag{3}$$

With E being the unknown parameter, Equation (3) is then fitted to the simulated curve applying a least square fit with MATLAB, as can be seen exemplarily in Figure 3 b).

The red dashed line in Figure 3 b) indicates the onset of a deviation between the fitted and simulated curve. This effect is caused by boundary effects (at a certain penetration the rigid boundaries start influencing the curve), indicating that the sample size in this example is chosen too small. To bypass this, either the sample size has to be increased or the range in which the Hertz fit is calculated has to be truncated (in this example the latter was already performed).

4.1 Preliminary Investigations

In preliminary investigations, AFM simulations on pure boehmite crystals and epoxy resin are performed in order to reexamine the assumptions like e.g. the crystal structure or the chosen force field. Additionally, the goal is to find the minimum model sizes (tip size, sample size), since model sizes have to be reduced due to the numerical limitations. Furthermore, the effect of the boundary conditions on the calculated elastic modulus is analyzed.

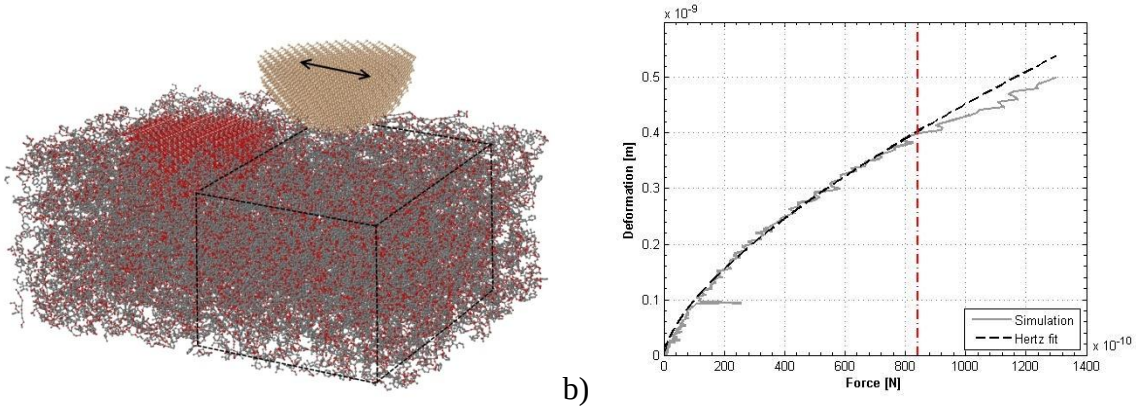


Figure 3: a) Example model of an AFM simulation (3 nm tip radius, 17.5 x 7.2 x 12 nm sample, approx. 90000 atoms, the black box indicates the minimum sample size, that is cut out prior to the loading step); b) Least square Hertz fit to simulated curve for the pure boehmite sample (red dashed line indicates onset of deviation between fitted and simulated curve caused by boundary effects).

To determine the minimum tip size, simulations with increasing tip radius (1 – 5 nm in steps of 1 nm) are performed for both materials. The boehmite is modeled with a sample size four times as big as the tip radius and the epoxy with a sample size of 15 x 9 x 15 nm. Both sample sizes are assumed to be large enough, not to impose any boundary effects on the measured modulus, which is shown to be valid in the following. The results concerning the effect of the tip radius on the calculated elastic modulus are depicted in Figure 4 a) for pure boehmite and Figure 4 b) for pure epoxy respectively.

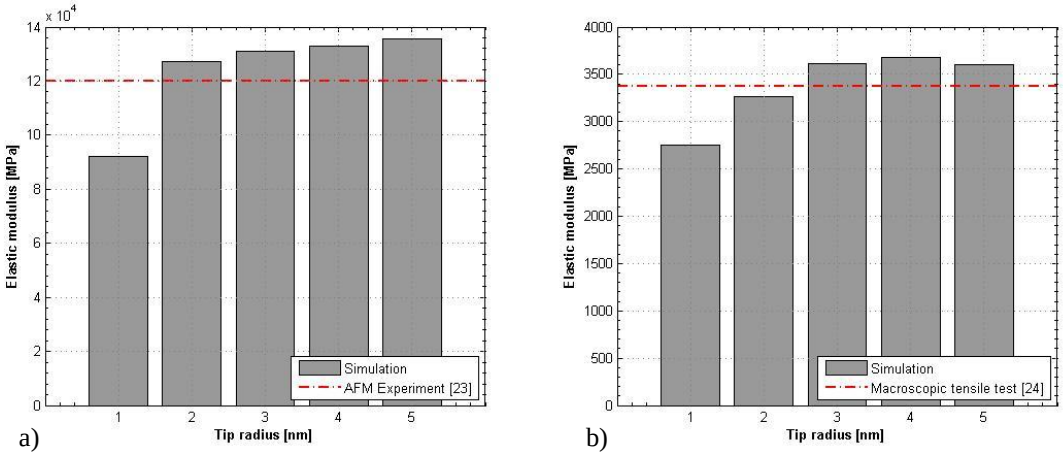


Figure 4: a) Influence of the AFM tip size on the calculated elastic modulus of pure boehmite crystals; b) Influence of the AFM tip size on the calculated elastic modulus of pure epoxy resin.

It is apparent, that with increasing tip radius the calculated moduli converge for both materials. Small tips underestimate the moduli, but already for 2 nm in case of the boehmite and for 3 nm in case of the epoxy the deviations between smaller and bigger tip sizes are smaller than 4 %. It can also be seen that the presented values are slightly above the displayed reference values from experimental investigations (13 % for the pure boehmite and 7 % for the pure epoxy). However, comparatively little literature on AFM tests on boehmite or the elastic properties of boehmite is available, making the comparison rather less reliable. The results for the pure epoxy are compared with results from macroscopic tensile tests.

The minimum sample size is determined using a similar approach. Starting with the full model size of 8 x 4.8 x 8 nm for boehmite and 15 x 9 x 15 nm for the epoxy resin, the sample size is gradually reduced to a minimum of 5 x 3 x 5 nm (boehmite) and 9 x 5.4 x 9 nm (epoxy). The aspect ratio remains the same for all samples. The simulations are performed with a 2 nm tip for the boehmite and a 3 nm tip for the epoxy. Figure 5 a) displays the resulting force-deformation curves for boehmite. As expected, with decreasing sample size, the point where the simulated curves start deviating from the Hertz curve is shifted to a lower deformation (marked by the straight lines in Figure 5 a)). Interestingly, reducing the model size from 8 to 6.5 nm and from 6.5 to 5 nm leads to an almost equal deformation shift of about 0.15 nm. The dashed simulation curve line in Figure 5 a) (denoted with Simulation "5x3 nm rcb", rcb = remove constrained bonds) represents a model, where some of the elements in the boundary region are deleted, as it is often done in MD simulations to soften the boundary conditions. All bond stretch elements and all angle bend and torsion elements with at least two nodes lying in the boundary region are deleted. With this model the deviation point can be shifted by approximately 0.09 nm, showing a good opportunity for model size reduction.

The results for pure epoxy in Figure 5 b) show significantly greater fluctuations, particularly the small models, which makes fitting the Hertz curve more difficult, especially for very small deformations. Thus, the deviation point is not as obvious as it is for boehmite. Therefore, simulations with greater tip displacements of 1.5 nm have been performed. The straight lines again represent the mark the onset of deviation from the Hertz curve. The line for the 9x5.4 nm variation is not displayed because it behaves too stiff from the beginning, whereas variation 15x9 nm shows good agreement in the whole range examined. A sample size of 12 x 7.2 x 12 nm is chosen to be the reference size for further investigations as a compromise between accuracy and efficiency. The "rcb" version has due to the fluctuations not been regarded during the following.

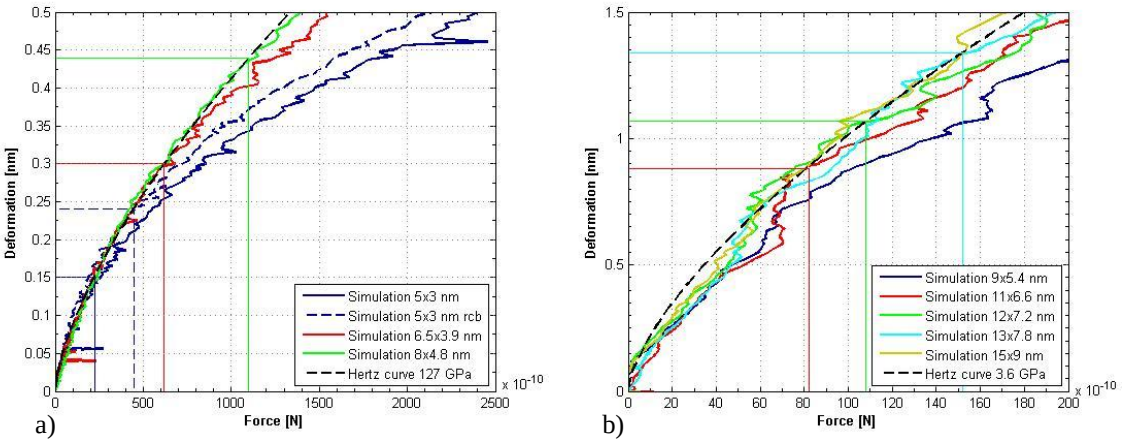


Figure 5: a) Influence of the sample size on the calculated elastic modulus of the pure boehmite crystal; b) Influence of the sample size on the calculated elastic modulus of the pure epoxy.

4.2 Nanocomposite AFM Simulation Results

Based on the findings of the previous chapter, in the following, the simulation results are shown for the combined particle-matrix samples. The tip size for the following models is chosen to be 3 nm and the sample size is 12 x 7.2 x 12 nm. The corresponding results are displayed in Figure 6. For each of the curves presented, 15 virtual AFM experiments are performed in a line in radial direction between 0 and 40 Å originating on the particle surface as illustrated in Figure 6 a). The region close to the particle surface can, as in real AFM exper-

iments, not be characterized, since depending on the tip size, tip flank and particle overlap. Thus a mix of particle and interphase modulus is measured in this case. Therefore, the particle surface appears to be shifted outward. Hence, in Figure 6 b) the region of 15 Å around the particle is neglected. Comparing the obtained curves of the interphase stiffnesses, the differences are clearly visible. The reference configuration, where the particle is not chemically but only physically bonded to the matrix, exhibits a constant stiffness and thus no visible interphase in the examinable region. This leads to the conclusion, that the particle does not influence the cross-linking in the surrounding matrix. Figure 6 c) shows the aforementioned behavior more demonstratively. It displays a stiffness plot, for which not the complete surface is screened, as in real AFM experiments, but the curve depicted in Figure 6 b) is rotated around the center of the particle. The coloring indicates the stiffness, with brighter regions being stiffer as darker regions. The red dashed circle displays the real particle surface. For the reference configuration, no interphase is visible. Variation 2 (hydroxyl groups on the particle surface are enabled to react during curing reaction, see Figure 6 d)) in contrast shows a clearly visible interphase. The region between 18 and 38 Å originating on the particle surface exhibits reduced stiffnesses by an amount of approximately 1 GPa (corresponding to a reduction of about 30 %). Further away from the particle surface the stiffness reaches the magnitude of the undisturbed matrix material. Variation 3 (with the pre-etherification step, see Figure 6 e)) exhibits an even more distinct interphase. The average reduction of the matrix stiffness is about 1.5 to 2 GPa (approximately 50 %). At the border of the used models the interphase stiffness is still reduced and thus the size of the interphase could not be determined.

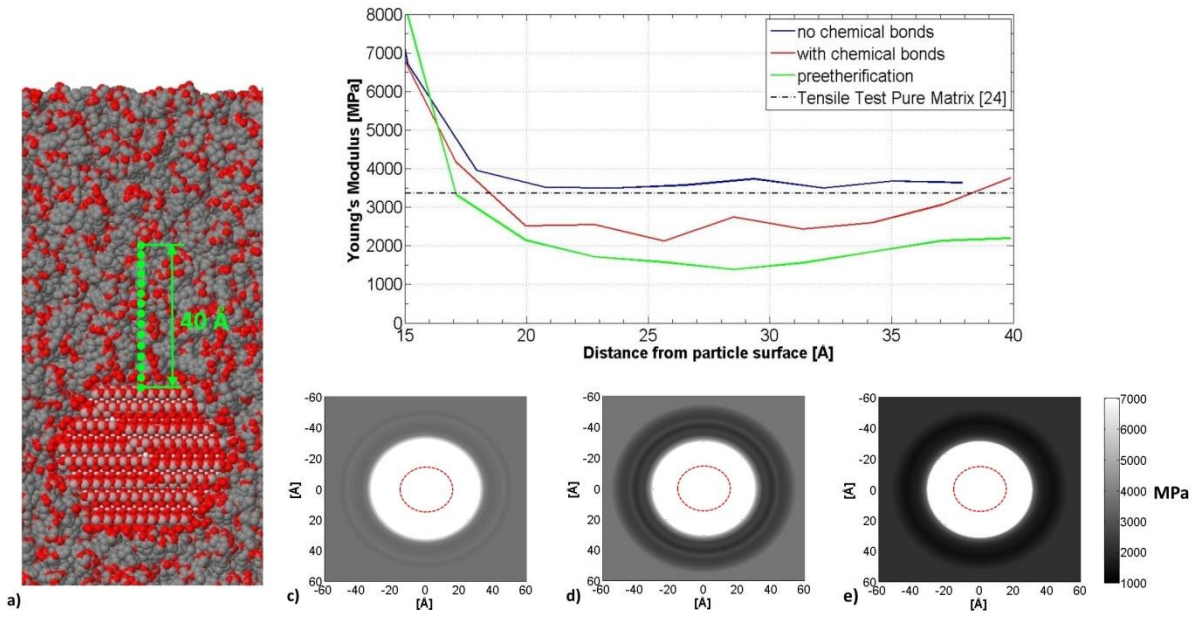


Figure 6: a) AFM measuring points; b) distribution of the interphase stiffness in dependency of the particle distance; c) stiffness plot without chemical interactions; d) stiffness plot with chemical interactions; e) stiffness plot with pre-etherification step (the red dashed circle indicates the real particle surface).

These results and their origins become more distinct in the context of the curing statistics of the underlying models (see Table 1). All models contain the same amount of epoxy groups and the same amount of hardener molecules is available for the cross-linking reaction. This results from the same particle geometry and weight fraction, leading to an equal amount of packed prepolymer chains. The values presented in Table 1 generally differ only by small

amounts, because the statistics refer to the complete model, not just to the interphase. However, the change of the values results mainly from the changed morphology of the interphase. All models have an almost equal amount of etherified connections. Interestingly the remaining curing statistics of the models of variation 2 and 3 show an opposite trend compared to the reference variation. This indicates different underlying mechanisms. For variation 2 (with chemical bonds but without a pre-etherification step), more hardener molecules are placed, leading to fewer monoesters. This is caused by the bigger amount of available hydroxyl groups, since particle surface hydroxyls are available additionally to the epoxy hydroxyls. Therefore in total more cross-links are created, but some cross-links are created between particle and matrix instead of within the matrix. This leads to a reduction of the matrix stiffness in the vicinity of the particle. In case of the pre-etherification step, the amount of hardeners used in the esterification step is significantly reduced. At the same time the number of monoesters and the number of epoxy molecules used show the highest value of the three variations. This is caused by the effect that many hydroxyl and epoxy groups are already occupied and no longer available for reaction. This leads to the conclusion, that in this case the cross-linking is really impeded, not just shifted to other reaction sites.

variation	Agent molecules used in esterification step [%]	Number of etherified connections	Number of Monoesters	Number of epoxy molecules used [%]
1: No chemical bonds	93.81	197	705	97.45
2: With chemical bonds	93.98	196	686	97.52
3: pre-etherification	93.45	196	746	97.67

Table 1: Curing statistics of the three compared variations.

4.3 Validation of Nanocomposite Materials

Since up to now the experimental results for a one-to-one comparison to the simulated AFM curves are not available yet, a statement about the variation that most likely fits the reality is made on the basis of simulated tensile tests. Therefore, cubic unit cells with 10 wt-% boehmite particles of 3 nm particle size are created and calculated (see Figure 7 a)). For each variation, three simulations with displacement boundary conditions in the three spatial directions have been performed and averaged. The results are shown and compared to the experimental result in Figure 7. It is obvious, that variation 2 matches the experimental result best. Variation 1 and 3 however, show good agreement among each other, but yield in a softer behavior than variation 2 and the experimental curve. The Young's modulus is extracted by fitting a straight line to the simulated curves in the strain range of 0.0002 to 0.01, leading to values of 3644 MPa, 3786 MPa and 3652 MPa for the variations 1, 2 and 3. The resulting Young's modulus of the experiment is calculated in the usual way between 0.0005 and 0.0025 with a value of 3806 MPa. Variation 2 shows an excellent agreement with the experimentally measured results.

The remaining question, why variation 3 does not show an even stiffer behavior than variation 3, but behaves almost similar to variation 1 can be answered as follows. In case of no chemical interactions, the particle is bonded to the matrix in a very weak way. This prevents an appropriate load transfer from the matrix into the particle. As can be seen from the curves in Figure 6 as well as from the curing statistics, variation 3 produces a very soft interphase. This leads to the conclusion, that in this variation in fact the particle-matrix interactions are

quiet strong, but at the cost of a very weak interphase, that, analogously to the weak bonding in case of variation 1, causes an unfavorable load transfer and thus a softer elastic behavior.

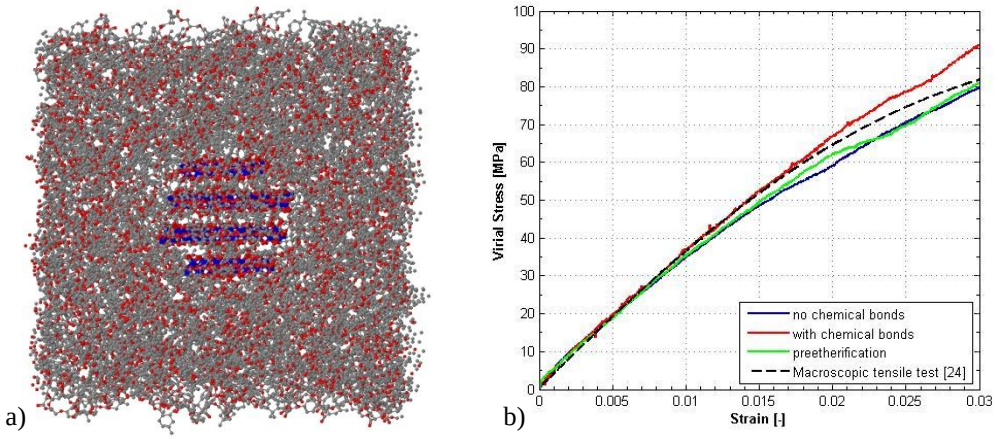


Figure 7: a) Example model for simulation of tensile test (3 nm particle, 8 nm boxsize, approx. 26000 atoms); b) Averaged stress-strain curves of a simulated tensile tests of an epoxy with 10 wt-% unmodified boehmite particles with 3 nm particle size in comparison with the experimentally obtained stress-strain curve.

5 CONCLUSION

The pure boehmite crystal and the pure epoxy have been investigated by atomic force microscopy simulations. The elastic moduli have been determined with 135 GPa for the boehmite and 3.6 GPa for the epoxy respectively. Both results show a good agreement with experimentally obtained values.

Furthermore, the influence of different particle-matrix interactions on the development of the mechanical properties in the particle matrix interphase has been studied. Three different variations were examined. Firstly, the particle was bonded by physical interactions only, secondly the surface hydroxyl groups of the boehmite particle were able to develop dative bonds during the curing reaction of the matrix. As a third variation, the surface hydroxyl groups were allowed to react with the epoxy groups of the matrix through etherification prior to the actual curing reaction. The first variation shows no visible interphase, leading to the conclusion that in this case the particle does not influence the cross-linking in the surrounding matrix. For the second variation, an interphase with a radius of approximately 38 Å and a mean stiffness reduction of approximately 30 % compared to the undisturbed matrix was obtained. Variation three shows an even more distinct interphase with a stiffness reduction of about 50% and an even bigger interphase size.

Since up to now no experimental data for the comparison to the simulated curves is available, numerical tensile test have been performed. The simulated tensile test of the variation containing chemical interactions between particle and matrix but no pre-etherification step yields in a quiet accurate prediction of the elastic behavior, whereas the variation with the pre-etherification step does not. Therefore in general the goal should be to find a balance between a good particle matrix interaction and the weakening of the interphase. Otherwise the problem of a poor load transfer is simply shifted away from the particle surface into the matrix.

The presented method can provide insight in the underlying mechanisms of nanoparticle reinforcements and significantly contribute to the optimization of nanocomposite materials, e.g. by investigating the influence of different particle surface modifications on the interphase properties. At the same time, in long term perspective, the technique is able to adequately replace AFM experiments and thus reduce the need of experimental effort. The goal of ongoing

studies is the investigation of varying surface modifications with varying parameters, like e.g. surface loading. Through parallel simulations of tensile tests, the influence of the composition and the developed mechanical properties within the interphase on the mechanical properties of the whole composite can be studied systematically. Furthermore, a direct comparison with real AFM experiments is pursued.

6 ACKNOWLEGEMENTS

This work originates from the Research Unit FOR 2021 "Acting Principles of Nano-Scaled Matrix Additives for Composite Structures". The authors would like to thank Benedikt Finke, Dr. Carsten Schilde and Professor Arno Kwade of the Institute for Particle Technology at the TU Braunschweig for providing and analyzing X-ray diffraction data of boehmite nanoparticles and Maximilian Jux and Professor Michael Sinapius of the Institute of Adaptronics and Function Integration at DLR Braunschweig for providing the experimental data of tensile tests of the pure/reinforced epoxy resin.

The Research Unit FOR 2021 is funded by the German Research Foundation (DFG). Furthermore the authors acknowledge the support by the RRZN scientific computing cluster, which is funded by the Leibniz Universität Hannover, the Lower Saxony Ministry of Science and Culture (MWK) and the German Research Foundation (DFG). The authors wish to express their gratitude for the financial support.

The pictures of the chemical unit cells and models in Figure 1, Figure 3 and Figure 7 are generated using Jmol [25], an open-source Java viewer for chemical structures in 3D.

REFERENES

[1] C. Arlt, Wirkungsweisen nanoskaliger Böhmite in einem Polymer und seinem Kohlenstofffaserverbund unter Druckbelastung. *Dissertation*, 2011.

[2] M.F. Uddin, C.T. Sun, Strength of unidirectional glass/epoxy composite with silica nanoparticle-enhanced matrix, *Composites Science and Technology*, 68(7-8), 1637-1643, 2008.

[3] A.K. Subramaniyan, C.T. Sun, Enhancing compressive strength of unidirectional polymeric composites using nanoclay, *Composites Part A: Applied Science and Manufacturing*, 37(12), 2257-2268, 2006.

[4] N. Shahid, R.G. Villate, A.R. Barron, Chemically functionalized alumina nanoparticle effect on carbon fiber/epoxy composites, *Composites Science and Technology*, 65(14), 2250-2258, 2005.

[5] N.A. Siddiqui, R.S.C. Woo, J.K. Kim, C.C.K. Leung, A. Munir, Mode I interlaminar behavior and mechanical properties of CFRPs with nanoclay-filled epoxy matrix, *Composites Part A: Applied Science and Manufacturing*, 38(2), 449-460, 2007.

[6] U. Landman, W.D. Luedke, N.A. Burnham, R.J. Colton, Atomistic Mechanisms and Dynamics of adhesion, Nanoindentation, and Fracture, *Sience*. 248, 454-461, 1990.

[7] T.-H. Fang, C.-I. Wenig, J.-G. Chang, Molecular dynamics simulation of nanolithography process using atomic force microscopy, *Surface Science*, 501(1-2), 138–147, 2002.

[8] B. Heymann, H. Grubmüller, AN02/DNP-hapten unbinding forces studied by molecular dynamics atomic force microscopy simulations, Chemical Physics Letters, 303 (1-2), 1-9, 1999.

[9] L. Nasdala, A. Kempe, R. Rolfes, The Molecular Dynamic Finite Element Method (MDFEM), *Computers, Materials & Continua*, 19(1), 57-104, 2010.

[10] L. Nasdala, A. Kempe, R. Rolfes, Are finite elements appropriate for use in molecular dynamic simulations?. *Composites Science and Technology*, 72(9), 989-1000, 2012.

[11] Q. Wang, Effective in-plane stiffness and bending rigidity of armchair and zigzag carbon nanotubes. *Int J Solids Struct*, 41, 5451-5461, 2004.

[12] H.W. Zhang, Z. Yao, J.B. Wang, W.X. Zhong. Phonon dispersion analysis of carbon nanotubes based on inter-belt model and symplectic solution method. *Int J Solids Struct*, 44, 6428-6449, 2007.

[13] S.Q. Zhu, X. Wang, Effect of Environmental Temperatures on Elastic Properties of Single-Walled Carbon Nanotube. *J Therm Stress*, 30, 1195-1210, 2007.

[14] C.Y. Li, T.W. Chou, A structural mechanics approach for the analysis of carbon nanotubes. *Int J Solids Struct*, 40, 2487-2499, 2003.

[15] M. Chwal, Influence of vacancy defects on the mechanical behavior and properties of carbon nanotubes, *Proc Eng*, 10, 1579-1584, 2011.

[16] E.I. Saavedra Flores, S. Adhikari, M.I. Friswell, F. Scarpa, Hyperelastic finite element model for single wall carbon nanotubes in tension. *Comput Mater Sci*, 50(3), 1083-1087, 2011.

[17] L. Nasdala, A. Kempe, R. Rolfes, An elastic molecular model for rubber inelasticity. *Comput Mater Sci*, 106, 83-99, 2015

[18] X. Bokhimi, J.A. Toledo-Antonio, M.L. Guzmán-Castillo, F. Hernández-Beltrán, Relationship between Crystallite Size and Bond Length in Boehmite. *J Solid State Chem*, 159(1), 32-40, 2001.

[19] B. Finke, Institute for Particle Technology, TU Braunschweig (private communication)

[20] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, and G.R. Hutchison. Open Babel: An open chemical toolbox. *J. Cheminf*, 3-33,2011

[21] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez. Packmol: A package for building initial configurations for molecular dynamics simulations. *J Comp Chem*, 30(13), 2157-2164, 2009.

[22] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations. *J Phys Chem*, 94(26), 8897-8909, 1990.

[23] R.C. Streller, Boehmite als Nanofüllstoffe für Polypropylen-Nanocomposites und Nanopartikel-modifizierte Polypropylen/Kautschuk-Blends. *Dissertation*, 2008.

[24] M. Jux, Institute of Adaptronics and Function Integration, DLR Braunschweig (private communication)

[25] Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MULTI-SCALE FINITE ELEMENT ANALYSIS OF GRAPHENE/POLYMER NANOCOMPOSITES: ELECTRICAL PERFORMANCE

Asimina Manta[1], Matthieu Gresil[1,2], and Constantinos Soutis[2]

[1] i-composites lab, School of Materials, University of Manchester, James Lighthill Building, Sackville Street, Manchester M13 9PL, United Kingdom.
e-mail: asimina.manta@postgrad.manchester.ac.uk

[2] Aerospace Research Institute, Faculty of Engineering and Physical Sciences, University of Manchester, James Lighthill Building, Sackville Street, Manchester M13 9PL, United Kingdom.
constantinos.soutis@manchester.ac.uk
matthieu.gresil@manchester.ac.uk

Keywords: Graphene, Finite Element Analysis, Electrical Conductivity, Nanocomposites, Computational methods

Abstract. *This paper presents a multi-scale, multi-physics finite element model (FEM) simulating the electrical performance of single layer graphene/insulator under DC loading, which would be easily applied to most cases, extended to a more sophisticated material architecture in respect of being scientifically structured and proven. The approach consists of the creation of a unit cell and a representative volume element (RVE) micro-scale nanocomposite model on a commercially available FE package. The implementation of these models has been considered satisfactory and successful, as the variation of nanocomposite's electrical conductivity in respect to the volume content was in accordance with experimental data. Moreover, the numerical simulation was in accordance with the classical percolation law predictions, while the obtained percolation thresholds in terms of aspect ratio obey the Excluded Volume Theory. The graphene shape was considered in the analysis as a geometric parameter, being proved that the shape does not exhibit any profound effect on the electrical performance of the graphene/insulator nanocomposite. Finally, this model is capable to predict the full percolation response of the nanocomposite and can be applied to any nanocomposite architecture in contrast to general theories that can only estimate the percolation threshold rather than the full response.*

1 INTRODUCTION

Conductive polymers have been extensively studied for their potential applications in light emitting devices, batteries, electromagnetic shielding, and piezoresistive sensors. At first, carbon black [1], [2], metallic powder [3–5], polyaniline [6] and graphite [7] were used as electrical reinforcement in polymer, but high concentration was necessary to achieve the percolation threshold which endangered the mechanical properties of the nanocomposites due to the formation of agglomerations. Later, several researchers proposed polymer nanocomposites reinforced with graphene nanoparticles and its derivatives (expanded graphite, graphene nanoplatelets, graphite oxide, functionalized graphene/expanded graphite), which are able to form more stable 3D conductive networks in lower volume content as a consequence of their high aspect ratio (AR-ratio of main particle dimension to minor one) [8–15].

However, despite the high financial and time cost of material preparation and experimental work, there are a few computational models predicting the electrical response of graphene/polymer nanocomposites. Most of them are based on molecular dynamics and geometrical programming routines, being able mainly to predict the percolation threshold of the nanocomposites with high computational cost. Oskouyi et al. [16] appeared to be the first to apply the method of Monte Carlo model to study the percolation threshold for disk-shaped fillers, simulating the conductive network formed by inclusions like graphene nanoplatelets. Later, Hicks et al. [17] developed a tunnelling-percolation model to investigate electrical transport in graphene-based nanocomposites, covering the need of a suitable model able to predict the full electrical response of semiconducting 2D element reinforced materials. This model, however, is applicable only to rectangle shaped nanoparticles forming 2D networks, while in common graphene reinforced nanocomposites, fillers exhibit a wide range of shapes, mainly circular-ellipsoid ones, and the conductive network formed is considered to be a 3D one. In simulation work [18], the percolation threshold for circular and ellipsoid fillers was investigated for a 2D network, while in [19] the percolation threshold of composites filled with intersecting circular disks and particles of various morphology was studied under a 3D Monte-Carlo simulation. Otten et al. [20] developed an analytical approach to investigate the percolation behaviour of polydisperse nanofillers, specifically focusing on the percolation threshold's sensitivity on the polydispersity in length, diameter and the level of conductivity of mainly needle-like filler nanocomposites. After the work of Hicks et al. [17] on the variation of electrical conductivity in respect of the volume fraction, Ambrosetti et al. [21] approached the electrical conductivity of an insulating matrix reinforced with conductive ellipsoids of revolution by assuming that an expected curve of electrical conductivity variation would be applied and finally being reduced in a geometrical form taking into account the inter-particle distance and the tunnelling distance. Mathew et al. [22] studied the percolation threshold of hard platelets in a 3D continuum system in terms of isotropic-nematic transition, while Hashemi et al. [23] proposed a continuum model developed, which embodied the most fundamental characteristics of the graphene nanocomposites (percolation threshold, interface effects and additional contribution of electron hopping, microcapacitor structures to interfacial properties) to determine the effective AC and DC electrical properties of graphene nanocomposites.

Knowing that, such numerical models are not easily applicable on industrial and engineering case study, as their employment is complicated and their results are limited to certain case, in this paper it is presented a multi-scale multi-physics finite element model (FEM) which would be easily applied to most cases, extended to more sophisticated material architecture in respect of being scientifically structured and proven. The current approach on multi scale modelling consists of the creation of a unit cell and a micro-scale nanocomposite model

(MSNM - Representative Volume Element RVE) on a commercial finite element modelling environment. The unit cell consists of a single rectangle graphene plate surrounded by a thin layer of polymer. This polymer layer represents the inter-plate volume between successive graphene reinforcements, in which conduction phenomena (tunnelling effect, electron hopping etc.) take place. The unit cell is loaded with a constant electric potential to compute the resistance matrix representing this system. The MSNM is a rectangle block, on whose resistance matrix elements previously obtained through the unit cell is randomly distributed. This distribution represents the random position of graphene on the bulk volume of polymer nanocomposites. The orientation of graphene is simulated by the 3D random orientation of the corresponding element local coordinate system.

2 FINITE ELEMENT MODELLING

2.1 Unit Cell

In similar terms with the approach of hard core-penetrable shell has been assumed on several studies predicting the percolation threshold [17], [24], the unit cell consists of a single-layer graphene plate surrounded by a thin layer of polymer. From this description, three main geometrical parameters could be derived-graphene's plate shape and size (expressed by the Aspect Ratio defined as the ratio of the length to the width) and the polymer layer thickness d_t. Graphene is modelled as an isotropic electrical conductive material with electrical conductivity of $\sigma=10^7$S/m, while the polymer layer represents the polymer volume between consecutive graphene, where the effect and the electron hoping phenomenon takes place. As a consequence, the conduction mechanism on polymer volume would be simulated by applying an exponentially varied resistivity ρ_{tunnel} expressed by the equation (1) in accordance to the findings of [25] for the electric tunnel effect between similar electrodes separated by a thin insulating film.

$$\rho_{tunnel} = \frac{h^2}{e^2\sqrt{2m\lambda}} \exp\left(\frac{4\pi d}{h}\sqrt{2m\lambda}\right) \quad (1)$$

Since it has been assumed that the conductive phenomena taking place between two successive graphene plates would be taking into account in the unit cell analysis, it is necessary to modify the application of the tunnelling resistivity in the unit cell, so as to simulate the contribution of a single plate to the conduction. In many studies the cut-off distance, otherwise the maximum inter-particle distance above this the tunnelling effect is eliminated, was considered around 2 nm [17], [21], [26], leading to the polymer thickness range between 0 and 1.0 nm. In addition to this, it is crucial to note that the architecture of graphene plates forms conductive paths inside the insulating polymer, which could be represented as a typical 2D or even 3D resistor network. As a consequence, if the resistance of each unit cell was used to from a network, the result would be to add the resistance of two graphene plates and the tunnelling resistance assigned to each plate for every pair of plates. However, due to the fact that the tunnelling resistance is an exponential function of tunnelling distance, the sum of assigned tunnelling resistance of each plate would not be equal to the actual resistance. This observation implied the need to modify the equation (1) by multiplying it with the equation (2), so as to be able to simulate the actual tunnelling resistance when formed in a 3D real nanocomposite structure. In the proposed function (2), it is noticeable that the definition of the tunnelling resistivity applied on the unit cell would not be a deterministic variable but a stochastic one. This statistical feature, then, should be transferred to the next scale and be applied on the representative volume implying complicated distributions on material properties. In order to

eliminate this effect, it is proposed to use an equivalent constant C in combination with the modified equation (3) of tunnelling resistivity. The constant C would be given by the equation (4), depicted in Figure 1 and its equivalent value would be the result of the equation (5). The equivalent value of C for each case is presented in Table 2, where it is possible to conclude that assuming C being equal to 0.3 would be reasonable to represent each case.

$$\frac{R_{tunnel}(d_1+d_2)}{R_{tunnel}(d_1)+R_{tunnel}(d_2)}=\frac{d_1}{d_1+d_2}\exp\left(\frac{-4\pi d_2}{h}\sqrt{2m\lambda}\right)+\frac{d_2}{d_1+d_2}\exp\left(\frac{-4\pi d_1}{h}\sqrt{2m\lambda}\right) \quad (2)$$

$$\rho_{tunnel}=C_{equivalent}\frac{h^2}{e^2\sqrt{2m\lambda}}\exp\left(\frac{8\pi d}{h}\sqrt{2m\lambda}\right) \quad (3)$$

$$R_{tunnel}(d_1+d_2)=C\left(R_{tunnel}(d_1)\exp\left(\frac{4\pi d_1}{h}\sqrt{2m\lambda}\right)+R_{tunnel}(d_2)\exp\left(\frac{4\pi d_2}{h}\sqrt{2m\lambda}\right)\right) \quad (4)$$

$$C_{equivalent}(d_{\max}-d_{\min})=\int_{d_{\min}}^{d_{\max}}C(\delta)d\delta \quad (5)$$

Plank's constant	h	m^2kg/s	$6.62607004\cdot10^{-34}$
Electron charge	e	Coulomb	$1.60217662\cdot10^{-19}$
Electron mass	m	kg	$9.10938356\cdot10^{-31}$
Height of barrier	λ	eV	0.5-2.5
Polymer layer thickness/Tunnelling distance	d_t	nm	0.0-1.0

Table 1: Tunnelling resistivity variables and parameters

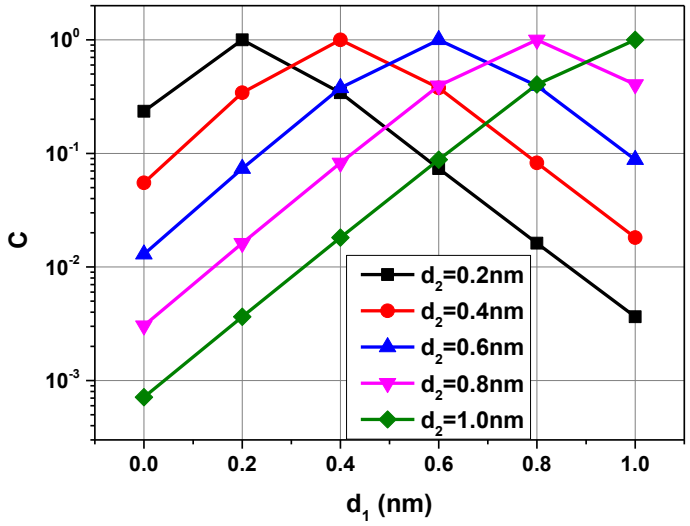


Figure 1: Variation of constant C in function of polymer thickness surrounding graphene layer.

d_2 (nm)	$C_{equivalent}$
0.2	0.37
0.4	0.37
0.6	0.38
0.8	0.34
1.0	0.20

Table 2: $C_{equivalent}$ in function of d_2.

In terms of finite element analysis, the proposed model was built in commercial FEA program ANSYS 16.2. A 3D 20-Node Hex couple-field solid element was used with its piezoresistive behaviour activated through the available key-option 101, while both for graphene and polymer isotropic electrical behaviour was assumed. It is important to mention that the piezoresistive coupling induced by this key-option was neglected by not defining the piezoresistive matrix. The values of the parameters used on the analysis of the unit cell are presented on the Table 3 and Table 4. Unit cell is loaded in the longitudinal, transverse and through thickness direction by a constant voltage, simulating DC loading and resulting to the calculation of the resistance matrix by the use of Ohm's law.

Graphene Shape		Rectangular	Ellipse
Graphene plate thickness	t (nm)	0.45	0.45
Graphene minor side dimension	a (nm)	10	2·5.64
Aspect Ratio	AR (-)	1,5,10,50	1,2,5,7,10
Graphene major side dimension	b (nm)	AR·a	AR·a

Table 3: Graphene geometrical parameters

Material	Electrical behaviour		
Graphene	Electrical conductivity	S/m	10^7
Polymer	Tunnelling resistivity	Ωm	Equation (3)
	$C_{equivalent}$	-	0.3
	Height of barrier (λ)	eV	0.5-2.5
	Tunnelling distance (d)	nm	0.2-1.0

Table 4: Unit cell material properties

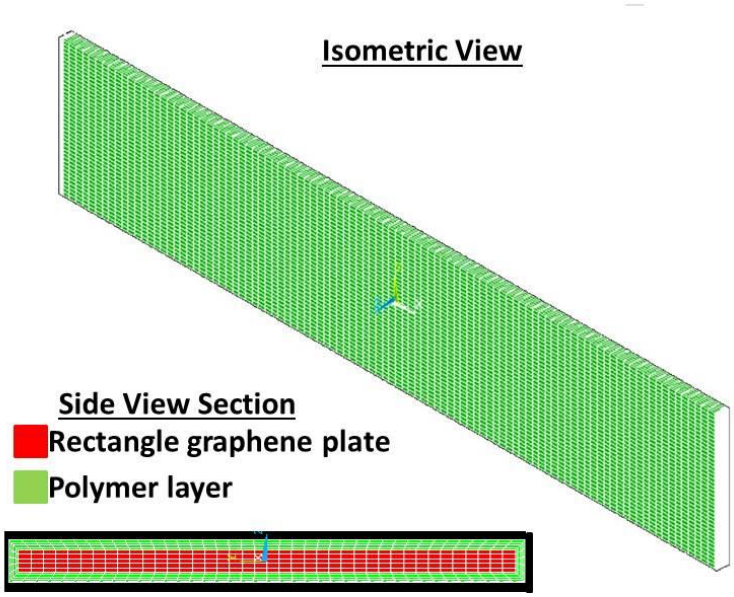


Figure 2: Unit Cell finite element model.

2.2 Representative Volume Element (RVE)

The nanocomposite structure was approached by a square block of side equal to multiplied b (k·b) and thickness of multiplied a (k·a). The number of graphene seeds on the block is defined by equations (6) and (7) as a function of nanocomposites volume fraction and graphene's geometry. The number of elements representing each graphene layer (γ) would equal to the number of division in the direction of loading (eix) divided by the factor k. As a consequence the number of elements representing graphene reinforcement in total in the volume (n_g) would be given by the equation (8). The number of elements in the perpendicular to loading direction (eiz) would be given by the equations (9) and (10), while the number of elements through the block's thickness (eiy) would be eiz/AR.

Number of graphene seeds in respect of the graphene shape		
Square - Rectangle	$n = V_f \frac{k^3 ARa}{t}$	(6)
Circle - Ellipse	$n = V_f \frac{k^3 ARa}{\pi t}$	(7)

Table 5: Equations describing the number of graphene seeds in respect of the graphene shape.

$$n_g = \gamma n = n\frac{eix}{k} \quad (8)$$

Number of element on perpendicular to loading direction in respect of graphene shape		
Square - Rectangle	$eiz = kAR\sqrt{\frac{a}{t}}$	(9)
Circle - Ellipse	$eiz = kAR\sqrt{\frac{a}{\pi t}}$	(10)

Table 6: Equations describing the number of elements on perpendicular to loading direction in respect of the graphene shape.

The nanocomposite's architecture is approached by the distribution of the graphene plates being enforced by choosing n_g non-repeatable integer numbers. Each number indicates the element ID number in the FE model. Each selected element has the electrical properties resulted by the analysis of the unit cell. Especially, assuming DC linear electrical behaviour, the electrical response of the unit cell in each element is introduced in accordance to equations presented in Table 7. The tunnelling distance between two successive graphene particles is assigned by random choice of material types obtained by unit cell and related to a specific tunnelling distance. Finally, the orientation of graphene layers in the volume is approached by the orientation of the element local coordinate system with rotation angles of $THXY \in [0,90]$, $THZY \in [0,90]$ and $THXZ \in [0,90]$. The elements, which have not been chosen to represent graphene sheets, are simulated as pure insulating matrix with electrical resistivity of $10^{16}\Omega$m. The RVE model was built up under the same principles as the unit cell one. The volume is loaded with constant voltage, while the result of the analysis is the reaction current. With the use of the Ohm's law and under the conduction of 50 separate analysis per case, the variation of electrical conductivity in respect of volume fraction is created, while in the probability function of percolation the inflection point is calculated to be the percolation threshold in accordance to the suggested procedure of [22].

Electrical resistivity distributed in the RVE in respect of the direction		
x-axis	$\rho_x = R_x^{unitcell} \frac{k^2 ARa}{eiz^2}$	(11)
y-axis	$\rho_y = R_y^{unitcell} \frac{k^2 ARa}{eix^2}$	(12)
z-axis	$\rho_z = R_z^{unitcell} \frac{k^2 ARa}{eix^2}$	(13)

Table 7: Equations for the electrical resistivity introduced to the elements representing graphene particles in respect of the direction.

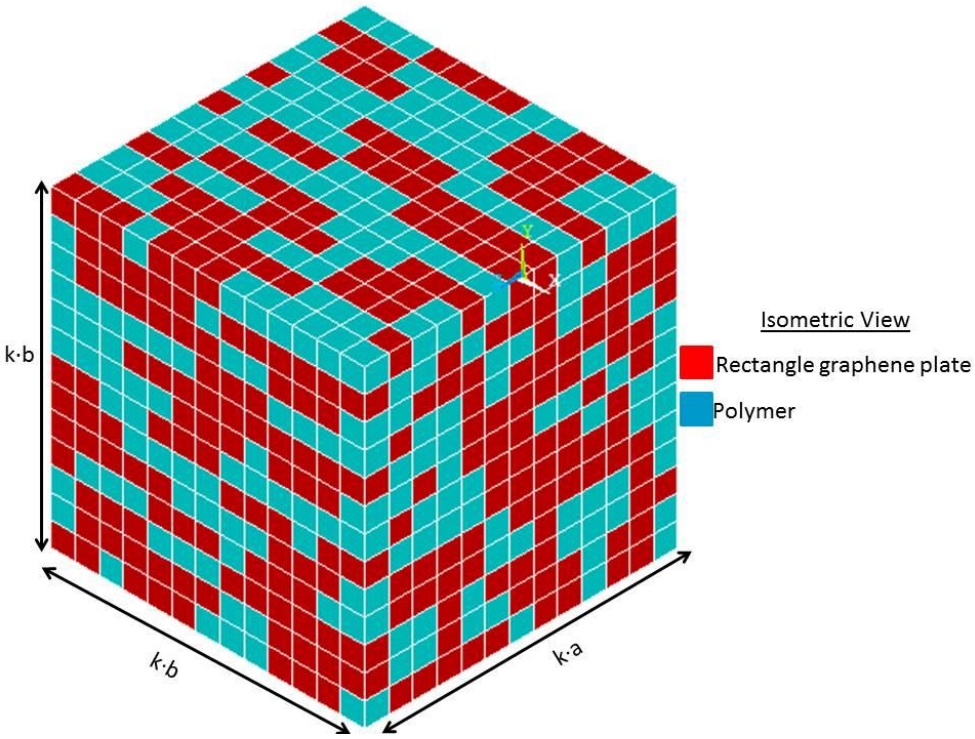


Figure 3: Representative Volume Element finite element model.

3 RESULTS

3.1 Unit cell analysis

After the analysis of the unit cell in respect of all the parameters mentioned on Table 3 and Table 4, it is possible to make some notes on the effect of these feature on the unit cell resistance, which could be assumed to represent the local resistance shown on a nanocomposite structure. As far as the effect of the aspect ratio is concerned, in Figure 4 this effect is depicted in respect of the loading direction and tunnelling distance for rectangle shaped graphene and height of barrier of 0.5eV. It could be noted that the rise of the aspect ratio leads to significant decrease in the resistance, as the resistance of the graphene has been the dominant conduction mean in the unit cell. Especially in the case of aspect ratio being greater than 50, in the longitudinal direction of the unit cell (x-direction) the resistance saturates to a constant value in function of the tunnelling distance.

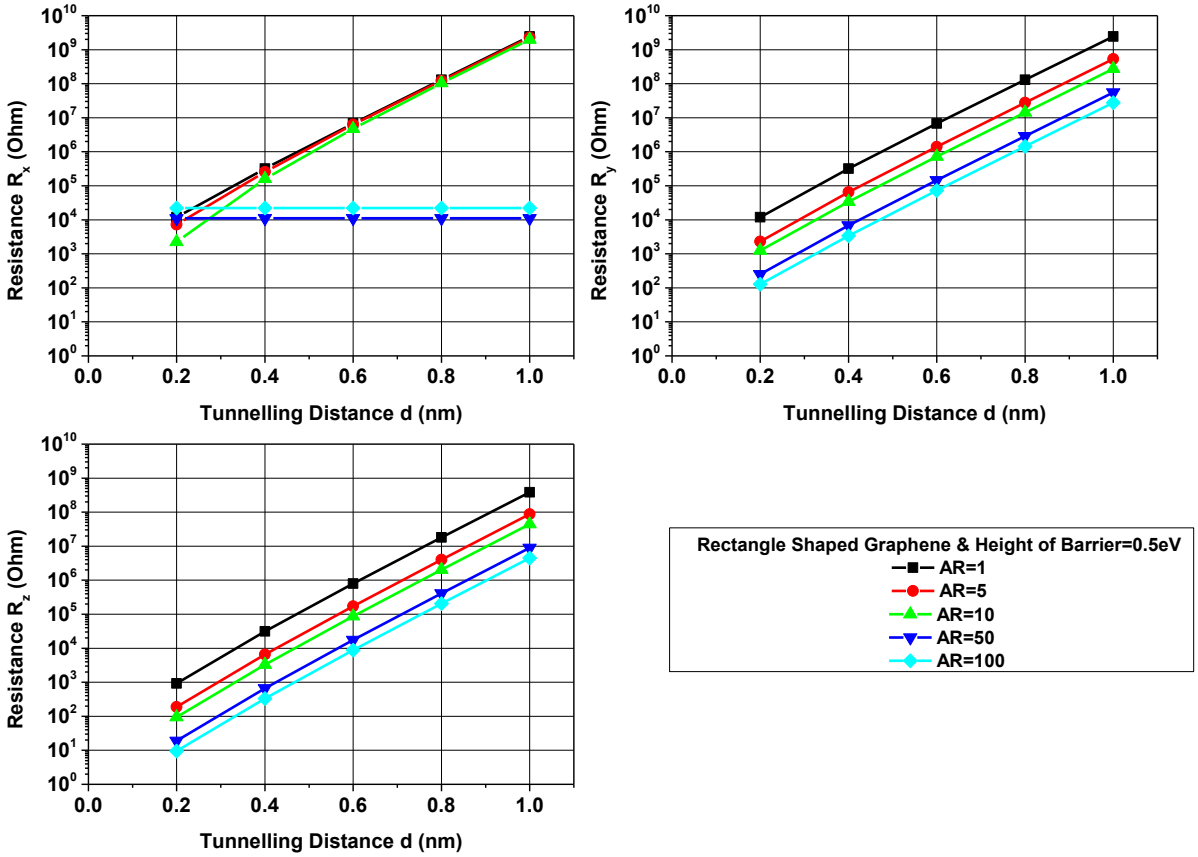


Figure 4: Effect of aspect ratio on unit cell resistance for the case of rectangular shaped graphene and height of barrier of 0.5eV in respect of the direction and the tunnelling resistance.

Polymer and all the features controlling its response under any excitation (degree of polymerization, voids, humidity, temperature etc.) have serious effect on the electrical conductivity of a nanocomposite through the energy quantity of height of barrier appearing on the tunnelling resistivity. As it is thoroughly explained in [25], when two conductive particles are separated by an insulating film (polymer film in this case), the equilibrium conditions require that the top of the energy gap of the insulator to be positioned above the Fermi level of the conductive particles. As a result, the action of the insulating film is to introduce a potential barrier between the electrodes which impedes the flow of the electron between the conductive particles. This description of the physical system of conductive particles separated by an insulator suggests that electronic current would flow through the insulating region between the conductive particles if the electron had enough thermal energy to overcome the potential barrier and flow in the conduction band or even if the barrier was that low to permit its penetration by the electric tunnel effect. This potential barrier is described in this study as height of barrier and usually takes values of a few electron volts (0.5-2.5eV). In this part of our study, the effect of height barrier on the unit cell resistance in respect of tunnelling distance and for the case of rectangle shaped graphene with aspect ratio of 1 and 5 is depicted in Figure 5. The rise of the height of barrier has as a consequence the rise of the local resistance even in orders of magnitude. This increasing trend is common for every case of unit cell studied, while becomes more evident with the increase of tunnelling distance.

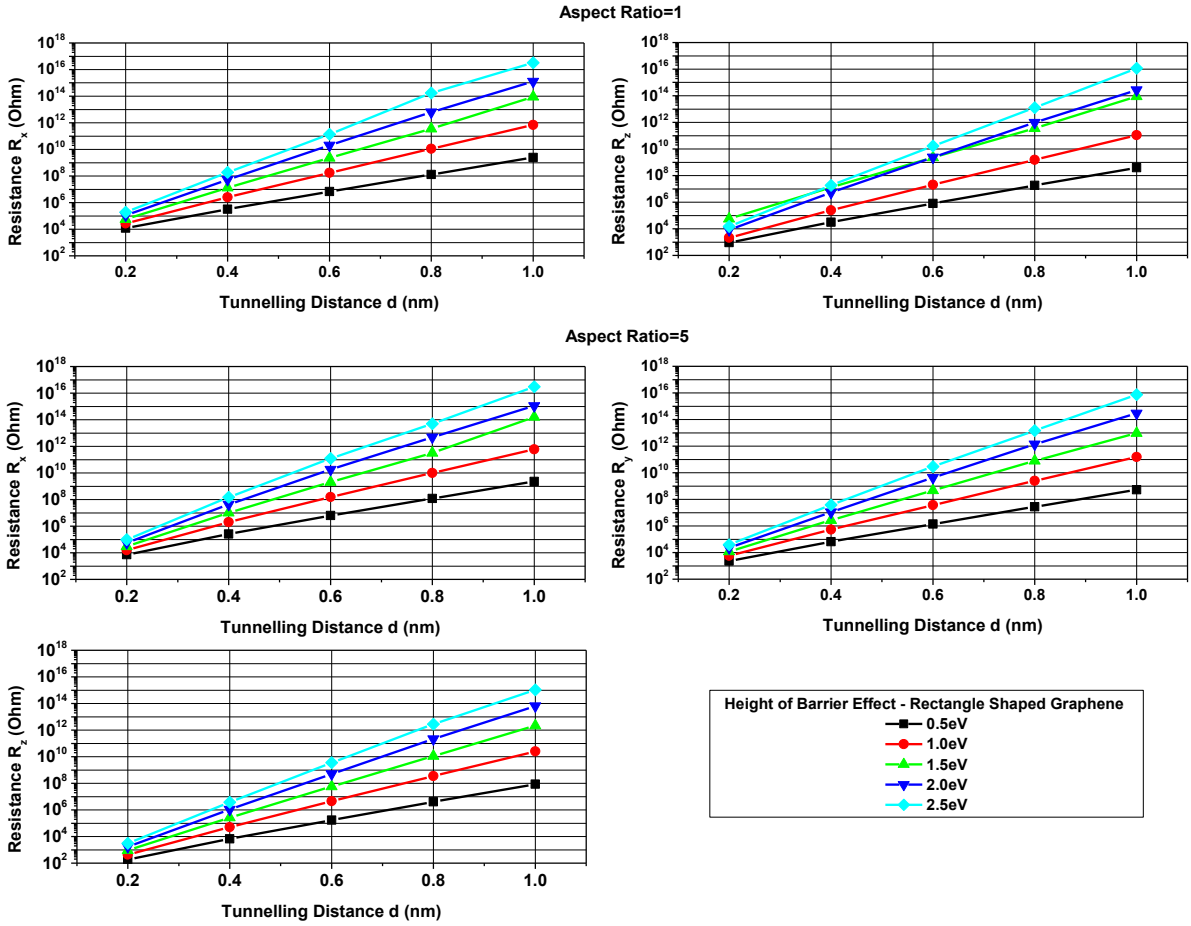


Figure 5: Effect of height of barrier on unit cell resistance for the case of rectangular shaped graphene filler and aspect ratio of 1 & 5 in respect of the direction and the tunnelling resistance.

Finally, the shape of the graphene layer is considered. In Figure 6, the effect of graphene shape on the unit cell resistance in respect of the direction and the tunnelling distance is depicted for all examined aspect ratios and height of barrier of 0.5eV. It could be stated that ellipse shaped graphene sheets exhibit higher local resistance in the main axis directions than the rectangle shaped ones. However, for the case of the through thickness resistance, for the common aspect ratios ellipse and rectangle shaped graphene plates show the same electrical resistance. It is important to mention that for the purposes of comparison, the area of the rectangle shaped graphene layers is equal to that of the elliptical shaped graphene layers for each aspect ratio.

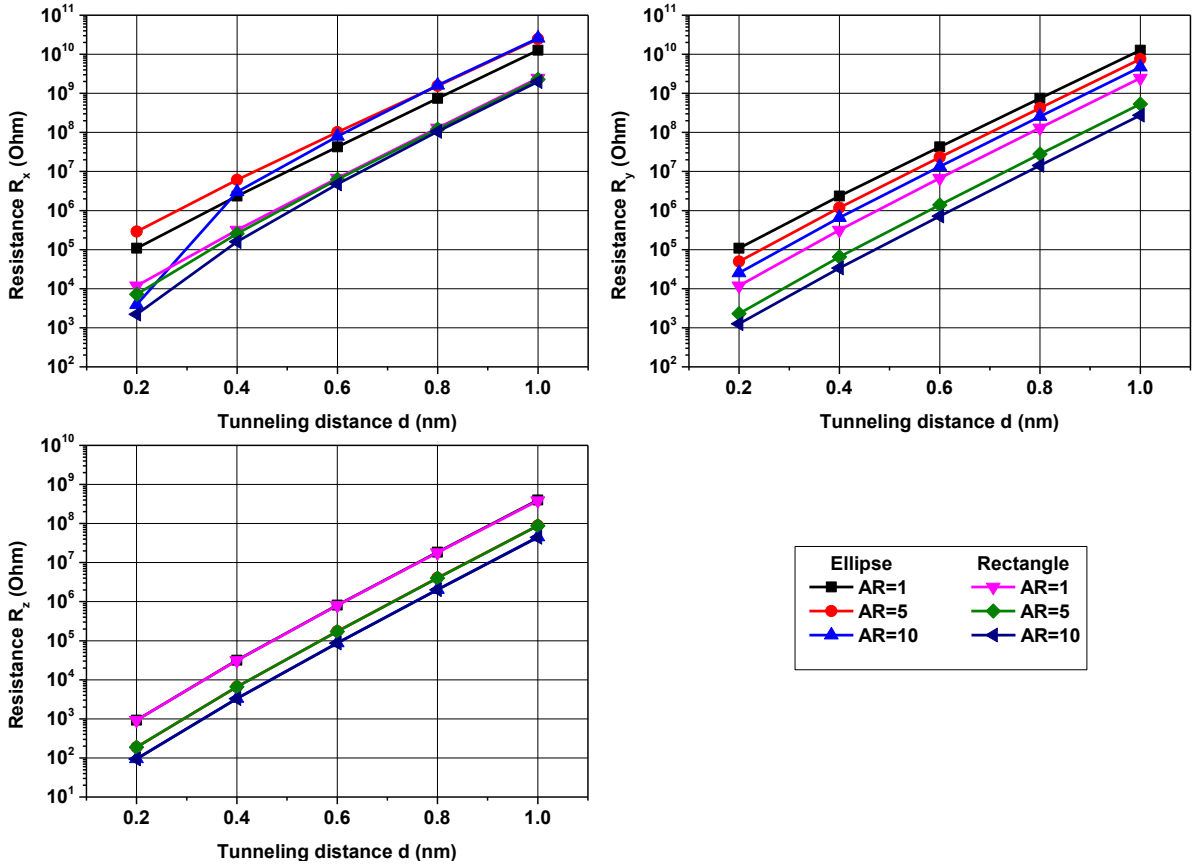


Figure 6: Effect of graphene shape on the unit cell resistance for all examined aspect ratios, height of barrier of 0.5eV in respect of the direction and the tunnelling resistance.

3.2 RVE analysis

Since it is not reasonable to model infinite material and in order to eliminate the edge effect, it is important to conduct a convergence study on the size of the square block. Having considered as value identical to the nanocomposite structure the ratio of the average of the electrical conductivity of the sample to its standard deviation, as it could be noticed in the Figure 7, for k=5 the structure of the nanocomposite has converged.

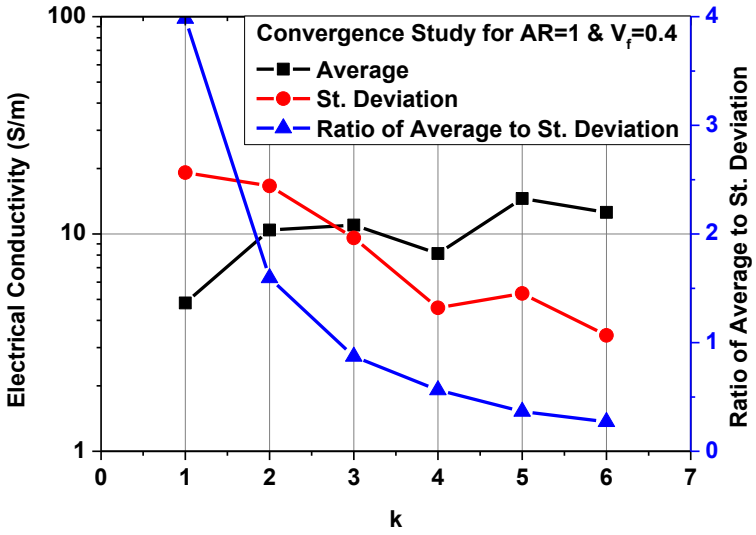


Figure 7: Convergence study on the size of the representative volume for the case of the graphene filler with aspect ratio=1 and V_f=0.4.

As it has been already stated the percolation probability, in accordance to the procedure suggested by the [22], is fitted by the function (14) and the inflection point of this function is considered to be the percolation threshold. In Figure 8, the percolation probability in respect of the volume fraction is depicted for the case of rectangle graphene sheets and λ=0.5eV. As it could be noted, the proposed fitting function is suitable for describing this kind of data set achieving 99.999% accuracy on fitting process.

$$P(V_f) = \frac{1}{2}\left(1 + \tanh\left(A\left(V_f - B\right)\right)\right) \qquad (14)$$

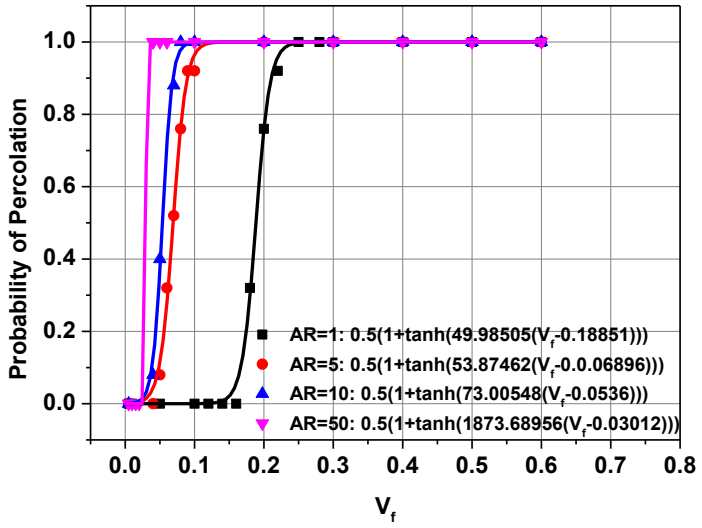


Figure 8: Percolation probability in respect of volume fraction for studied graphene aspect ratios for the case of λ=0.5eV and the corresponding functions.

In Figure 9, the effect of aspect ratio on the variation of electrical conductivity in respect of the graphene volume fraction for the case of rectangle shaped filler. It is notable that, the rise in aspect ratio leads to decrease on the volume fraction on which the onset of conductance occurs. In addition to this, for a constant value of volume fraction, nanocomposite's electrical conductivity increases with rising aspect ratio. These results suggest that fillers with high aspect ratio are able to form more stable and efficient conductive network in the volume of nanocomposite in lower volume fraction. The above conclusion is supported also by the fact that with the rise of volume fraction the deviation on the statistical sample is significantly reduced leading to more uniform sample close to uniquely defined electrical conductivity for a nanocomposite of a specific volume fraction.

In Figure 10, the effect of height of barrier on nanocomposite's electrical conductivity is depicted for the case of rectangle shaped graphene with aspect ratio equal to 1 and 5. Although, the percolation threshold seems no to be affected by the height of barrier, since the onset of percolation is mainly geometrically oriented as it has been proven by [23], for the case of constant volume fraction the rise of height of barrier causes rapid decrease on the nanocomposite's electrical conductivity even by a decrease of an order of magnitude in the case of increasing the height of barrier by 0.5eV.

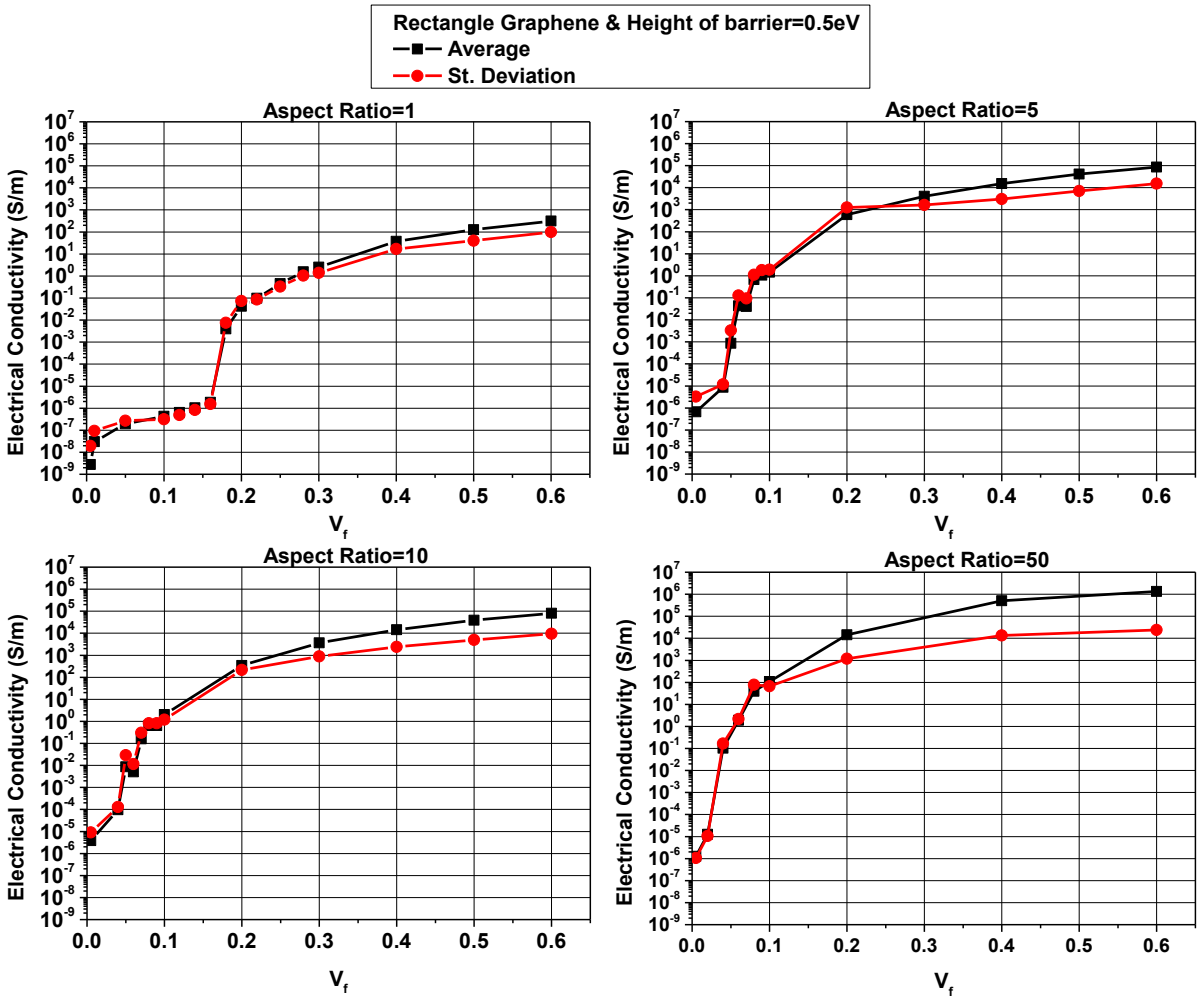


Figure 9: Effect of aspect ratio on the electrical conductivity in respect of volume fraction V_f for rectangle shaped graphene and height of barrier λ=0.5eV.

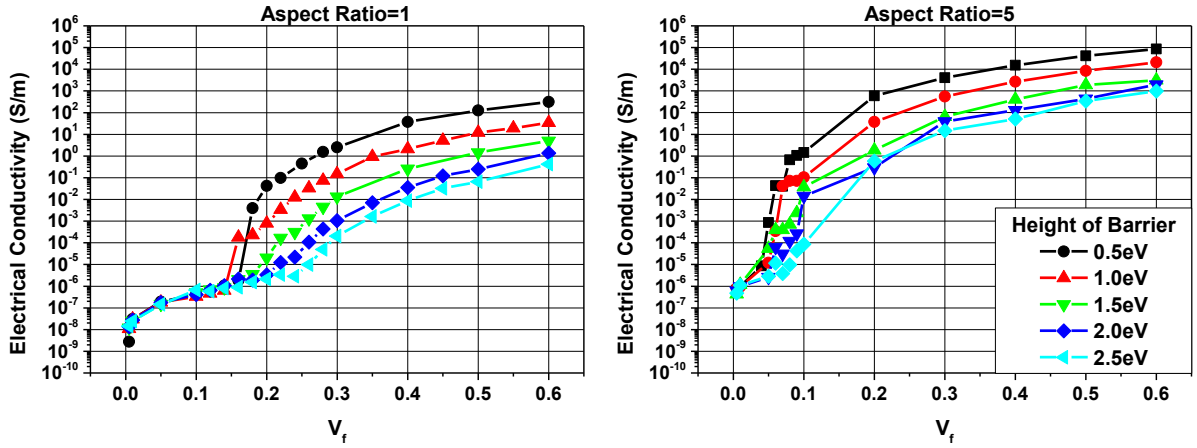


Figure 10: Effect of height of barrier on the electrical conductivity in respect of volume fraction V_f for rectangle shaped graphene and aspect ratios of 1 &5.

In Figure 11 the effect of the graphene shape is depicted in respect of the volume fraction for the case of height of barrier being 0.5eV. The graphene shape does not affect the trend of the nanocomposite electrical response in function of the V_f, although for graphene particles of the same aspect ratio and volume, the nanocomposite reinforced with rectangle shaped exhibits higher electrical conductivity than that of a nanocomposite reinforced with ellipse shaped graphene particles, while the rectangle shaped graphene particles show lower percolation threshold than that obtained for the ellipse graphene. [27] suggested that this observation could be explained as the corner angles of squares and rectangles make it easier for the plates to touch each other, therefore enhancing the current passing from the one particle to another.

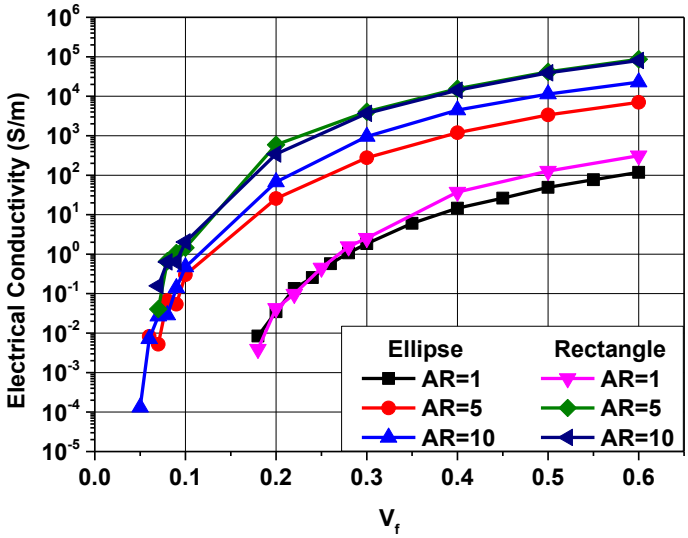


Figure 11: Effect of graphene shape on the electrical conductivity in respect of the volume fraction V_f for height of barrier of 0.5eV.

3.3 Percolation model

For the description of the percolation behaviour of the nanocomposites, several theories have been suggested with more popular the percolation law, which describes the variation of

the electrical conductivity of the nanocomposite in respect of the volume fraction after the percolation threshold having being obtained, and the excluded volume theory, which is able to predict the percolation threshold of a nanocomposite in respect of the geometrical features of the reinforcement (dimensions, shape, 2D-3D conductive networks).

Taking into consideration the excluded volume theory and assuming that the each particle could be considered equivalent to a circular one, the excluded volume theory for the circular disk [28] could be modified so as to define the upper and lower boundaries of the volume fraction in which the percolation threshold should be expected (15). In Figure 12, it could be obvious that the percolation threshold is reduced with increasing aspect ratio, the values were successfully found to lie between the assumed percolation bounds. In addition to this, the shape of the graphene seems not to show any effect on the percolation threshold, since the corresponding values for each case are in good agreement.

$$\left.\begin{array}{l} 1-\exp\left(-1.4\dfrac{V}{V_{ex}}\right) \le V_c \le 1-\exp\left(-2.8\dfrac{V}{V_{ex}}\right) \\ V=\pi\left(\dfrac{d}{2}\right)^2 t \\ V_{ex}=\pi^2\left(\dfrac{d}{2}\right)^3 \end{array}\right\} \xrightarrow{\frac{d}{2}=\alpha\sqrt{\frac{AR}{\pi}}} 1-\exp\left(-1.4t\frac{\sqrt{\pi}}{\alpha\sqrt{AR}}\right) \le V_c \le 1-\exp\left(-2.8t\frac{\sqrt{\pi}}{\alpha\sqrt{AR}}\right) \quad (15)$$

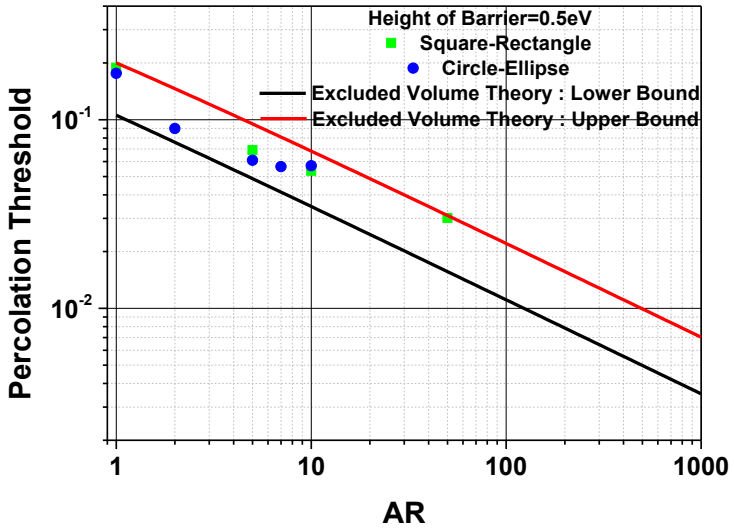


Figure 12: Percolation Threshold in respect of the graphene shape and boundary functions (15).

The percolation law is a power equation (16) introduced in [29], which is able to describe to electrical behaviour of an insulating material reinforced with conductive particles after the percolation threshold. In the equation (16) σ_c is the electrical conductivity of the composite, V_f is the volume fraction, V_p is the percolation threshold, σ_o is a factor with value close the electrical conductivity of the conductive phase measured in units of electrical conductivity and t is the critical exponent. For a three dimensional conductive network, t usually takes value between 1.65 and 2.0, which is accepted as a universal value. Higher values of t indicate

the presence of tunnelling phenomena and the filler having extreme geometries (i.e. high aspect ratio). In Figure 13, the fitting of the finite element model results in accordance to the percolation law is presented, showing a good correlation, while in Figure 14 the effect of the geometry of the filler and the height of barrier on the percolation law variables is depicted. As far as the σ_o is concerned, it is increased with rising aspect ratio converging to the filler electrical conductivity, while it is fallen with increasing height of barrier. The increase on aspect ratio leads to more stable electrical networks which are mainly governed by the fillers conductance and not the tunnelling resistance formed in the interparticle region, leading σ_o to converge to the electrical conductivity of the filler. On the other hand, the rise of height of barrier which induces a rise on the tunnelling resistance in the interparticle volume, has as a consequence the decrease of the σ_o for the case of low aspect ratios. One of the most significant findings related to the percolation law is the agreement of the observation to the general law regarding the critical exponent. An increase on the aspect ratio (defining an extreme geometry) or even any rise of the height of barrier (defining a more intense tunnelling phenomenon) has a result an important increase of the value of the critical exponent far above the prediction for a common three dimensional conductive network.

$$\sigma_c = \sigma_o \left(V_f - V_p\right)^t \tag{16}$$

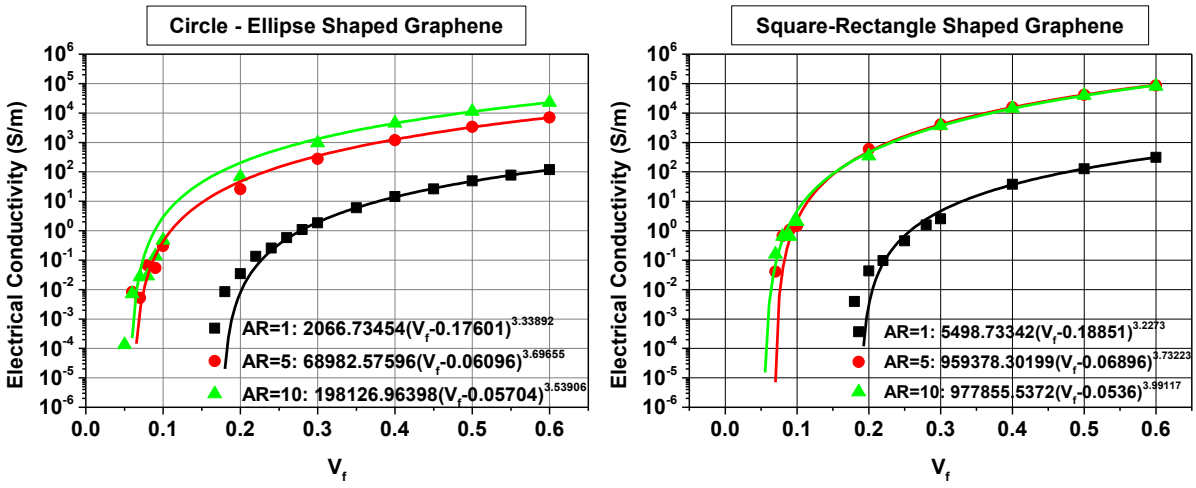


Figure 13: Percolation Law in respect of the aspect ratio and the shape of the graphene particle for the case of height of barrier of 0.5eV.

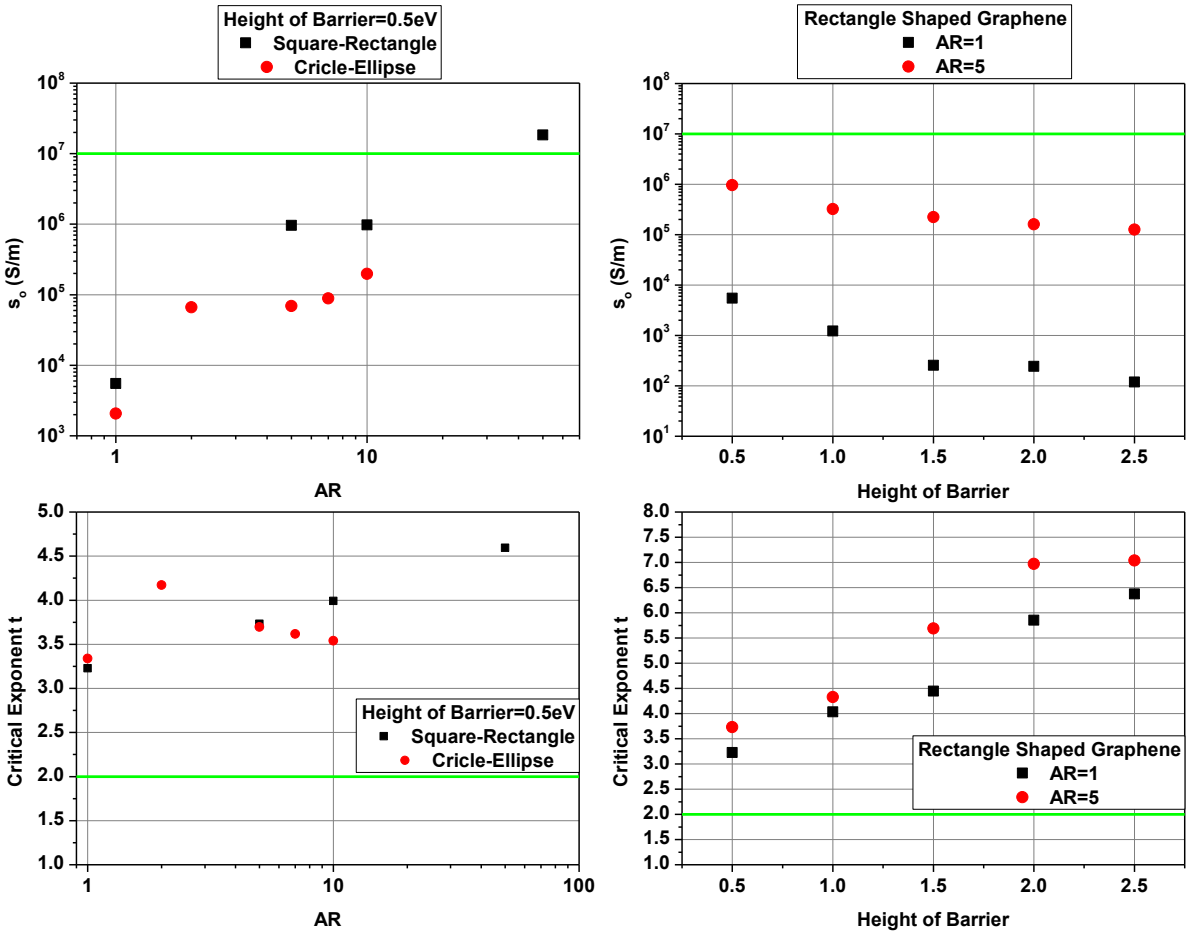


Figure 14: Variation of the percolation law variables in respect of the aspect ratio, the shape of graphene and the height of barrier.

4 CONCLUDING REMARKS

- The finite element model has been shown to effectively simulate the electrical response of the selected graphene/polymer composite under DC loading.
- The results are in accordance to theoretical predictions.
- Increasing the Aspect Ratio reduces the percolation threshold and increases the electrical conductivity of the nanocomposite for a given value of volume fraction.
- The tunnelling resistance exhibited in the inter-particle volume affects the overall performance of the nanocomposite, especially due to the height of barrier, whose rise increases the inter-particle resistance and decreases the electrical conductivity of the nanocomposite.
- The shape of the graphene fillers do not show any significant effect in terms of percolation, but the formation of sufficient contact between particles for the charge transfer is enhanced for the case of rectangular shaped graphene.

REFERENCES

[1] E. K. Sichel, J. I. Gittleman, and P. Sheng, "Electrical properties of carbon-polymer composites," *J. Electron. Mater.*, vol. 11, no. 4, pp. 699–747, Jul. 1982.

[2] Y. Ishigure, S. Iijima, H. Ito, T. Ota, H. Unuma, M. Takahashi, Y. Hikichi, and H. Suzuki, "Electrical and elastic properties of conductor-polymer composites," *J. Mater. Sci.*, vol. 34, no. 12, pp. 2979–2985.

[3] G. Pinto and A. Jiménez-Martín, "Conducting aluminum-filled nylon 6 composites," *Polym. Compos.*, vol. 22, no. 1, pp. 65–70, Feb. 2001.

[4] V. . Roldughin and V. . Vysotskii, "Percolation properties of metal-filled polymer films, structure and mechanisms of conductivity," *Prog. Org. Coatings*, vol. 39, no. 2–4, pp. 81–100, Nov. 2000.

[5] L. Flandin, J. Y. Cavaillé, G. Bidan, and Y. Brechet, "New nanocomposite materials made of an insulating matrix and conducting fillers: Processing and properties," *Polym. Compos.*, vol. 21, no. 2, pp. 165–174, Apr. 2000.

[6] S. S. Ray and M. Biswas, "Water-dispersible conducting nanocomposites of polyaniline and poly(N-vinylcarbazole) with nanodimensional zirconium dioxide," *Synth. Met.*, vol. 108, no. 3, pp. 231–236, Feb. 2000.

[7] A. Quivy, R. Deltour, A. G. M. Jansen, and P. Wyder, "Transport phenomena in polymer-graphite composite materials," *Phys. Rev. B*, vol. 39, no. 2, pp. 1026–1030, Jan. 1989.

[8] W. Zheng, S.-C. Wong, and H.-J. Sue, "Transport behavior of PMMA/expanded graphite nanocomposites," *Polymer (Guildf).*, vol. 73, pp. 6767–6773, 2002.

[9] W. Zheng and S.-C. Wong, "Electrical conductivity and dielectric properties of PMMA/expanded graphite composites," *Compos. Sci. Technol.*, vol. 63, pp. 225–235, 2003.

[10] G. Chen, W. Weng, D. Wu, and C. Wu, "PMMA/graphite nanosheets composite and its conducting properties," *Eur. Polym. J.*, vol. 39, pp. 2329–2335, 2003.

[11] H. Kim, Y. Miura, and C. W. Macosko, "Graphene/Polyurethane Nanocomposites for Improved Gas Barrier and Electrical Conductivity," *Chem. Mater*, vol. 22, pp. 3441–3450, 2010.

[12] Y. Song, J. Yu, L. Yu, F. E. Alam, W. Dai, C. Li, and N. Jiang, "Enhancing the thermal, electrical, and mechanical properties of silicone rubber by addition of graphene nanoplatelets," *JMADE*, vol. 88, pp. 950–957, 2015.

[13] C. Yun, Y. Feng, T. Qiu, J. Yang, X. Li, and L. Yu, "Mechanical, electrical, and thermal properties of graphene nanosheet/aluminum nitride composites," *Ceram. Int.*, vol. 41, pp. 8643–8649, 2015.

[14] M. H. Al-Saleh, "Electrical and mechanical properties of graphene/carbon nanotube hybrid nanocomposites," *Synth. Met.*, vol. 209, pp. 41–46, 2015.

[15] E. Kandare, A. A. Khatibi, S. Yoo, R. Wang, J. Ma, P. Olivier, N. Gleizes, and C. H. Wang, "Improving the through-thickness thermal and electrical conductivity of carbon fibre/epoxy laminates by exploiting synergy between graphene and silver nanoinclusions," *Compos. PART A*, vol. 69, pp. 72–82, 2015.

[16] A. B. Oskouyi and P. Mertiny, "Monte Carlo model for the study of percolation thresholds in composites filled with circular conductive nano-disks," *Procedia Eng.*, vol. 10, pp. 403–408, 2011.

[17] J. Hicks, A. Behnam, and A. Ural, "A computational study of tunneling-percolation electrical transport in graphene-based nanocomposites," *Appl. Phys. Lett.*, vol. 95, no. 21, pp. 213103:1–213103:3, 2009.

[18] W. Xia and M. F. Thorpe, “Percolation properties of random ellipses,” *Phys. Rev. A*, vol. 38, no. 5, pp. 2650–2656, Sep. 1988.
[19] L. Vovchenko and V. Vovchenko, “Simulation of percolation threshold in composites filled with conducting particles of various morphologies,” *Materwiss. Werksttech.*, vol. 42, no. 1, pp. 70–74, Jan. 2011.
[20] R. H. J. Otten and P. Van Der Schoot, “Connectivity percolation of polydisperse anistropic nanofillers,” *J. Chem. Phys.*, vol. 134, pp. 094902:1–094902:15, 2011.
[21] G. Ambrosetti, C. Grimaldi, I. Balberg, T. Maeder, A. Danani, and P. Ryser, “Solution of the tunneling-percolation problem in the nanocomposite regime,” *Phys. Rev. B*, vol. 81, no. 15, p. 155434, Apr. 2010.
[22] M. Mathew, T. Schilling, and M. Oettel, “Connectivity percolation in suspensions of hard platelets.,” *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, vol. 85, no. 6 Pt 1, p. 061407, Jun. 2012.
[23] R. Hashemi and G. J. Weng, “A theoretical treatment of graphene nanocomposites with percolation threshold, tunneling-assisted conductivity and microcapacitor effect in AC and DC electrical settings,” *Carbon N. Y.*, 2016.
[24] G. Ambrosetti, N. Johner, C. Grimaldi, A. Danani, and P. Ryser, “Percolative properties of hard oblate ellipsoids of revolution with a soft shell,” *Phys. Rev. E*, vol. 78, pp. 061126:1–061126:11, 2008.
[25] J. G. Simmons, “Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film,” *J. Apllied Phys.*, vol. 34, no. 6, pp. 1793–1803, 1963.
[26] A. B. Oskouyi, U. Sundararaj, and P. Mertiny, “Tunneling Conductivity and Piezoresistivity of Composites Containing Randomly Dispersed Conductive Nano-Platelets,” *Materials (Basel).*, vol. 7, pp. 2501–2521, 2014.
[27] Y. B. Yi and E. Tawerghi, “Geometric percolation thresholds of interpenetrating plates in three-dimensional space,” *Phys. Rev. E*, vol. 79, no. 04, pp. 041134:1–041134:6, 2009.
[28] I. Balberg, C. H. Anderson, S. Alexander, and N. Wagner, “Excluded volume and its relation to the onset of percolation,” *Phys. Rev. B*, vol. 30, no. 7, pp. 3933–3943, 1984.
[29] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, Revised Se. London: Taylor & Francis, 1994.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

THE BOOTSTRAP APPROACH TO THE STATISTICAL SIGNIFICANCE OF PARAMETERS IN RSM MODEL

Jacek Pietraszek[1], **Leszek Wojnar**[2]

[1] Department of Software Engineering and Applied Statistics
Faculty of Mechanical Engineering, Cracow University of Technology
37 Jana Pawla II, 31-864 Krakow, Poland
e-mail: pmpietra@mech.pk.edu.pl

[2] Chair of Computer Science and Biocybernetics
Faculty of Mechanical Engineering, Cracow University of Technology
37 Jana Pawla II, 31-864 Krakow, Poland
e-mail: leszek.wojnar@gmail.com

Keywords: Bootstrap, Statistical Significance, Response Surface Model, ANOVA, Materials Science, Biomechanics.

Abstract. *Since G.E.P. Box introduced central composite designs in early fifties of 20th century, the classic design of experiments (DoE) utilizes response surface models (RSM), however usually only in the simple form of low-degree polynomials. The typical procedure assumes the normal distribution of the noise and uses the least square method (LSQ) to identify parameters of the model with a priori assumed structure. The terms of model are repeatedly eliminated in the specific backward stepwise regression, while three indicators (the least significance of parameters, the significance of the lack of fit and the conformity of residuals with the normal distribution) are simultaneously observed to make decision to stop or to continue elimination procedure.*

Practically, in the case of small size datasets, the conformity with the normal distribution has very weak reliability and it leads to very uncertain assessment of parameters statistical significance. The bootstrap approach with simulation-based identification of parameters confidence intervals (CI) appears to be better solution than theoretically proved but only asymptotically equal t-distribution based evaluations.

The case study presented in this paper utilizes data obtained during investigation on compression vertebral fractures prediction based on computer tomography (CT) and microtomography (μCT) images. The significant difference in a resolution between these two class of images leaded to different prediction models. The small sample size (23 compressed and scanned vertebraes) and the high dimensionality of detected properties imposed the necessity of an alternative approach to the analysis, other than classic one derived with a requirement of the normality.

1 INTRODUCTION

The research investigation in a materials science always leads to the datasets. The large (or even huge) datasets should be analyzed to obtain summary results expressed in qualitative or quantitative forms. Such conclusions should be semi-proved by a statistical analysis.

Methods of approximation and prediction rapidly evolved in recent years. New approaches came from the artificial intelligence area: non-parametric, data-driven, stochastic. Apart from this, the classic approach of the response surface methodology (RSM) is still very useful and popular. The significance analytical improvements have appeared in the background of RSM in recent decades: statistics based on widespread strict requirements of the normal distribution have been replaced by so-called robust statistics based on the weak requirement of the continuous distribution or even on the statistics based on numerical simulation like e.g. bootstrap approach.

The recognized phenomenon may be studied in two possible ways:

- *a priori* selected treatments are imposed to the phenomenon (through a set of controlled factors) to observe changes in phenomenon behavior – such kind of study is especially specific to engineering research and usually coordinated by the design of experiments (DoE) methodology,
- the phenomenon (one or more objects) is observed passively and its properties are collected to dataset to study regressions between them however potential cause-and-effect relationship may be unknown – such kind of study is especially specific for research on live organisms, which individual properties may vary widely.

The models typically used in the first kind of research approaches are fixed-effects or response surface [1]. The second approaches usually used random-effects or response surface [1]. Additionally, Runkler [2] distinguished three aspects of models:

(a) parametric, which structure is *a priori* known and invariant through identification process,
(b) nonparametric, which structure is a priori unknown and it is constructed during identification process (other name: *data-driven*),
(c) semiparametric [3], which is sequential combination of parametric and nonparametric models.

The identification of a model is only the first step of analysis. The second one is a statistical estimation of uncertainty to determine the sensitivity of the model and statistical significance of its terms. However this definition is intuitively clear, it requires some quantitative measure and its evaluation. Typical procedure requires some important assumptions about mutual independence of observations, a normality of residuals and so on. If these conditions are met then algorithm leads simultaneously to the analysis of variance (ANOVA) [1] and tests of significance [1] – both result in the same conclusions.

The typical test of significance checks if the tested parameter may be treated as zero. The null hypothesis H_0 states that the parameter is zero against the alternative hypothesis H_1 that the parameter is not zero. If previously mentioned assumptions are met and the least squares method was used to identification then the ratio of the parameter divided by its standard deviation is tested against t-Student distribution [1]. But the problem arises if the assumptions are not met or the reliability of tests is very weak due to small size of a sample.

The proposed solution is the replacement of the classic test of parameters significance with the equivalent bootstrap-based checking of zero existence inside the parameter confidence intervals. This solution is more practical for datasets with the unknown distribution because it does not require to meet the mentioned assumptions.

Authors have been tried [4] previously to combine a bootstrap approach with an artificial neural network approximator to analyze materials science data processed by image analysis however they have with many numerical artifacts. Such approach was more successfully for processing non-parametric statistical analysis for surface layer [5] and analysis of human vertebraes scannings [6].

The bootstrap approach appears to be a promising solution for some aspects of fuzzy statistics which is developed mainly by Buckley [7, 8]. The general workflow for such implementations has been proposed by Grzegorzewski [9]. Some preliminary investigations in the field of a design of experiments (DoE) have been made by Pietraszek [10, 11].

2 COMPUTATIONAL METHODS

Two main methods were used: the response surface model (RSM) [12] and the bootstrap method [13]. The RSM model is used to construct prediction model to fit observed strength of vertebrae with the crushing force. The bootstrap method is used to evaluate confidence intervals for effects. Finally, the existence of zero inside the intervals is inspected. If any of intervals contains zero, the null hypothesis of is rejected.

2.1 RSM model

The RSM model consist of three terms [14]:

$$R_c = b_{const} + b_{BV/TV} \cdot BVTV + b_{branches} \cdot Branches + b_{Junctions} \cdot Junctions \tag{1}$$

where:
R_c – observed strength of a vertebrae
b_{const} – constant term,
$b_{BV/TV}$ – coefficient of relative density of a trabecular bone,
$b_{branches}$– coefficient of a average number of branches in trabecular bone,
$b_{junctions}$– coefficient of a average number of junctions on the branches in trabecular bone.
All coefficients were evaluated using the ordinary least squares (OLS) [1].

2.2 Bootstrap

The bootstrap approach [13] to RSM model (Eq.1) is described by the following workflow. The original dataset **D** is a source for iteratively made drawing with replacement which results in the bootstrapped dataset $\mathbf{D}_{\mathrm{Boot}}$. Then, the parameters $\hat{b}_i$ of the model are identified from the dataset $\mathbf{D}_{\mathrm{Boot}}$ using the least squares criterion L

$$\min L(b_i) = L(\hat{b}_i) \tag{2}$$

and obtained parameters are collected until large number of iterations will be reached. Finally, the quantiles are evaluated (typically 2.5% and 97.5%) for datasets of parameters as a range of confidence intervals and at last zero existence inside confidence intervals is checked.

3 MATERIALS AND DATASET

The data (Table 1) were obtained during investigation on the development and the optimization of diagnosis methods for the estimation risk of fractures in osteoporosis based on a three-dimensional images of trabecular bones obtained *in vivo* [14].

No	BV/TV	Branches [mm^{-3}]	Junctions [mm^{-3}]	Rc [MPa]
1	0.1166	5.8627	3.4724	8.08
2	0.2341	10.2597	5.9767	18.81
3	0.1603	11.1857	6.5363	7.31
4	0.2495	9.4670	5.5925	18.61
5	0.1704	9.3460	5.5847	12.27
6	0.1959	7.8127	4.5305	16.21
7	0.0727	3.1667	1.8487	6.50
8	0.2181	8.0648	4.7760	14.76
9	0.1156	2.4981	1.4564	10.87
10	0.1459	6.0010	3.5627	8.91
11	0.1392	8.3790	4.8488	10.52
12	0.0956	3.2275	1.9075	7.13
13	0.2139	9.8098	5.8136	14.10
14	0.1061	5.2762	3.1694	6.01
15	0.1022	4.0528	2.4041	6.36
16	0.1065	7.8878	4.5475	7.11
17	0.0837	6.7055	3.8857	6.26
18	0.1546	4.7831	2.8448	10.92
19	0.1957	9.6873	5.6839	16.32
20	0.0938	5.8117	3.4954	6.48
21	0.1410	9.2607	5.4136	9.55
22	0.1407	6.8481	4.0794	9.21
23	0.1220	6.0651	3.6250	8.57

Table 1: Raw data (source [14])

4 ANALYSIS

The model was identified from the equation (1) and the dataset (Table 1). The obtained parameters and their statistics evaluated classically are presented in Table 2.

Term	Value	SE	t	p	CI	
					-95%	+95%
b_{const}	0.34	0.91	0.39	0.701	-1.55	2.26
$b_{BV/TV}$	92.61	8.08	11.45	0.000	75.69	109.53
$b_{Branches}$	7.26	3.27	2.22	0.039	0.42	14.10
$b_{Junctions}$	-13.16	5.64	-2.33	0.031	-24.98	-1.35

Table 2: Descriptive statistics of transformed data (source [14])

The ANOVA protocol leads to the classic ANOVA table which decomposes the total variation into part assigned to grouping factor and the remain assigned to all other factors grouped under name ‘error’ which should be treated rather as ‘unexplained’ than only ‘random error’. For the mentioned data, the ANOVA table (Table 3) revealed that critical p-Value is less than significance level $\alpha = 0.05$ for all terms, what means that all terms (except constant) are statistically significant.

Effect	SS	df	MS	F	p
$b_{BV/TV}$	222.04	1	222.04	131.2	0.000
$b_{Branches}$	8.34	1	8.34	4.93	0.039
$b_{Junctions}$	9.21	1	9.21	5.44	0.031
Error	32.15	19	1.69	–	–
Total	271.74	22		–	–

Table 3: ANOVA table for transformed data (source [14])

Now, the approach was switched to the bootstrap. The number of draw iterations was set to 10.000 to easy selection of quantiles from the bootstrapped dataset. After the full bootstrap procedure, the descriptive statistics were evaluated for model parameters (Table 4).

The bounds of the confidence intervals were easy identified due to the selected number of bootstrap iterations. They were values found at positions 250 and 9750 in the sorted bootstrapped results. Similarly, the bootstrapped p-Value was evaluated as relative position of sign switching inside the sorted bootstrapped results.

Effect	Parameter mean	SE	$\text{p-Value}_{\text{bootstrapped}}$ [1] / $\text{p-Value}_{\text{theoretical}}$ [2]	-95 CI	+95 CI
b_{const}	0.29	0.82	0.682 / 0.729	-1.42	1.84
$b_{BV/TV}$	91.98	8.93	0.000 / 0.000	74.46	109.42
$b_{Branches}$	7.60	3.27	0.016 / 0.031	1.57	14.57
$b_{Junctions}$	-13.72	5.66	0.010 / 0.026	-25.79	-3.39

Table 4: Parameters of fixed-effects model for transformed data
[1] bootstrapped p-Value was evaluated from relative position of sign switching inside a bootstrap table
[2] theoretical p-Value was evaluated from ratio (mean/SE) and t-Student distribution at d.o.f. = 19

The analysis of confidence intervals bounds revealed that an asymmetric exists between left and right side relative to means (Table 5). The coefficients were evaluated as studentized i.e. a quotient of a deviation and standard error, where deviation was difference between confidence interval bound and respective mean:

$$cf = \frac{\pm 95CI - mean}{SE}. \tag{3}$$

Coefficient	Parameters intervals coefficients	
	-95 CI	+95 CI
b_{const}	-2.09	1.89
$b_{BV/TV}$	-1.96	1.95
$b_{Branches}$	-1.84	2.13
$b_{Junctions}$	-2.13	1.83

Table 5: Studentized coefficients of confidence intervals for the bootstrap model

5 CONCLUSIONS

- Bootstrap approach appears to be effective computational method to identify parameters of RSM effects model and their statistical properties.
- Bootstrap approach does not require to make *a priori* inconvenient assumptions.
- Bootstrap approach is very convenient to automatize in computational workflow and further statistical postprocessing.

REFERENCES

[1] O. Kempthorne, K. Hinkelmann, *Design and analysis of experiments. Vol.1. Introduction to experimental design*. John Wiley & Sons, Hoboken, NJ, USA, 2007.

[2] T.A. Runkler, *Data Analytics. Models and Algorithms for Intelligent Data Analysis.* Vieweg+Teubner Verlag | Springer Fachmedien, Wiesbaden, 2012.

[3] J.L. Horowitz, *Semiparametric Models*. W: red. J.E. Gentle, W.K. Hardle, Y. Mori, Handbook of Computational Statistics, Springer, Heidelberg, 2012, 597-617.

[4] J. Pietraszek, A. Gadek-Moszczak, The Smooth Bootstrap Approach to the Distribution of a Shape in the Ferritic Stainless Steel AISI 434L Powders. *Solid State Phenomena* **197**, 162-167, 2013.

[5] J. Pietraszek, N. Radek, K. Bartkowiak, Advanced Statistical Refinement of Surface Layer's Discretization in the Case of Electro-Spark Deposited Carbide-Ceramic Coatings Modified by a Laser Beam. *Solid State Phenomena* **197**, 198-202, 2013.

[6] A. Gadek-Moszczak, J. Pietraszek, B. Jasiewicz, S. Sikorska, L. Wojnar, The Bootstrap Approach to the Comparison of Two Methods Applied to the Evaluation of the Growth Index in the Analysis of the Digital X-ray Image of a Bone Regenerate. *New Trends in Computational Collective Intelligence* **572**, 127-136, 2015.

[7] J.J. Buckley, *Fuzzy probability and statistics*. Springer Verlag, Heidelberg, 2006.

[8] J.J. Buckley, Fuzzy statistics: hypothesis testing. *Soft Computing* **9**, 512-518, 2005.

[9] P. Grzegorzewski, *Decision support under uncertainty. Statistical methods for imprecise data.* EXIT, Warszawa, 2006.

[10] J. Pietraszek, Fuzzy Regression Compared to Classical Experimental Design in the Case of Flywheel Assembly. *Lecture Notes in Artificial Intelligence* **7267**, 310-317, 2012.

[11] J. Pietraszek, A. Gadek-Moszczak, T. Torunski, Modeling of Errors Counting System for PCB Soldered in the Wave Soldering Technology. *Advanced Materials Research* **874**, 139-143, 2014.

[12] S. Heinz, *Mathematical Modeling*. Springer Verlag, Berlin Heidelberg, 2011.

[13] J. Shao, D. Tu, *The Jackknife and Bootstrap*. Springer, New York, 1995.

[14] E. Czerwiński, A. Gądek-Moszczak, T. Konopka, Z. Latała, R. Petryniak, Z. Tabor, L. Wojnar, J. Pietraszek, *Development and optimization of diagnosis methods for estimation risk of fractures in osteoporosis based on a three-dimensional images of trabecular bones obtained in vivo. [in Polish]*. Cracow University of Technology, Kraków, 2012.

[15] J. Pietraszek, A. Goroshko, The Heuristic Approach to the Selection of Experimental Design, Model and Valid Pre-Processing Transformation of DoE Outcome. *Advanced Materials Research* **874**, 145-149, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

INVERSE IDENTIFICATION OF THE BENDING STIFFNESS OF A BRAIDED POLYETHYLENE TWINE SUBJECT TO LARGE DEFORMATION: APPLICATION TO THE IDENTIFICATION OF THE MESH OPENING RIGIDITY OF FISHING NETS.

B. Morvan[1−2], G. Bles[2], N. Dumergue[1], and D. Priour[1]

[1] IFREMER
Plouzane, France
e-mail: {barthelemy.morvan, nicolas.dumergue, daniel.priour}@ifremer.fr

[2] ENSTA Bretagne
FRE CNRS 3744
IRDL
F-29200 Brest, France
e-mail: guilhem.bles@ensta-bretagne.fr

Keywords: Numerical model, experiments, braided twine, mesh opening rigidity, large deformation.

Abstract. *The evaluation of the mesh opening stiffness of fishing nets is an important issue in assessing the selectivity of trawls. It appeared that a larger bending rigidity of twines decreases the mesh opening and could reduce the escapement of fish. Nevertheless, netting structure is complex. A netting is made up of braided twines made of polyethylene or polyamide. These twines are tied with non-symmetrical knots. Thus, these assemblies develop contact-friction interactions. Moreover, the netting can be subject to large deformation. In this study, we investigate the responses of netting samples to different types of solicitations. Samples are loaded and unloaded with creep and relaxation stages, with different boundary conditions. Then, two models have been developed: an analytical model and a finite element model. The last one was used to assess, with an inverse identification algorithm, the bending stiffness of twines. In this paper, experimental results and a model for netting structures made up of braided twines are presented. During dry forming of a composite, for example, the matrix is not present or not active, and relative sliding can occur between constitutive fibres. So an accurate modelling of the mechanical behaviour of fibrous material is necessary. This study offers experimental data which could permit to improve current models of contact-friction interactions [4], to validate models for large deformation analysis of fibrous materials [1] on a new experimental case, then to improve the evaluation of the mesh opening stiffness of a fishing net.*

1 INTRODUCTION

Twines used in the manufacture of netting materials for trawl codends tend to be stiffer. This increased stiffness affects the mesh opening in fishing gears, and thus the ability of fish to escape. The strong influence of the codend on trawl selectivity has been demonstrated [15]. Moreover, previous studies have shown how mesh rigidity to opening affects the mechanical behaviour, and thus the selectivity of codends ([5], [16]). However, the mesh opening stiffness is quite difficult to evaluate. In fact, it depends on its definition and on the theoretical model of the netting. Generally, to assess the mesh opening stiffness, we consider that a mesh side must have the same behaviour as a beam ([16], [9], [12], [13]). With this theoretical model, it appears that the bending rigidity EI of the beam represents the mesh opening stiffness. O´ Neill demonstrated how an increase in twine bending stiffness reduces the diameter of the codend and so the mesh opening. In addition, Herrmann [5] reported that an increase in twine bending stiffness could lead to a reduction in selectivity. Likewise, Moderhak [7] theoretically demonstrated how changes in mesh size and mesh opening stiffness can impact the shape of a codend and its selectivity.

As far as the authors know, the first method implemented so far for the assessment of this stiffness was presented by Sala [16]. The method uses a prototype on which netting panels are mounted and deformed. It incorporates four tension load cells and four stepping motors and it is designed so that all twines of the netting panel have the same deformation. The bending stiffness of the netting twine is estimated from the data of forces and deformation using the beam theory. However, the prototype uses a relatively expensive and high-tech device to ensure the uniform deformation of the netting panel.

Alternative methods exist: Priour and Cognard [14] assumed that the twine bending stiffness could be given by the stiffness of a simple cantilever beam whose deformation equals the observed deformation of a netting panel subject to out-of-plan bending. However, the method does not take into account the effect of knots and requires a sample of close mesh netting. De la Prada and Gonzalez [12] offered a simple uniaxial experimental set-up, which stretches a netting sample in the T-direction of the meshes while leaving free its deformation in the N-direction. Nevertheless, they assumed that the deformation is identical in all meshes of the sample. This assumption is probably too strong because the panel is held in the vertical position during the experiment, so forces and deformation supported by meshes at the top of the panel are higher than those supported at the bottom due to gravity.

The final aim of this study is to propose a simple methodology to assess the mesh opening stiffness of fishing nets. This methodology has to be based on a simple experimental plan, which does not require expensive devices, to be easily used in laboratories and in the fishing industry.

In this paper, we investigate the responses of netting samples to different types of solicitations; samples are loaded and unloaded with creep and relaxation stages, with two different boundary conditions. Then, two models have been developed: an analytical model and a finite element model. The last one allows the simulation of the netting without uniform deformation. These models were used to assess, with an inverse identification algorithm, the bending stiffness of mesh sides. In this paper, experimental results and models for netting structures made up of braided twines are presented.

2 MATERIAL AND EXPERIMENTAL METHOD

2.1 Netting

The netting materials specifications (fibre type, twine construction) were selected according to the netting commonly used in trawl codends.

In this paper, we investigate a single twine netting made up of polyethylene (PE). This twine is made up of a core and a sheath (Fig. 1). The core and the sheath are composed of 20 and 64 fibres respectively. In the core, fibres are twisted, whereas in the sheath 16 threads made up of 4 fibres are braided. The linear density of the twine is $5590.4\ 10^{-6} \pm 10.6\ 10^{-6}\ kg.m^{-1}$, that is $5590.4 \pm 10.6\ tex$. The diameter of the fibres, measured with a digital microscope, is $302.1\ \mu m$. The diameter of the twine is $3.14 \pm 0.01\ mm$. The pitch of the braided sheath, which is the longitudinal distance required for one revolution of a thread around the twine, is $26.12\ \pm 0.62\ mm$.



Figure 1: Left: single twine netting made up of polyethylene. Right: the braided sheath.

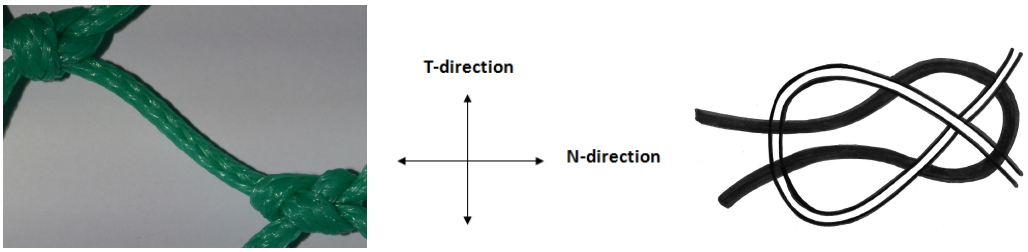


Figure 2: Left: mesh side in a single twine netting made up of polyethylene braided twines. Right: scheme of the knot.

Each 4x10 mesh netting panel was subject to a pretension step: each sample was loaded in tensile with $392.4\ kg$ in the N-direction during 30 minutes.

In order to define the input data of models, the sample characteristics are obtained:

- The mass m of the netting panel is measured using electronic scale.
- The characteristic length of the netting mesh sides is obtained from the mean value of some mesh side lengths measured with a caliper. The length of the mesh side is $40\ mm$.
- The diameter of the beam elements is calculated so that the mass of the model of the netting panel is equal to the mass of the real one. The material density of high-density polyethylene used for the calculation of the diameter is $950\ kg.m^{-3}$.

- The initial position of the netting panel (i.e. the position at rest) is derived from the length and the width of the panel at rest. The panel position at rest is performed by vibrating a horizontal plane where the netting panel is put down free of load, until a stable position is reached. Then, the dimensions (width W_{panel0}, length L_{panel0}) of the netting panel are measured and the mesh angle α_0 (angle between a mesh side and the N-direction) is derived with a simple cosinus equation using the characteristic length of the mesh sides (Fig. 3).

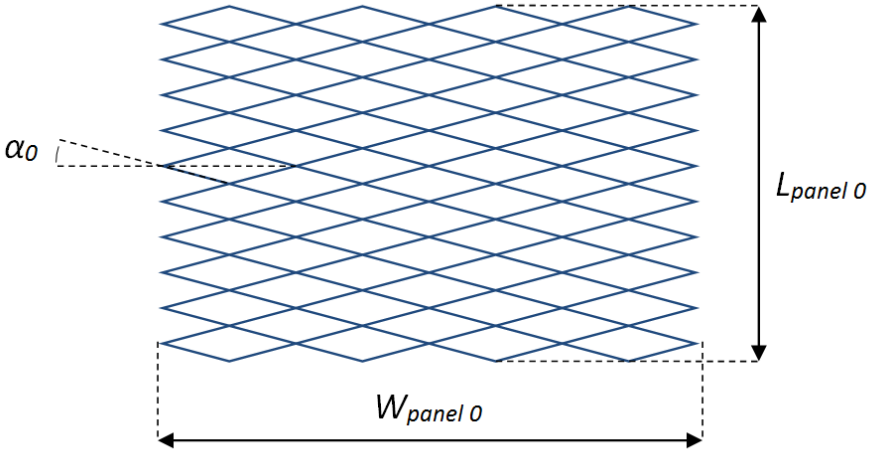


Figure 3: The dimensions W_{panel_0} and L_{panel_0} of the netting panel allow the mesh angle α_0 at rest to be calculated.

For the study, it is convenient to define some parameters. First, the T-direction and the N-direction in a mesh of the netting have been defined in [6]. The two directions are defined relatively to the knot orientation (Fig. 2 and 4).

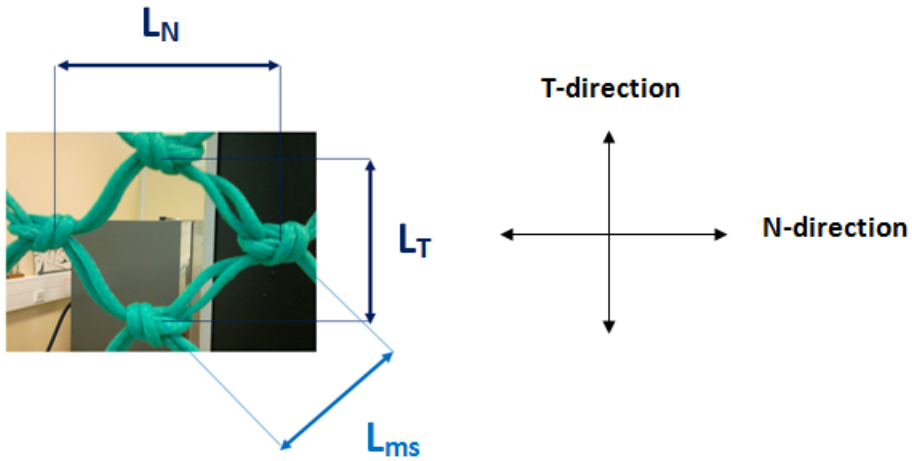


Figure 4: Definition of the N-direction and of the T-direction, and definition of the parameters L_N, L_T and L_{ms} in a diamond netting mesh.

Then, we can define, in one mesh, the distances L_N and L_T between opposite knots, in the N-direction and the T-direction respectively. L_{ms} is the length of the mesh side at rest.

To work with dimensionless parameters which do not depend on the mesh side length, we introduce the openings o_N and o_T in the N-direction and the T-direction respectively. The opening o_i, in the direction i, is the ratio of the distance L_i by the mesh side length L_{ms} (Eq. 1).

$$o_i = \frac{L_i}{L_{ms}} \tag{1}$$

2.2 Evaluation of the axial stiffness

Tensile tests have been performed on a LR5Kplus tensile testing machine to evaluate the axial stiffness of twines constituting the studied fishing nets. Because of the complex visco-elasto-plastic behaviour of the polyethylene, it is not straightforward to evaluate the axial stiffness of the instantaneous elasticity. We proposed to assess this stiffness by measuring the moduli of short-time behaviour according to [2]. G. Bles [2] noted that the measurements after relaxations and creeps and those of the initial modulus were in agreement and provided a characteristic evolution of the modulus. In our study, the strain rate was $2.10^{-4}\ s^{-1}$ and the load at the beginning of the relaxation stages increased from 100 to $900\ N$, by $100\ N$. The duration of the eight relaxation stages was 15 minutes.

We calculated the moduli of short-time behaviour by calculating the slope at the beginning of each load step, just after a relaxation stage (Fig. 5). The evolution of the moduli with the logarithmic strain in the case of the studied braided twine is shown in figure 5. For a small logarithmic strain (< 0.055), we will suppose that the modulus of short-time behaviour, which means the axial stiffness, is inferior to $14000\ N$.

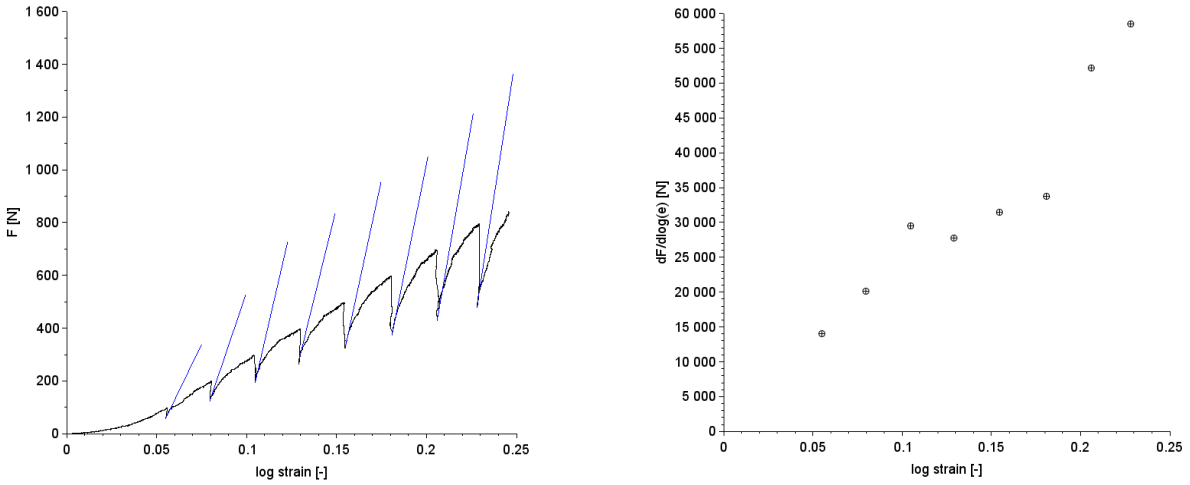


Figure 5: Left: the black line presents the load as a function of the logarithmic strain, and the blue segments indicate the slope (short-time modulus) just after each relaxation stage. Right: evolution of the short-time modulus, obtained with the slopes in the figure on the left.

2.3 Experiments

Two types of experiments were performed: a tensile test, on a classical testing machine, and a suspending test of the same type as [11].

2.3.1 Tensile test

In the first type, measurements were performed on a LR5Kplus tensile testing machine, with a 250 N load cell, of the company Lloyd instruments. The uniaxial tension tests were controlled by the jaw displacement. The relaxation stages were performed by blocking the jaw movement. A LASERSCAN 200 non-contacting extensometer (Lloyd instruments) allowed the measurement of the height of the mesh in the middle of the sample (Fig 6).

Concerning the load, we measured the sum of the effects of the displacement, the weight of the netting panel and the weight of the device which allowed the fixation of the sample. We defined the loads Fm_N and Fm_T as the loads applied on one mesh in the N-direction and the T-direction respectively. So, knowing the load F applied on the netting panel, the weight P_{panel} of the sample and the weight P_{device} of the device, we calculated the **force by mesh** Fm_T applied on the mesh in the middle of the netting sample:

$$Fm_T = \frac{F - \frac{P_{panel}}{2} - P_{device}}{4} \tag{2}$$

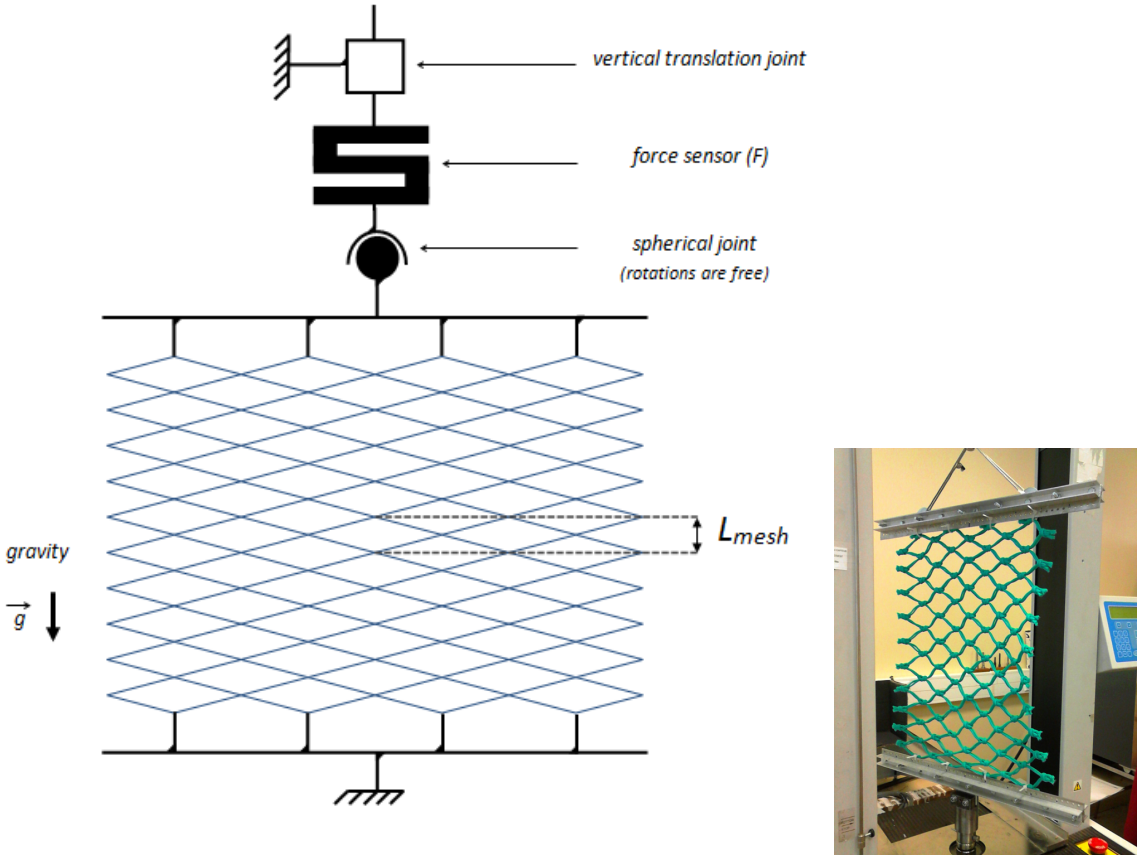


Figure 6: Plan (left) and photograph (right) of the experimental set up of the tensile test.

2.3.2 Suspending test

In the second type of experiment, the panel was suspended from one of its ends and was subject to its own weight and to forces applied on the four nodes at the bottom (Fig. 7).

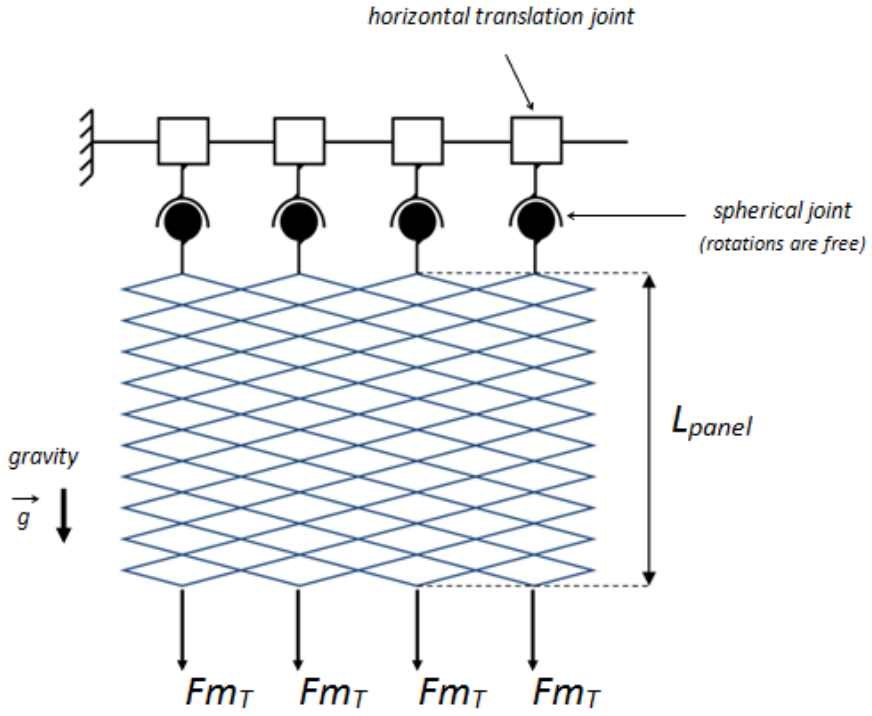


Figure 7: Experimental set up plan of the suspending test. A panel is suspended from one of its ends and is subject to its own weight and to forces applied on its bottom end.

Because of the possible variation in the results, 10 samples of 4 x 10 mesh netting panels were tested in the same conditions and loadings.

When a panel was suspended, the positions of all the nodes of the netting panel were measured. The use of a camera with a software designed and implemented in the laboratory allowed the recording of pictures with a chosen frequency. Note that the software allowed the application of optical corrections. Then, targets defined by the user on the first picture were identified in all the pictures. So the displacements of all these targets were measured during the recording. The height of the netting panel L_{panel} is measured using the position of the uppermost knot and the position of the lowermost knot.

3 EXPERIMENTAL RESULTS

3.1 Tensile test

Results of a tensile test are shown on figures 8 and 9. A load and unload cycle interrupted by several relaxation stages was imposed. The controlled jaw displacement rate was $100\ mm.min^{-1}$. The duration of the 9 relaxation stages was 15 minutes.

Figure 10 presents the result of a cyclic tensile test on a double twine netting sample.

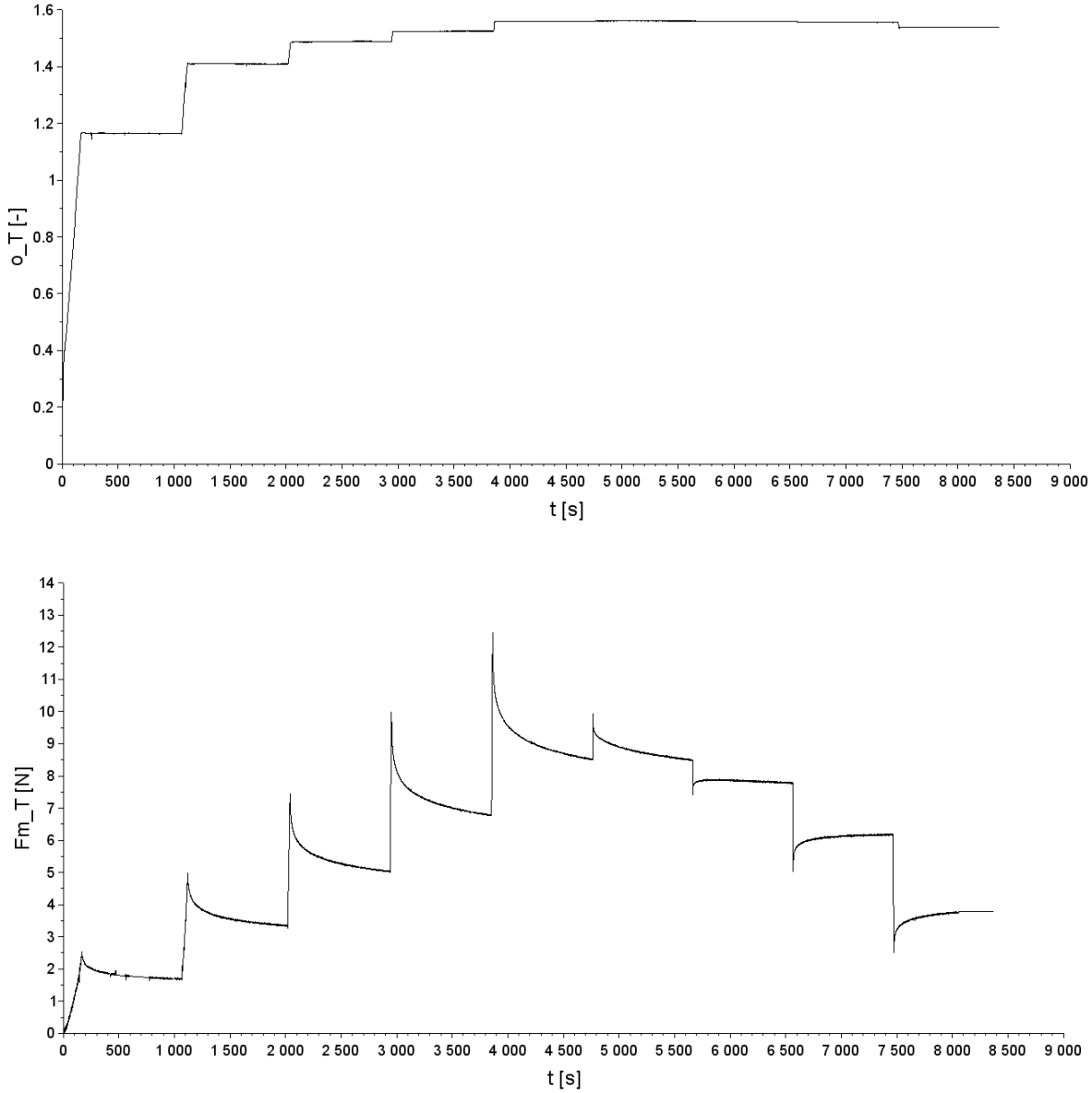


Figure 8: Result of a tensile test including relaxation stages of 15 minutes. Above: evolution of the opening in the T-direction of the mesh in the middle of the netting panel. Below: evolution of the force by mesh Fm_T applied on the mesh in the middle of the panel.

3.2 Suspending test

During suspending tests, different steps were performed: netting panels were submitted to forces added at the bottom, one after another, of 1.3, 3.257, 7.18 and 11.1 N. For each different level of solicitation, there was a creep period of 30 min. Figure 11 shows the mean value ($\pm$ standard deviation) of the heights L_{panel} measured on the 10 suspended netting samples.

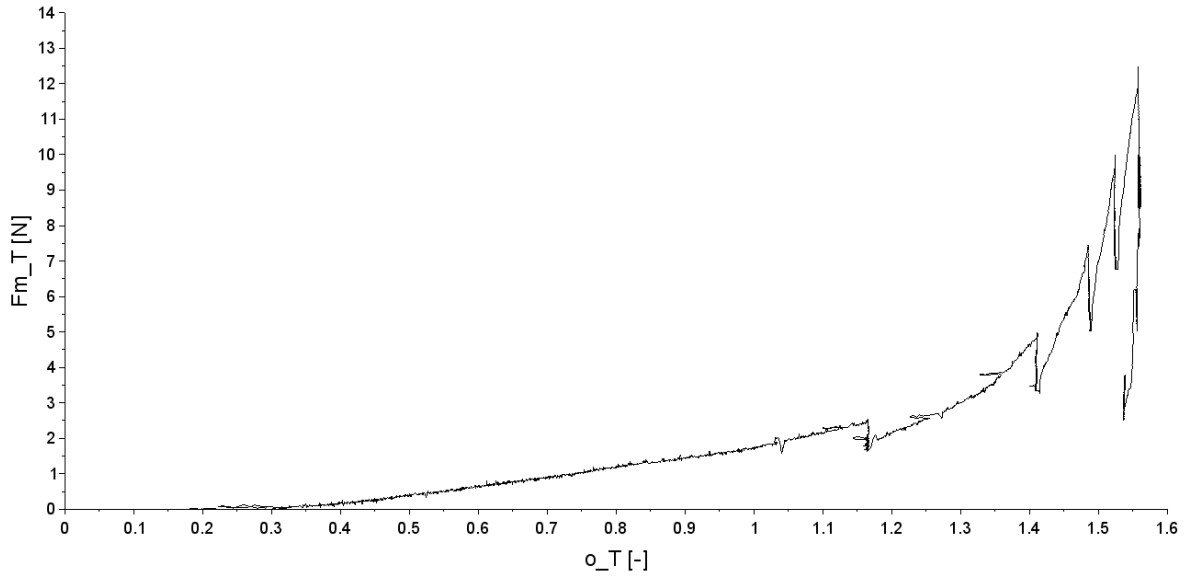


Figure 9: Result of a tensile test including relaxation stages.

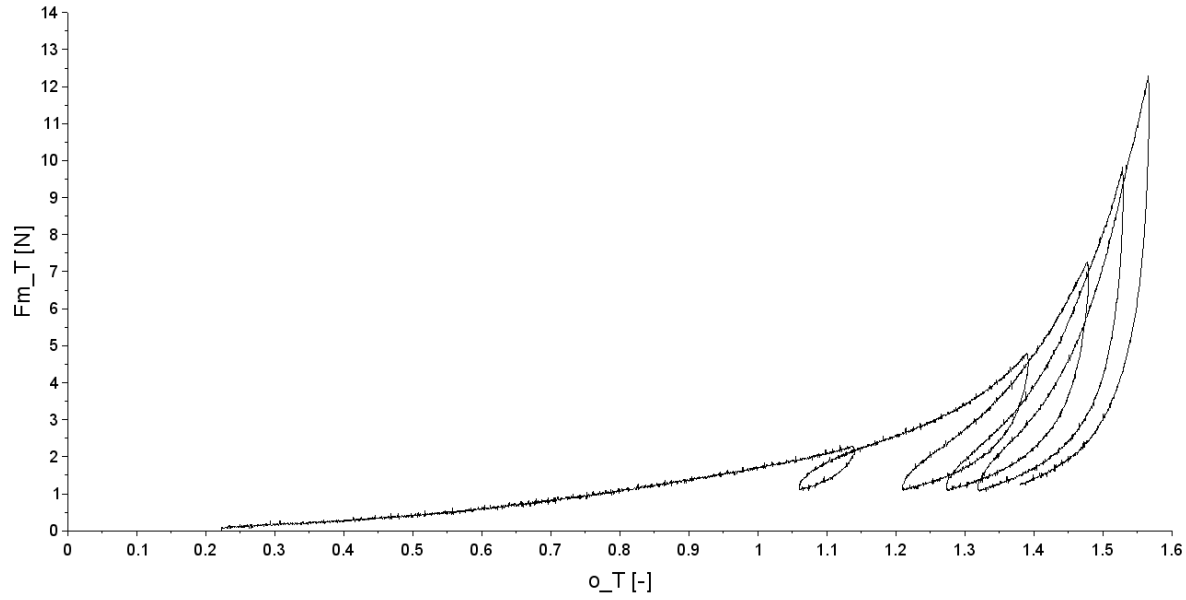


Figure 10: Result of a cyclic tensile test on a double twine netting sample.

4 NUMERICAL METHODS

4.1 Analytical model

An analytical model based on the beam theory and allowing large rotations is proposed to simulate the tensile test, assuming that the load applied on the mesh in the middle of the netting panel is supposed to be uniaxial, only in the T-direction. The notations used are defined in figure 12.

First, the bending moment M is related to the curvature as follows:

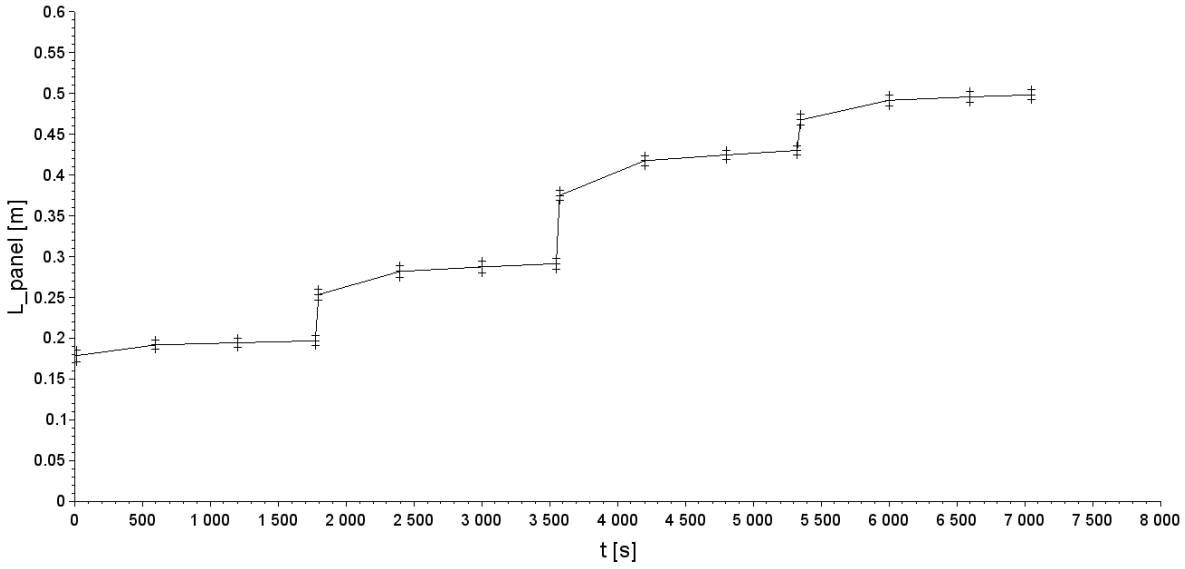


Figure 11: Evolution of the mean value ($\pm$ standard deviation) of the heights L_{panel} measured on the 10 suspended netting samples.

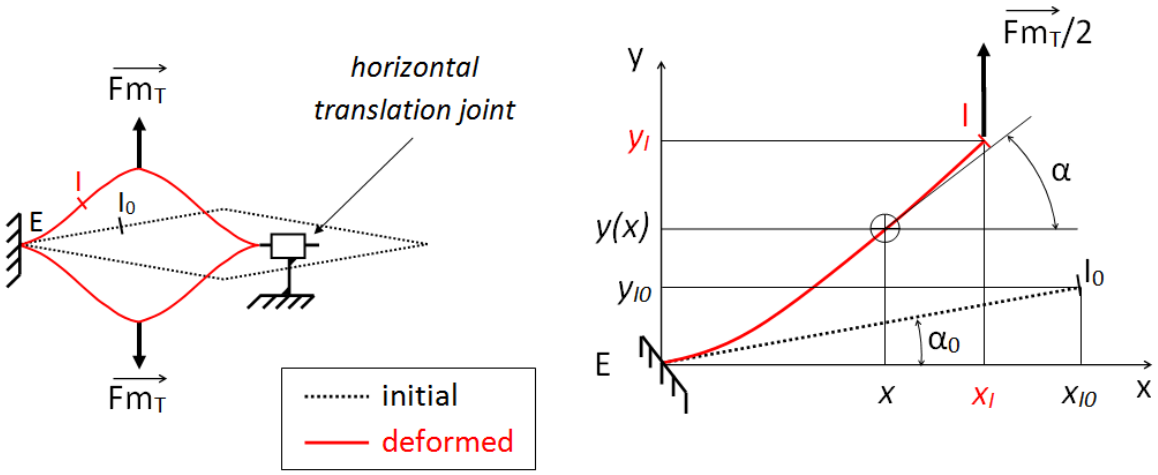


Figure 12: Left: application of a load Fm_T on a mesh. Right: kinematic of the twine between the points E and I.

$$M = EI\frac{d\alpha}{ds} \tag{3}$$

where EI is the bending stiffness, s is the curvilinear abscissa, α is the mesh angle, and $\frac{d\alpha}{ds}$ is the curvature.

The bending moment is given by:

$$M = \frac{Fm_T}{2}(x_I - x) \tag{4}$$

along the center line of the beam.
The curvature is a function of $y(x)$ as follows:

$$(\frac{d\alpha}{ds}) = \cos^2(\alpha) y''(x) \frac{1}{\sqrt{1+[y'(x)]^2}} = \frac{y''(x)}{[1+[y'(x)]^2]^{3/2}} \tag{5}$$

By using relations 3, 4 and 5, the following expression is obtained:

$$\forall x \in [0; x_I] \qquad \frac{Fm_T}{2}(x_I - x) = EI \frac{y''(x)}{[1+[y'(x)]^2]^{3/2}} \tag{6}$$

The boundary conditions are:

$$y(x=0) = 0 \tag{7}$$

$$\alpha(x=0) = \alpha_0 \Leftrightarrow y'(x=0) = \tan(\alpha_0) \tag{8}$$

The length of the mesh sides is assumed to be constant:

$$\forall t \in \mathbf{R}^+ \qquad L_{ms} = L_{ms0} \tag{9}$$

$$\int_E^I \mathrm{d}s = \int_{E_0}^{I_0} \mathrm{d}s = \frac{L_{ms0}}{2} \tag{10}$$

$$\frac{L_{ms0}}{2} = \int_0^{x_I} \frac{\delta s}{\delta x}\,\mathrm{d}x = \int_0^{x_I} \sqrt{1+[y'(x)]^2}\,\mathrm{d}x \tag{11}$$

To work with dimensionless parameters, $\bar{x}$ and $\bar{y}$ are defined by:

$$\bar{x} = \frac{2x}{L_{ms0}} \qquad \bar{y} = \frac{2y}{L_{ms0}} \tag{12}$$

And u is defined by:

$$\forall x \in [0; x_I] \qquad u(\bar{x}) = y'(x) \tag{13}$$

Using equations 12 and 13, relations 6, 8 and 11 become 14, 15 and 16 respectively.

$$\forall \bar{x} \in [0; \bar{x}_I] \qquad \frac{Fm_T}{2}(\frac{L_{ms0}}{2})^2(\bar{x}_I - \bar{x}) = EI \frac{u'(\bar{x})}{[1+[u(\bar{x})]^2]^{3/2}} \tag{14}$$

$$u(\bar{x}=0)=\tan(\alpha_0) \tag{15}$$

$$\int_0^{\bar{x}_I}\sqrt{1+[u(\bar{x})]^2}\,\mathrm{d}\bar{x}=1 \tag{16}$$

The solutions of the differential equation 14 are:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad \frac{Fm_T}{2EI}\frac{{L_{ms0}}^2}{4}\bar{x}[\bar{x}_I-\frac{\bar{x}}{2}]=\frac{u(\bar{x})}{\sqrt{1+[u(\bar{x})]^2}}+K \tag{17}$$

where $K\in\mathbf{R}$.
Using relation 15 with relation 17, we obtain:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad \frac{Fm_T}{2EI}\frac{{L_{ms0}}^2}{4}\bar{x}[\bar{x}_I-\frac{\bar{x}}{2}]=\frac{u(\bar{x})}{\sqrt{1+[u(\bar{x})]^2}}-\frac{\tan(\alpha_0)}{\sqrt{1+\tan^2(\alpha_0)}} \tag{18}$$

To simplify the equation, we defined v by:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad v(\bar{x})=\frac{u(\bar{x})}{\sqrt{1+[u(\bar{x})]^2}} \tag{19}$$

So u can also be defined by:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad u(\bar{x})=\frac{v(\bar{x})}{\sqrt{1-[v(\bar{x})]^2}} \tag{20}$$

where $v\in]-1;1[$.
Then relation 18 leads to:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad v(\bar{x})=\frac{Fm_T}{2EI}\frac{{L_{ms0}}^2}{4}\bar{x}[\bar{x}_I-\frac{\bar{x}}{2}]+sin(\alpha_0) \tag{21}$$

By using relations 20, 21 and 16, the value of $\bar{x}_I$ can be evaluated by an iterative algorithm (e.g. dichotomy algorithm).

Using the relations 12 and 13, $\bar{y}$ becomes:

$$\forall\bar{x}\in[0;\bar{x}_I] \qquad \bar{y}(\bar{x})=\int_0^{\bar{x}}u(\bar{x})\,\mathrm{d}\bar{x} \tag{22}$$

So $\bar{y}_I$ can be calculated by:

$$\bar{y}_I=\int_0^{\bar{x}_I}u(\bar{x})\,\mathrm{d}\bar{x} \tag{23}$$

Finally, the displacement of the half mesh side, and so the opening of the mesh in the middle of the netting panel during the tensile test, submitted to a tensile force Fm_T, can be calculated by relations 20, 21 and 23.

4.2 Finite element model

The mechanical behaviour of the studied netting was modeled using the Abaqus Standard finite element code. A mesh side was assumed to behave like a Timoshenko beam. In Abaqus, we model a mesh side with a planar beam that uses linear interpolation and a 2D hybrid formulation (B21H type in Abaqus). Timoshenko beams allow for transverse shear deformation and can be subject to large axial strains. Hybrid beam element types are used for geometrically nonlinear analysis when the beam undergoes large rotations. In this study, we modeled a mesh side with 20 B21H elements. The shear modulus was supposed to be very high so that the transverse shear deformation was not taken into account. The axial stiffness EA of the beam elements was determined using the results presented in figure 5.

4.3 Inverse identification methods

The global scheme for the identification of one mechanical parameter, here the bending stiffness, is presented on Fig. 13.

In case of the tensile test, the objective function was the difference between the experimental and the numerical height of the mesh in the middle of the netting panel. In case of the suspending tests, the objective function was calculated by using two different methods: first, the objective function was the mean distance between the coordinates of the experimental and the numerical knots; then, the objective function was the difference between the experimental and the simulated suspended sample height. There was a difference of 3.5 per cent between the bending stiffness identified using the measured coordinates of all the knots and the bending stiffness evaluated using the height of the netting panel. Thus, it was decided to identify the parameter with only the height of the netting sample.

The optimization of the search of the parameter EI is performed with the Nelder-Mead (or downhill simplex) algorithm proposed by John Nelder and Roger Mead ([8]).

This algorithm is used within the Python library SciPy, which contains an optimization module. The convergence is usually reached after about ten loop iterations of the Nelder-Mead algorithm.

5 NUMERICAL RESULTS

5.1 Tensile test

Using the experimental results and the identification method presented previously, we identified the bending stiffness EI of braided twines constituting the netting sample. The evolution of the bending stiffness EI is presented on figure 14. The evolutions of this parameter during the whole test, including relaxation stages, identified with the finite element model and the analytical model, are presented.

Figures 15 and 16 show, respectively, the influence of the force by mesh in the T-direction and the influence of the opening o_T on the bending stiffness. For these two figures, the bending stiffness was identified using the finite element model at the begining (0 min) and at the end (15 min) of the relaxation stages.

5.2 Suspending tests

The identified evolution of the bending stiffness, in the case of suspending tests, is presented on figure 17. As previously described, the identifications of the bending stiffness have been carried out using results of suspending tests on 10 netting samples.

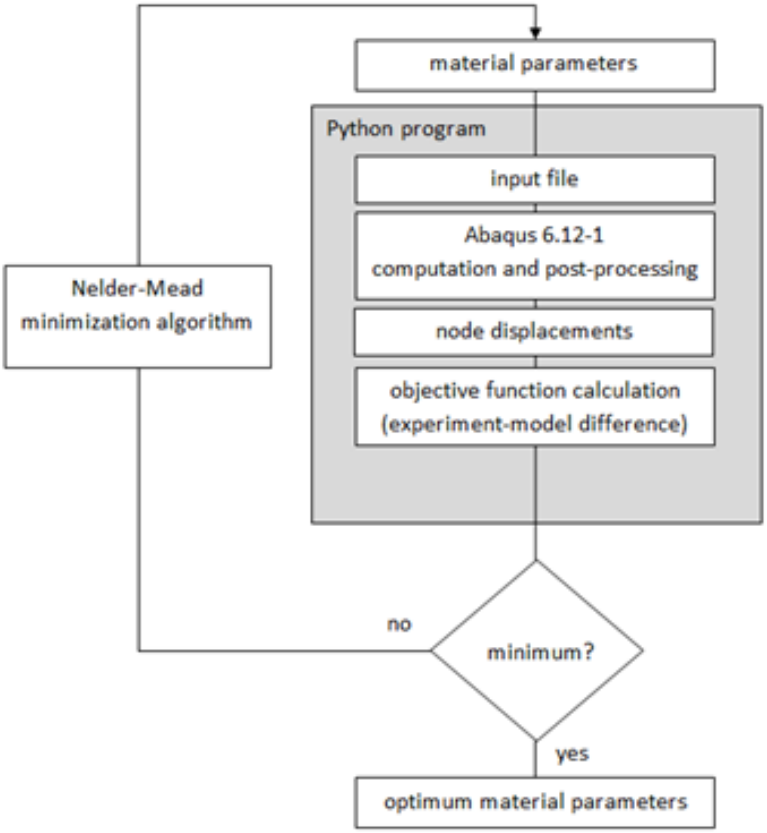


Figure 13: Inverse analysis strategy.

6 DISCUSSION

The tensile test provides results which reveal the viscosity of the material (Fig. 8 and 9). Figure 14 shows the importance of taking into account the viscosity of the material, which influences greatly the identified bending stiffness. Thus, with the assumption of a viscoelastic behaviour, the relaxation stages permit to reduce the effect of viscosity and to improve the evaluation of the bending stiffness; indeed, the higher the relaxation duration is, the less the rate dEI/dt is, then the less the effect of the time on the value of EI is. Figure 16 shows that the opening o_T was still high during the unload (with relaxation stages of 15 minutes). So the mesh did not come back to its initial state. The structure presented a permanent deformation or a long-term viscosity.

Figure 14 shows that the bending stiffness identified with the analytical model is lower than the one evaluated with the finite element model. The relative difference between the two curves ranges from 4.40 % for a force by mesh value of $1.59N$ to 12.77 % for a force by mesh value of $12.38N$. According to the finite element model of tensile test, the mesh in the middle is submitted to a tensile force in the T-direction and a compression force in the N-direction. However, the analytical model only takes into account the tensile force. We can assume that to obtain the same mesh opening in T-directon without compression force in the N-direction, a lower bending stiffness is necessary. We can note that for a low force level, in case of a tensile test, the bending stiffness identified with the analytical model is close to the one evaluated with the

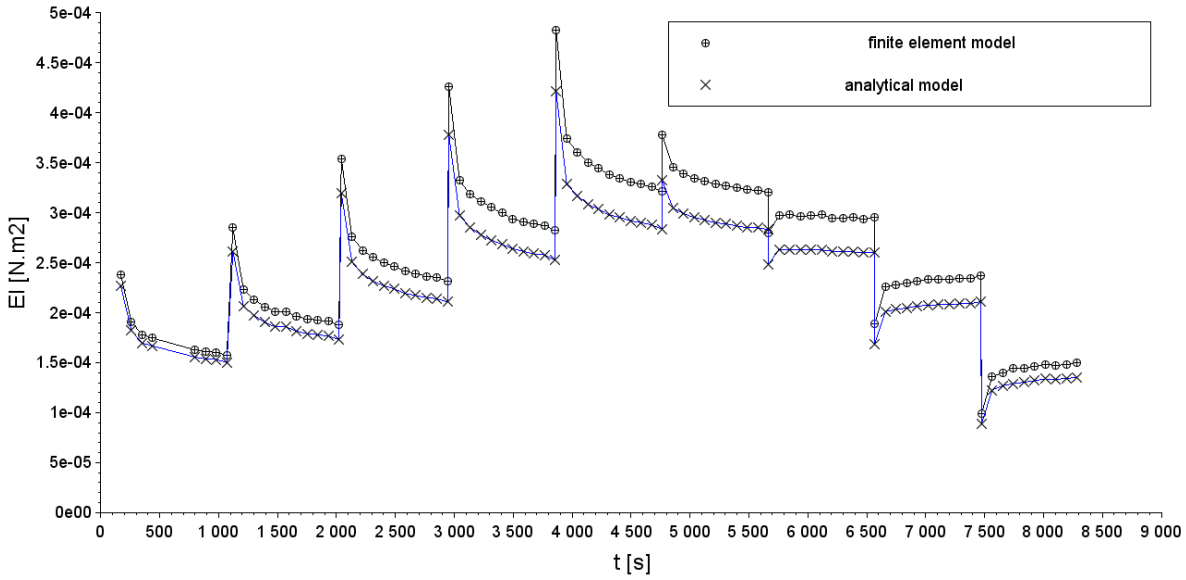


Figure 14: Tensile test. Evolution of the bending stiffness EI, identified with the finite element model ('$\oplus$', black line) and the analytical model ('$\times$', blue line).

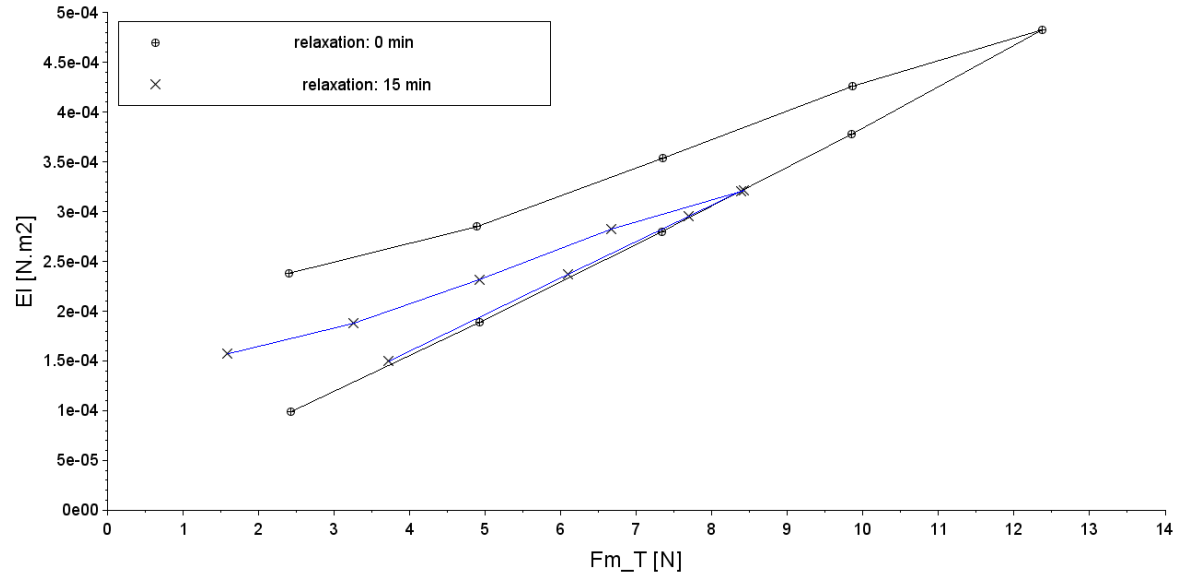


Figure 15: Tensile test. Evolution of the bending stiffness EI, identified using the finite element model at 0 minute ('$\oplus$', black line) and at 15 minutes ('$\times$', blue line) of each relaxation stage, as a function of the applied load Fm_T.

finite element model.

Figures 15 and 16 show the evolution of the identified bending stiffness as a function of the force applied on the mesh Fm_T and as a function of the opening o_T respectively. We can note that the variation of the parameter EI during the test is smaller when it has been calculated at the end of each relaxation step. However, the evolution of Fm_T (Fig. 8) indicates that the contribution of the viscosity is not completely relaxed after a relaxation stage of 15 minutes.

When the opening increases (and so the force), the tensile force in mesh sides increases so

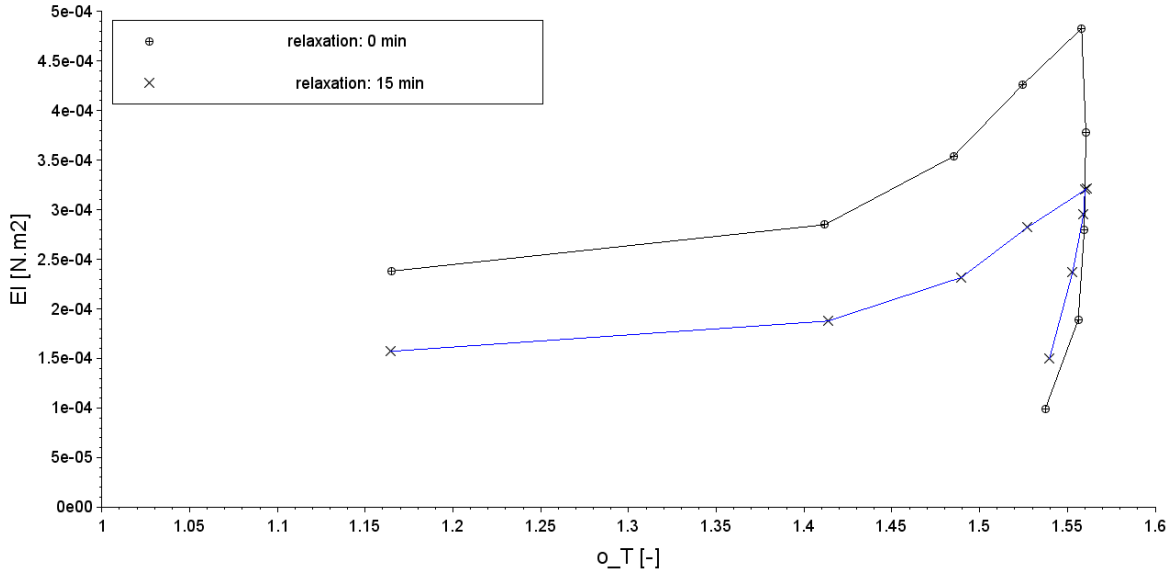


Figure 16: Tensile test. Evolution of the bending stiffness EI, identified using the finite element model at 0 minute ('$\oplus$', black line) and at 15 minutes ('$\times$', blue line) of each relaxation stage, as a function of the opening in the T-direction, o_T.

we could suspect that the twine diameter decreases. And a smaller diameter means a smaller moment of inertia I. Thus, we could suspect a decrease of EI when the opening increases. But figure 16 shows that EI increases when the opening o_T of the mesh increases. It is necessary to take into account the complexity of the twine structure. When the twine diameter decreases, interactions between fibres (directions) and interactions between the core and the sheath become different. The complexity of the behaviour was already revealed on figure 5 which showed that it was not possible to identify a constant Young modulus E.

During the tensile test, the force by mesh at the beginning of relaxation stages ranged from $2.40\ N$ to $12.38\ N$. During suspending tests, netting panels were submitted to forces by mesh added at the bottom ranging from $0.3235\ N$ to $2.775\ N$. The force by mesh applied on the uppermost meshes was higher and ranged from $0.46\ N$ to $2.91\ N$ due to the effect of gravity. The load applied on meshes during suspending tests were only around a quarter of the load applied during the tensile test. With this loading level, a smaller opening o_T of the mesh is observed during the suspending test than during the tensile test, therefore a smaller variation of the identified bending as presented on figure 17. We can note that the bending stiffness identified for the smallest value of force by mesh with the tensile test is close to the ones identified with the suspending tests.

The suspending tests on 10 netting samples allow us to calculate the first and third quartiles, and the standard deviation of the bending stiffness (Fig. 17). The standard deviation of the bending stiffness decreases when the load increases, so when the opening of meshes increases. We can remember that each sample was subject to a pretension step (loaded in tensile with $392.4\ N$ in the N-direction during 30 minutes). Thus, we can suppose that the decrease of the standard deviation is not caused by the tightening of knots.

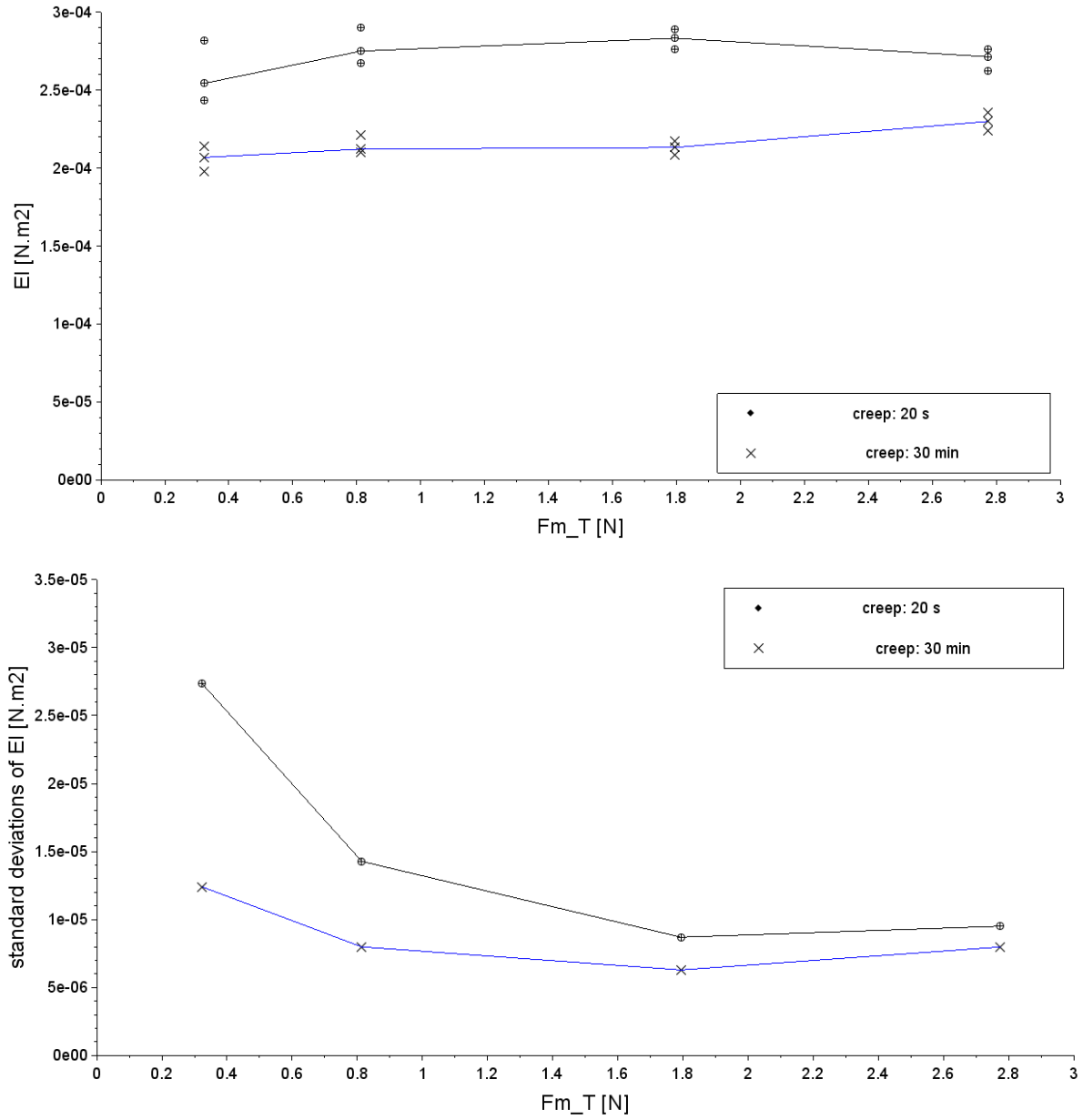


Figure 17: Suspending test. Medians of the bending stiffness EI with first and third quartiles (above) and standard deviation (below) of EI at the begining (20 s) and at the end (30 min) of each creep stage.

7 CONCLUSION

A simple method to evaluate the bending stiffness of braided twines, so the mesh opening stiffness of fishing nets, is presented. In this study, twines were made up of polyethylene.

Two different tests were performed: a tensile test and a suspending test. In case of a tensile test, a load and unload cycle interrupted by several relaxation stages was imposed. In case of suspending test, netting panels were submitted to forces added at the bottom, with creep stages. The experimental results show the importance of taking into account the viscosity of the material.

Two models based on the beam assumption were developped: an analytical model and a finite

element model. The results obtained with these models are close, even if the bending stiffness identified with the analytical model, in case of the tensile test, is lower than the bending stiffness evaluated with the finite element model. The finite element model is more accurate because it models the compression force in the N-direction.

The results of the identifications show an evolution of the bending stiffness as a function of time (viscosity), of the load level and of the mesh opening. To take into account the effect of the structure of the braided twine, one could model the fibres constituting the core and the sheath of each twine, and the interactions between the fibres. This work provides experimental results that could be used to study the contact-friction interactions, to develop a model for large deformation analysis of fibrous material.

Finally, the identification of the bending stiffness EI on suspending test, proposed here, could be used as a simple methodology to evaluate the mesh opening stiffness of fishing nets.

ACKNOWLEDGMENT

This research received the support of Region Bretagne under a financing agreement for doctoral research (no. 13/2.213 747/F).

REFERENCES

[1] P. Badel, E. Vidal-Sall, P. Boisse, Large deformation analysis of fibrous materials using rate constitutive equations. *Computers and Structures*, **86**, 1164–1175, 2008.

[2] G. Bles, W.K. Nowacki, A. Tourabi, Experimental study of the cyclic visco-elasto-plastic behaviour of a polyamide fibre strap. *International Journal of Solids and Structures*, **46**, 2693–2705, 2009.

[3] M. K. Broadhurst, D. J. Sterling, R. B. Millar, Increasing lateral mesh openings in penaeid trawls to improve selection and reduce drag. *Fisheries Research*, **170**, 68–75, 2015.

[4] D. Durville, Contact-friction modelling within elastic beam assemblies: an application to knot tightening. *Comput Mach*, **49**, 687–707, 2012.

[5] B. Herrmann, F.G. O'Neill, Theoretical study of the influence of twine thickness on haddock selectivity in diamond mesh cod-ends. *Fisheries research*, **80**, 221–229, 2006.

[6] ISO, Fishing nets — Netting — Basic terms and definitions. *International Standard Organisation*, **1107**, ISO Geneva, 1974.

[7] W. Moderhak, Influence of twine parameters on the shapes of meshes and T90 codends. *International Workshop Methods for the Development and Evaluation of Maritime Technologies(DEMaT'07)*, ROSTOCK, Germany, 2007.

[8] J.A. Nelder, R. Mead, A Simplex Method for Function Minimization. *The Computer Journal*, **7**, 308–313, 1965.

[9] F.G. O'Neill, Bending of twines and fibres under tension. *Journal of the Textile Institute*, **93**, 1–8, 2002.

[10] F.G. O'Neill, A theoretical study of the factors which influence the measurement of fishing netting mesh size. *Ocean engineering*, **30**, 2053–2063, 2003.

[11] A. de la Prada, M. Gonzales, Nonlinear stiffness models of a net twine to describe mesh resistance to opening of flexible net structures. *J Engineering for the Maritime Environment*, 2014.

[12] A. de la Prada, M. Gonzales, Quantifying mesh resistance to opening of netting panels: experimental method, regression models, and parameter estimation strategies. *ICES Journal of Marine Science*, 2014.

[13] D. Priour, A finite element method for netting: application to fish cages and fishing gears. *Springer*, 2013.

[14] D. Priour, J.-Y. Cognard, Investigation of methods for the assessment of the flexural stiffness of netting panels. *10th International Workshop Methods for the Development and Evaluation of Maritime Technologies (DEMaT'11)*, SPLIT, Croatie, 2011.

[15] J.H.B. Robertson, P.A.M. Stewart, A comparison of size selection of haddock and whiting by square and diamond mesh codends. *J. Cons. Int. Explor. Mer*, **44**, 148–161, 1988.

[16] A. Sala, F.G. O'Neill, G. Buglioni, A. Lucchetti, V. Palumbo, RJ. Fryer, Experimental method for quantifying resistance to the opening of netting panels. *ICES Journal of Marine Science: Journal du Conseil*, **64**, 1573–1578, 2007.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MULTI-SCALE FRAMEWORK FOR MODELLING CARBON FIBRE WEAVES BASED ON STOCHASTIC REINFORCEMENT GEOMETRY

Andy Vanaerschot[1], Karen Soete[1], Stepan V. Lomov[2] and Dirk Vandepitte[1]

[1]KU Leuven, Department of Mechanical Engineering
Celestijnenlaan 300, 3001 Leuven, Belgium
e-mail: {andy.vanaerschot,karen.soete,dirk.vandepitte}@kuleuven.be

[2] KU Leuven, Department of Materials Engineering
Kasteelpark Arenberg 44, 3001 Leuven, Belgium
e-mail: stepan.lomov@kuleuven.be

Keywords: Textile composites, Statistics, Probabilistic methods, Multi-scale modelling.

Abstract. *The internal geometry of textile composites is subjected to a significant amount of variability. An improved assessment of the quality of any composite material is achieved by identification and simulation of the inherent uncertainty in the reinforcement geometry. By following this strategy, many random realistic representations of the textile reinforcement can be generated based on experimental data of only a few textile samples. This work presents such a roadmap with its successive steps to characterise the spatial scatter in the internal structure of any textile composite (short- and long-range) and simulate random models possessing the measured statistical information on average. Three main steps can be distinguished: (1) collection of experimental data and statistical analysis, (2) stochastic multi-scale modelling of the reinforcement, and (3) construction of virtual specimens in a geometrical pre-processor such as WiseTex. The methodology is demonstrated for a typical carbon-epoxy 2/2 twill woven composite produced by resin transfer moulding, but is applicable to any woven textile or other topology with minor modifications.*

1 INTRODUCTION

The internal geometry of fibre-based composites is subjected to a significant amount of scatter. Variability in the macroscopic performance is dominated by the spatial randomness in the geometrical characteristics at the lower scale, especially for textile composites. By identifying the irregularity in the tow reinforcement, an improved assessment of the quality of a composite can be obtained and a virtual model with a more realistic representation of the reinforcement can be constructed. Among the different strategies for simulating variability in composites using appropriate scaling techniques [1, 2], those that will lead to the most accurate predictions of the statistical distributions of composite properties are calibrated by experimental quantification of the reinforcement variability. However, an inadequacy of experimental data on textile geometry is exists, while methods are lacking for reliable modelling of the effects of geometrical scatter on the randomness in the mechanical properties.

This article presents a stochastic modelling procedure for textile composites that enables to generate as many random virtual specimens as desired based on the statistical analysis of the reinforcement of only one or a few samples. The ensemble of virtual specimens possess the same statistical information on average as measured from the physical samples. The methodology consists of three steps: (1) collection of experimental data about reinforcement geometry, (2) multi-scale modelling of the reinforcement using advanced simulation strategies, and (3) creation of virtual specimens in a geometrical pre-processor such as WiseTex [3]. While step 1 is essential in replicating the real geometry, it is often omitted and researchers are forced to make assumptions about the input information for step 2 and 3.

Next sections describe the general approach and successive steps to build random virtual specimens, starting from the collection of experimental data till the generation of random samples. The methodology is applied on a carbon-epoxy 2/2 twill woven textile composite.

2 MULTI-SCALE FRAMEWORK

Scatter in the reinforcement is considered at the meso-scale (or the unit-cell scale) and macro-level (or the sub-component scale); variation in the matrix and fibre properties is not considered. The variability of each tow path is expressed in terms of its centroid coordinates (x, y, z), tow cross-sectional aspect ratio AR, tow cross-sectional area A and tow cross-sectional orientation θ. Three main steps can be distinguished in obtaining realistic descriptions of the internal geometry of textile composites:

1. **Collection of experimental data with statistical analysis**
   (a) Quantification of the short-range variation
   (b) Quantification of the long-range variation
   (c) Statistical analysis in terms of average trends, standard deviation and correlation information
2. **Stochastic multi-scale modelling of the reinforcement**
   (a) Combination of systematic and handling trends from the experimental data
   (b) Monte Carlo Markov Chain method for simulating auto-correlated deviations
   (c) Cross-correlated Series Expansion for simulating auto- & cross-correlated deviations

3. **Construction of virtual specimens in WiseTex**

   (a) Construction of the nominal model with manufacturer's data

   (b) Update tow path descriptions with generated instances

3 COLLECTION OF EXPERIMENTAL DATA WITH STATISTICAL ANALYSIS (STEP 1)

3.1 Methodology

Non-destructive inspection techniques are applied to measure the fabric architecture in a reliable and efficient way across the composite volume. Scatter in geometry is inspected on the short-range (meso-scale) and the long-range (macro-scale).

It is recommended to acquire short-range data on uncertain tow path parameters using X-ray micro-computed tomography (micro-CT). A three-dimensional (3-D) reconstructed volume is obtained, where from two-dimensional (2-D) slices are extracted in warp and weft direction. Tow cross-sectional shapes can be fitted with ellipses to extract data that fully describe the reinforcement: centroid coordinates (x, y, z), tow cross-sectional aspect ratio AR, tow cross-sectional area A and tow orientation θ in cross-section. After the collection of geometrical data, the reference period collation method [5] is applied. This approach groups tows that should be identical given the nominal periodicity of the textile, with such a representative tow named *genus*. Each tow parameter of each genus is decomposed in a non-stochastic, average trend and non-periodic stochastic fluctuations. These stochastic deviations are further quantified in terms of standard deviation and correlation length.

In addition to the short-range data, the collection of long-range can be advised when one or more tow path parameters have a correlation length ξ that exceeds the unit cell dimensions. Depending on the tow path parameter(s) that need such long-range quantification, additional micro-CT scans, optical imaging techniques or digital image correlation need to be performed.

The statistical information of all tow path parameters, from short- and long-range characterisation, are used as input to the next step for calibrating the reconstruction algorithms.

3.2 Application to a 2/2 twill composite

The methodology is applied to a 2/2 twill woven composite. The dry reinforcement is a fabric from Hexcel (G0986) [4] which is impregnated with epoxy in a Resin Transfer Moulding (RTM) process. It is a balanced textile with four warp (x-axis) and four weft (y-axis) tows. A virtual representation of the unit cell is given in figure 1 with λ_x=11.43 mm and λ_y=11.43 mm, respectively the periodic lengths of warp and weft tows. Considering the manufacturing of the 2/2 twill woven fabric, warp tows can be represented by one *genus*, and similar for the weft tows.

All tow parameters, with exception of the in-plane centroid position, vary within the unit cell dimensions. The latter centroid position is also subjected to the largest variability, indicated by the high standard deviation σ. The micro-CT procedure and derived statistical information is described in [6]. Additional long-range information is acquired for the in-plane centroid positions. An optical scan of the in-plane dimension of a single-ply composite, spanning a region of interest of 10 by 10 unit cells, is performed. A detailed discussion of the procedure and results are given in [8].

Figure 2 present the average reinforcement of the 2/2 twill woven composite. Deviations for all tow path parameters are approximately represented by a normal distribution, with a summary

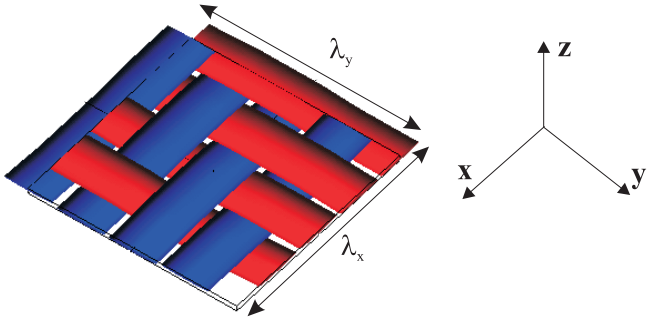


Figure 1: WiseTex model of a 2/2 twill woven reinforcement. The x-axis and y-axis of the coordinate system are respectively parallel to the warp and weft direction. The z-direction refers to the tickness direction.

of the statistical data given in table 3.2. Only the in-plane centroid position is cross-correlated between neighbouring tows.

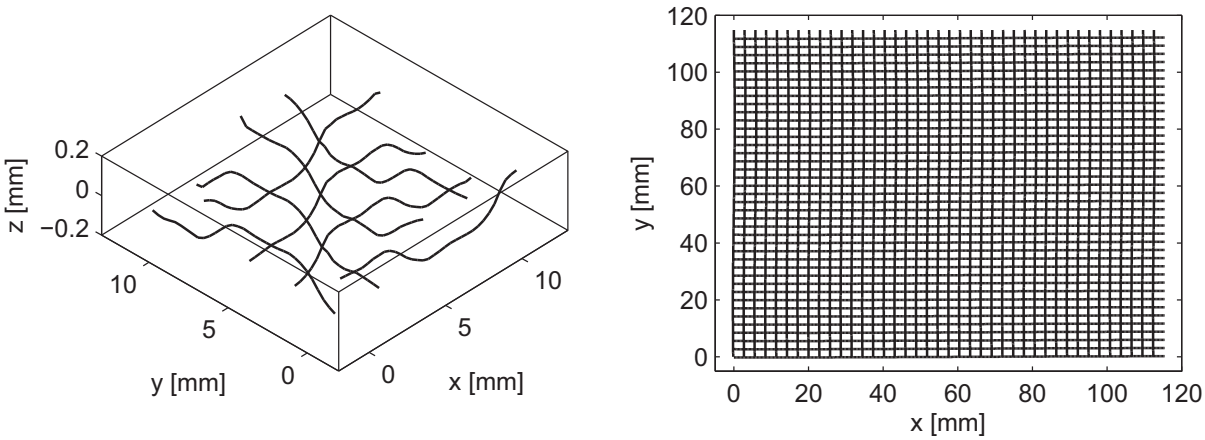


Figure 2: The average reinforcement description presented for the centroid coordinates of a 2/2 twill woven composite.

	x [mm]	**y [mm]**	**z [mm]**	AR **[-]**	A **[mm^2]**
σ^{warp} [mm]	-	0.106	0.014	1.774	0.023
ξ_{auto}^{warp} [mm]	-	114.89	1.78	7.26	2.53
ξ_{cross}^{warp} [mm]	-	4.49	-	-	-
σ^{weft} [mm]	0.615	-	0.015	1.440	0.024
ξ_{auto}^{weft} [mm]	52.89	-	1.62	5.48	1.01
ξ_{cross}^{weft} [mm]	13.16	-	-	-	-

Table 1: Statistical information of all tow path parameters: standard deviation, auto- and cross-correlation length.

4 STOCHASTIC MULTI-SCALE MODELLING OF THE REINFORCEMENT (STEP 2)

4.1 Methodology

A stochastic multi-scale modelling approach is developed to reproduce the measured variation in the tow reinforcement within the unit cell and between neighbouring unit cells. Random

instances of tow paths are acquired by following successive steps:

1. Construct a 2-D lattice of straight tows, with rows representing the warp tows and columns representing the weft tows (other definition could be required for non-woven textiles).

2. On each row or column, define an equidistant grid consisting of locations i' and with a length $N_{i'}$ of which (1) the grid spacing is smaller than the shortest correlation length and (2) the length at least approximates the largest correlation length.

3. Interpolate the average trends of all tow path parameters to match the grid locations. Periodicity is exploited for the repetitive systematic trend of the short-range parameters.

4. Generate and add a set of zero-mean deviations $\{\tilde{\epsilon}^{sr}_{i'},\ i' = 1..N_{i'}\}$ along the grid of a single tow for each tow path property that exhibits a short-range trend. This step is performed continuously along an entire tow path spanning multiple unit cells and does not need to be limited per unit cell length.

5. Generate and add a set of zero-mean deviations $\{\tilde{\epsilon}^{lr}_{j'},\ j' = 1..N_{j'}\}$ along a second grid with locations j' of the same length but with a considerably larger grid spacing, with at least one point per periodic length. Afterwards interpolate the produced values to the short-range grid with locations i'.

A necessary condition for superposition of the short-range deviations onto the long-range deviations is that the short-and long-range deviations are not correlated with each other.

Zero-mean deviations which are only correlated along the tow path are produced by the Monte Carlo Markov Chain for textile structures [10], while uncertain quantities that are dependent along and between tow paths are generated using a cross-correlated Series Expansion method [11].

4.2 Monte Carlo Markov Chain algorithm

Tow path parameters which vary without any type of cross-correlation are generated using the Markov Chain algorithm for textile structures [10]. Deviations of a single tow path parameter are produced independently from other tow parameters in a single process. The Markovian operation is the core computation within the Monte Carlo based scheme and generates the distribution vector P^{ϵ}_{i+1} of the particular parameter ϵ at the next grid location $i+1$ using a probability transition matrix A^{ϵ}_{trans}:

$$P^{\epsilon}_{i+1} = A^{\epsilon}_{trans} P^{\epsilon}_{i} \tag{1}$$

Each tow parameter has a different transition matrix A^{ϵ}_{trans} that is calibrated with the experimental standard deviation and nearest-neighbour correlation information to reproduce the statistical information. A post-processing smoothing operation is required to reduce low-amplitude short-range spikes present in the discretized tow path. More details on this algorithm can be found in [10, 7].

4.3 Cross-correlated Series Expansion

A methodology proposed by Vořechovský [11] based on Series Expansion simulates tow path parameters that share the same auto-correlation and of which the cross-correlation can be represented by a single scalar, called cross-correlation coefficient. Each uncertain parameter is represented by a Gaussian random field H_i. The method is calibrated with the experimental

standard deviation, auto-correlation and cross-correlation data, of which the correlation information is summarised in correlation matrices. A single realisation of tow parameter deviations for a certain tow path H_j, represented by $\tilde{H}_j$, is acquired using the Karhunen-Loève (K-L) Series Expansion [12], with λ and ϕ respectively the eigenvalues and -vectors of the auto-correlation matrix:

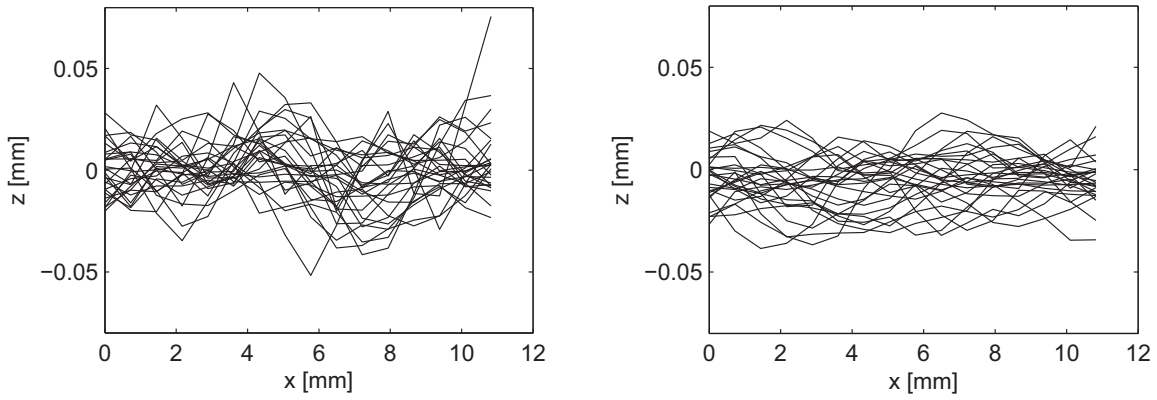
$$\tilde{H}_j(x) = \sum_{i=1}^{N_{KL}} \sqrt{\lambda_i^A} \chi_{j,i}^D \phi_i^A(x) \tag{2}$$

The random variables of χ^D are uncorrelated for each set of random variables that describe a single tow and at the same time cross-correlated between all sets of random variables belonging to different tows. Random fields of the in-plane centroids of warp and weft tows are simulated as a truncated series with N_{KL} terms. Only the largest eigenvalues and corresponding eigenvectors are considered in the procedure. A detailed discussion of the method is given in [9].

4.4 Application to a 2/2 twill composite

Stochastic instances of the 2/2 twill reinforcement are generated for a region spanning ten by ten unit cells, which is sufficiently large to represent the short- and long-range variation quantified in this composite. Each tow is discretised in 320 points, corresponding to 32 points per unit cell. The model is representative for one ply within a laminate which is stacked with 2/2 twill carbon fabrics on each ply.

The Monte Carlo Markov Chain method is employed for the generation of zero-mean deviations of the out-of-plane centroid z, area A and aspect ratio AR deviations. Smoothing is applied using information of ± 2 neighbouring grid points to obtain a realistic tow path, as demonstrated for the warp out-of-plane centroid for one periodic length in figure 3.



(a) Experiment (b) Simulation

Figure 3: Warp out-of-plane centroid deviations trend for 28 warp tows: (a) experimental vs. (b) smoothed deviations obtained from simulations.

Tow path deviations for thousand virtual specimens are generated with the Markov Chain to verify if the experimental statistics are reproduced. The standard deviations and auto-correlation lengths of single unit cells are centred around the experimental target values. Figure 4 shows the histogram of the unit cell statistics for the z-centroid of the warp tows. Smoothing has a limited effect on the standard deviation, while the correlation lengths are slightly shifted to higher values. Similar conclusions are made for the other tow parameters of the warp and weft genus.

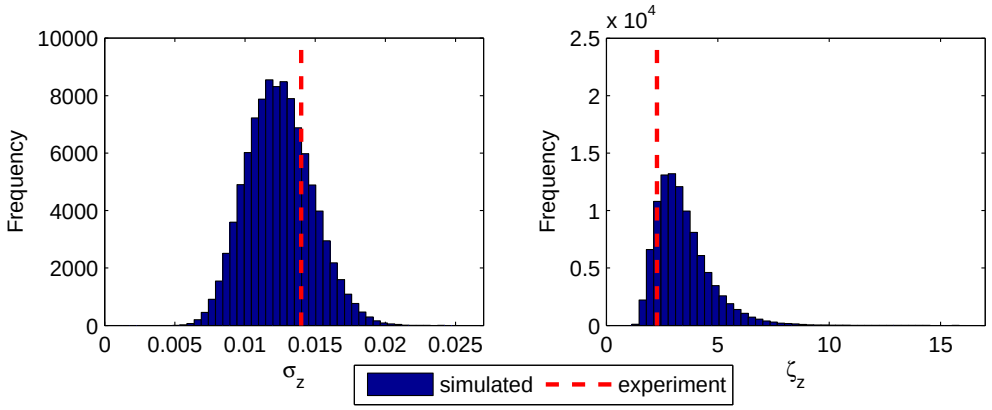
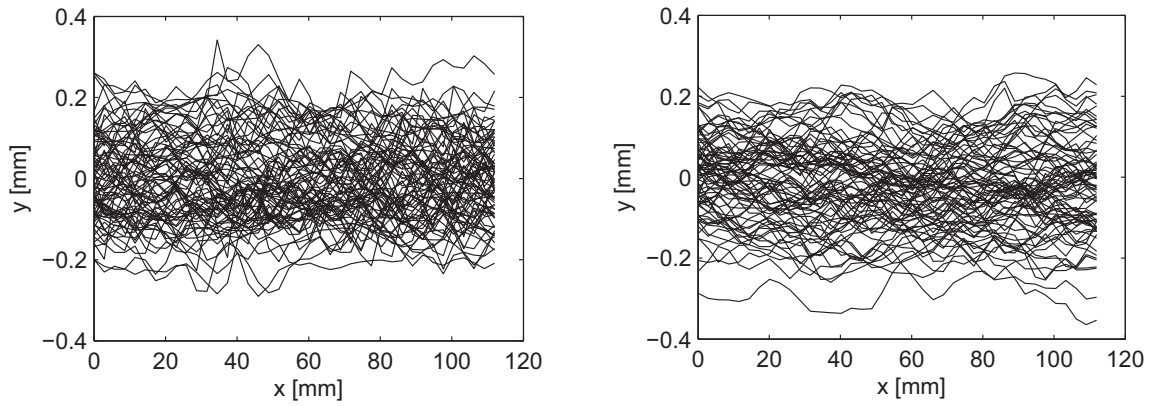


Figure 4: Unit cell standard deviations and correlation lengths of the out-of-plane warp centroid after smoothing.

All cross-correlated in-plane centroid deviations of the 2/2 twill woven composite are generated with the cross-correlated Series Expansion technique over a grid of 41 points. A good resemblance is found for the deviations pattern of the measured in-plane deviations and the produced deviation patterns, as demonstrated in figure 5 for the warp tows. The short wavelength of the experimental warp fluctuations and long wavelength of the measured weft deviations are reproduced without the need of an additional smoothing operation. As post-processing step, the generated deviations are interpolated over the grid of 320 points which is used for the short-range variations.

Comparison of the produced and target statistical information using the Series Expansion is demonstrated at the level of individual specimens. In figure 6, the histogram is presented of the generated auto- and cross-correlation lengths for the warp tows. The mean of the generated correlation lengths achieve the target statistics on average.



(a) Experiment (b) Simulation

Figure 5: Warp in-plane centroid deviations trend for 80 warp tows: (a) experimental vs. (b) simulated deviations.

5 CONSTRUCTION OF VIRTUAL SPECIMENS IN WISETEX (STEP 3)

5.1 Methodology

The last step of the multi-scale framework creates virtual textiles with random geometry. A virtual representation of the textile composite is obtained in WiseTex [3], which is a commer-

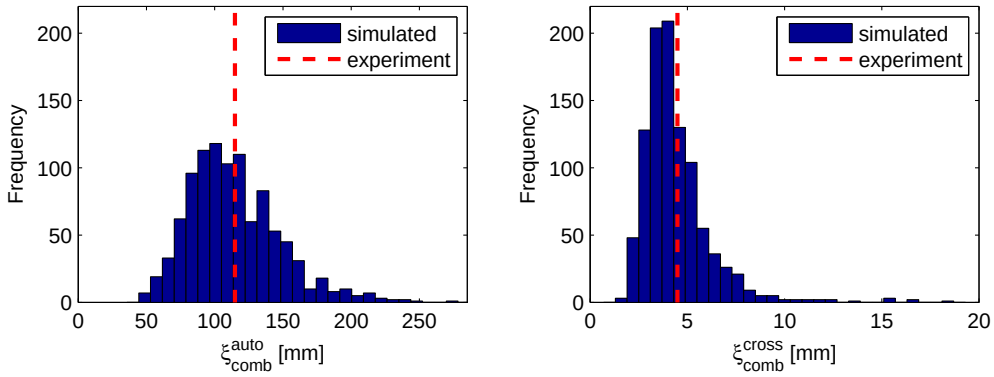


Figure 6: Auto- (left) and cross-correlation (right) lengths of produced warp in-plane positions.

cially available pre-processor for textile geometry. To obtain a random model, nominal tow path descriptions are overwritten with realistic tow representations obtained from the previous step, while preserving the original fibre mechanics and matrix properties.

5.2 Application to a 2/2 twill composite

A virtual specimen of the 2/2 twill woven composite is presented in figure 7, with warp and weft tows oriented respectively along the horizontal and vertical direction. The in-plane centroid position varies on the long-range with bundling behaviour appearing for the weft tows. The unit cell image demonstrates the variation in the out-of-plane centroid position and tow cross-sectional parameters.

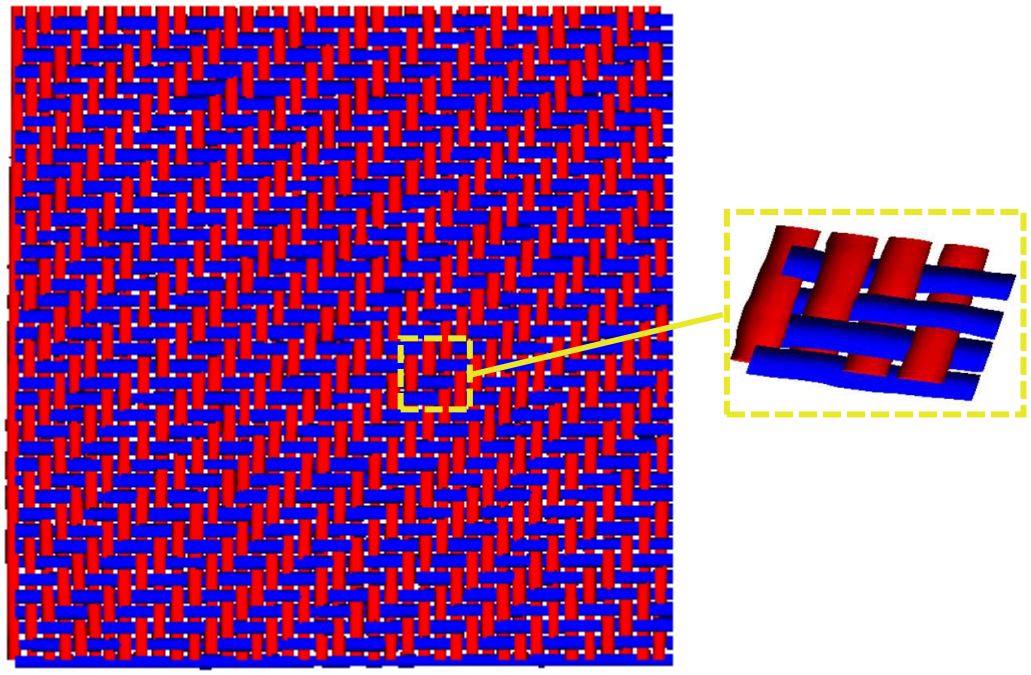


Figure 7: Virtual specimen in the WiseTex format. The in-plane dimension and a single unit cell is presented to demonstrate the short- and long-range variability.

6 CONCLUSIONS

A stochastic multi-scale strategy is presented to quantify and simulate the geometrical variability present in a textile composite. Three general steps are followed to acquire virtual specimens that possess the same statistical information on average as measured from experimental samples. In step 1, experimental data are collected on the centroid positions and cross-sectional dimensions over the short- and long-range distance. Step 2 continues with the construction of virtual tow paths as a combination of a mean trend with zero-mean deviations. If a tow path parameter is not cross-correlated with its adjacent tows, deviations for that parameter are produced with the Monte Carlo Markov Chain method. If cross-correlation is present, a cross-correlated Series Expansion technique is applied to generated the tow path properties. In the final step, random virtual specimens are created in the WiseTex software. These geometrical models will support material design in predicting a more realistic range in which the mechanical properties vary.

The methodology is demonstrated on a carbon-epoxy 2/2 twill woven composite. The in-plane centroid positions show the largest standard deviation of all tow properties with a correlation length exceeding the unit cell dimensions. Therefore, additional data are gathered by optical imaging of the in-plane dimension of multiple unit cell one-ply composites to inspect the in-plane positions over a larger distance. Virtual reinforcements are simulated that span a region of ten by ten unit cells and are representative for a ply within a laminate. Deviations of the out-of-plane centroid coordinate, aspect ratio and area are produced using the Markov Chain method, while the cross-correlated in-plane centroid position is generated by the Series Expansion procedure. A good comparison in terms of wavelengths and extreme values is obtained between the experimental and simulated deviations trends for all properties. Further, all simulated tow deviations achieve the target statistics on average. The virtual WiseTex model reflects the variation observed and quantified from the 2/2 twill composite sample.

ACKNOWLEDGEMENTS

This study is supported by the Flemish Government through the Agency for Innovation by Science and Technology in Flanders (IWT) and FWO-Vlaanderen. The authors would also like to acknowledge KU Leuven.

REFERENCES

[1] C.C Chamis, Probabilistic simulation of multi-scale composite behaviour, *Theoretical and applied fracture mechanics*, **41**, 51–61, 2004.

[2] M.M. Kamiński, Computational mechanics of composite materials: sensitivity, randomness and multiscale behaviour, *Springer*, London, 2005.

[3] I. Verpoest and S.V. Lomov, Virtual Textile Composites Software WiseTex: Integration with Micro-mechanical, Permeability and Structural Analysis, *Composites Science and Technology*, **65**, 2563–2574, 2005.

[4] Hexcel, *HexForce G0986 SB 1200*, USA, 2012.

[5] H. Bale, M. Blacklock, M.R. Begley, D.B. Marshall, B.N. Cox and R.O. Ritchie, Characterizing three-dimensional textile ceramic composites using synchrotron X-ray micro-computed-tomography, *Journal of the American Ceramic Society*, **95**, 392–402, 2012.

[6] A. Vanaerschot, B.N. Cox, S.V. Lomov, D. Vandepitte, Stochastic framework for quantifying the geometrical variability of laminated textile composites using micro-computed tomography, *Composites Part A*, **44**, 122–131, 2013.

[7] A. Vanaerschot, B.N. Cox, S.V. Lomov, D. Vandepitte, Stochastic multi-scale modelling of textile composites based on internal geometry variability, *Computers & Structures* , **122**, 55–64, 2013.

[8] A. Vanaerschot, B.N. Cox, S.V. Lomov, D. Vandepitte, Stochastic characterisation of the in-plane tow centroid in textile composites to quantify the multi-scale variation in geometry, in M. Papadrakakis, editor, *Proceedings of the IUTAM symposium on multiscale modeling and uncertainty quantification of materials and structures*, Santorini, Greece, September 8-11, 2013.

[9] A. Vanaerschot, B.N. Cox, S.V. Lomov, D. Vandepitte, Simulation of the cross-correlated positions of in-plane tow centroids in textile composites based on experimental data, *Composite Structures* , **116**, 75–83, 2014.

[10] M. Blacklock, H. Bale, M. Begley and B. Cox, Generating virtual textile composite specimens using statistical data from micro-computed tomography: 1D tow representations for the Binary Model, *Journal of the Mechanics and Physics of Solids*, **60**, 451–470, 2012.

[11] M. Vořechovský, Simulation of simply cross-correlated random fields by series expansion methods, *Structural Safety*, **30**, 337–363, 2008.

[12] R. Ghanem and P.D. Spanos, Stochastic Finite Elements: a Spectral Approach, *Springer-Verlag*, New York, 2000.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

MULTISCALE MODELING OF THERMO-ELASTIC PROPERTIES OF MICROCRACKED MATERIALS

Antonino Favata, Patrizia Trovalusci, and Renato Masiani

Department of Structural and Geotechnical Engineering
Sapienza University of Rome
via Antonio Gramsci 53
e-mail: {antonino.favata, patrizia.trovalusci, renato.masiani}@uniroma1.it

Keywords: Continua with microstructure, heterogeneous materials, non-classical thermal conduction, microtemperature, coarse-graining.

Abstract. *A general framework able to describe the mechanical behavior of thermo-elastic materials with microcracks is proposed. The main features of our model come from the definition of additional kinetics descriptors, both mechanical and thermal (multifield model). The model is thermodynamically consistent and turns out to be non-local, in that it retains memory of the fine material structure through internal lengths and dispersion properties. It is shown that the multifield model allows us to reveal the presence of the microcracks through a dispersive oscillating behavior and a temperature decrease of microcracks, circumventing well known problems related to standard wave propagation and heat conduction.*

1 Introduction.

The mechanical behavior of many materials of growing interest in materials science (such as composites, granular materials, alloys, liquid crystals) is often strongly influenced by an existing or emergent microstructure (such as microcracks, voids, defects, dislocations, phases in multiphase materials). Due to the different material properties of the microstructure, the macroscopic material turns out to be highly anisotropic and eventually inhomogeneous. Several approaches to the modeling of such materials can be found in the literature, and for a synthetic description the reader is referred to [30] and references therein.

A fine description of such materials often requires an excessive computational effort, otherwise, a gross description within the so-called *generalized continuum mechanics* can be more suitable, on extending the conventional mechanics in order to incorporate intrinsic microstructural effects in the mechanical behavior [19, 8, 10, 2]. In this contribution we fit into a general framework of this kind, able to describe the mechanical behavior of thermo-elastic materials with microcracks. In particular we propose a continuum model endowed with a mechanical and thermal structure defined for predicting the macroscopic behavior of a material with distributed microcracks (due to manufacturing defects or lack of cohesion). To fix ideas, we focus on porous metal-ceramic composites (MCC, CMC) such as tungsten or titanium/molybdenum carbides (WC/Co, TiC/Mo_2C), alumina/zirconia materials (Al_2O_3/ZrO_2).

Our model is connected to a discrete description of the matter, with the perspective to a proper constitutive identification. Besides the standard descriptors – displacement and temperature – additional terms are introduced in an enhanced principle of virtual powers formulation. These additional microstructural descriptors respectively represent: on the mechanical side, the mean jump of the displacement field due to the presence of microcracks; while on the thermal side, the nonstandard thermal state variable of the thermal displacement, firstly considered by Helmholtz, and an additional microstructural kinetic variable, whose time derivative represents the mean jump of the temperature field due to the presence of microcracks. This model is non-local, in that it retains memory of the fine material structure through internal lengths and dispersion properties, and can be considered as a *continuum with microstructure* according to the definition by Capriz [8].

The continuum here described, with just reference to the mechanical structure, has been proposed in [18] and developed in [31, 37, 30] with applications to particle/fiber reinforced polymer/ceramic matrix and masonry-like materials. Recently, the microstructure has been extended to the thermomechanical framework in [13], stemming from the ideas exposed in [11], where a simple and paradigmatic example of continua with micro-structure is exposed within the beams theory framework. An application of a two-temperatures model to rod theory can be also found in [1]. More in general, microtemperatures have been considered in [38] and the entropy balance postulated by Green and Naghdi [16], extensively discussed in [**?**, 6], is used to obtain a new theory of heat for materials with inner structure. A more exhaustive reference framework on this topic is reported in [13].

The PDEs governing equations obtained have been tested in a simple one–dimensional setting, where we have shown the potentiality of the theory laid down. The performed numerical tests show that the presence of microcracks has twofold effect: on one side it makes the displacement and the temperature field wave-like and oscillating, what reveals the presence of the microcracking itself; on the other side, it makes the final temperature lower than in the case without microcracks, as one would expect.

2 Discrete–to–scale–dependent continuum model: towards a constitutive identification

A model of generalized continuum, as the one we propose, may be computationally efficient provided that a proper physical meaning of the additional fields is provided and a constitutive identification for the internal and the external actions is performed [31, 37, 27, 33, 28, 3, 14, 20, 4, 12, 2, 36]. In [36, 29] it has been shown that the Principle of Virtual Powers (PVP) can be expedient to perform coarse-graining procedures in the framework of the homogenization techniques. These corpuscular-continuous approaches are based on the assumption that a selected microscopic level structure of matter can be described as discontinuous, and that the transition from the coarse scale is governed by an *a priori* map between the large set of degrees of freedom of discrete systems and the deformation fields of the continuum. In this sense the procedure we propose can be connected to homogenization techniques adopted to derive constitutive models with internal variables; nevertheless an important difference has to be pointed out: when dealing with these techniques an effective stiffness of the material is obtained, depending on the properties and the distribution of the microcracks (see for instance [7, 17, 23]); analogous results can be obtained for the effective thermal conductivity [24]. In our approach, both the classical mechanical and thermal properties are unaltered, while self-actions (both mechanical and thermal), related to additional material properties, are added. The results of the two different approaches can be compared in terms of macroscopic displacement [29] and temperature.

Here below a synthetic and revisited version of the developments in [13], is reported. For further details the reader is referred to the quoted article.

At the microscopic level, the material is supposed to be composed of two interacting superposed lattices; the former that we will call *macrolattice* and denote by $\mathcal{L}$, is constituted of rigid particles representing the matrix, connected in pairs by elastic and conductive links, which carry only axial forces; the second lattice that we will call *microlattice* and denote by $\mathcal{L}'$, is made of interacting slits of arbitrary shape with a predominant dimension; they are thought as a device able to transmit to the macrolattice additional forces and thermal interaction, due to crack opening displacement. The two lattices are linked together by elastic and conductive bonds; in a purely mechanical framework, this approach has already been used in [18].

Let $\mathcal{A}$ and $\mathcal{B}$ two interacting particles of the macrolattice, whose centers occupy the positions $\mathbf{a}$ and $\mathbf{b}$. Let us consider two slits $\mathcal{A}'$ and $\mathcal{B}'$, whose centers are individuated by the positions $\mathbf{a}'$ and $\mathbf{b}'$. We endow both lattices with two physical structures, the one mechanical the other thermal; for both the structures, we define the kinetic variables. Let $\mathbf{u}_i$ the relative displacement between two points $\mathbf{u}_i = \mathbf{u}^a - \mathbf{u}^b$ of $\mathcal{L}$. Let the crack opening displacement over $\mathcal{A}'$ be indicated by $\mathbf{d}^{a'}$, and let $\mathbf{d}_j = \mathbf{d}^{a'} - \mathbf{d}^{b'}$ be the relative displacement jump between two interacting slits $\mathcal{A}'$ and $\mathcal{B}'$. We denote by $\vartheta_i = \vartheta^a - \vartheta^b$ the difference in temperature between $\mathcal{A}$ and $\mathcal{B}$; the corresponding kinetic variable we consider is the difference in thermal displacement $\alpha_i = \alpha^a - \alpha^b$, defined as $\dot{\alpha}_i = \vartheta_i$. Let the jump of temperature between the two sides of the slit $\mathcal{A}'$ be $\theta^{a'}$, while the relative temperature jump between two interacting slits $\mathcal{A}'$ and $\mathcal{B}'$ be $\theta_j = \theta^{a'} - \theta^{b'}$; consistently, we define $\beta^{a'}$ and β_j, with $\dot{\beta}^{a'} = \theta^{a'}$ and $\dot{\beta}_j = \theta_j$.

The strain measures of the $\mathcal{L}$ is $\mathbf{u}_i$, while the strain measure for $\mathcal{L}'$ is $\mathbf{d}_j$; for each pair of interacting particle-slit, a further strain measure is assumed as the vector $\mathbf{u}_l = \mathbf{u}^a - (\mathbf{u}^{a'} + \gamma^{a'}\mathbf{d}^{a'})$, with γ a function of $(|\mathbf{a} - \mathbf{a}'|)$ and of the number of the particles interacting with $\mathcal{A}'$.

The mechanical interaction between two particles through the i-th bond are represented by the vectors $\mathbf{t}_i^a$ and $\mathbf{t}_i^b$, with $\mathbf{t}_i^a = -\mathbf{t}_i^b =: \mathbf{t}_i$; the corresponding thermal interaction is given by the scalar quantities h_i^a and h_i^b, with $h_i^a = -h_i^b =: h_i$. The vectors $\mathbf{z}_j^{a'}$ and $\mathbf{z}_j^{b'}$, with $\mathbf{z}_j^{a'} = -\mathbf{z}_j^{b'} =: \mathbf{z}_j$ indicate the interactions between two slits along the j-th direction; the corresponding

thermal interactions are given by the scalar quantities $h_j^{a'}$ and $h_j^{b'}$, with $h_j^{a'} = -h_j^{b'} =: h_j$. The interacting force between the l-th bond will be denoted by $\mathbf{m}_l^a = -\mathbf{m}_l^{a'} = -\mathbf{m}_l$. The force due to the displacement $\mathbf{d}^{a'}$ is represented by the vector $\mathbf{z}_0^{a'}$; the zero-th order thermal interaction corresponding to the temperature of the particle $\mathcal{A}$ is denoted by h_0^a, those corresponding to the temperature of the particles $\mathcal{A}'$ is denoted by $h_0^{a'}$. All the defined fields are considered time-dependent.

The dynamics of the two lattices is specified by an extended notion of the internal virtual power, defined over the collection of virtual velocities ('primal' variables)

$$(\delta\mathbf{u}_i, \delta\mathbf{d}^{a'}, \delta\mathbf{d}_j; \delta\alpha^a, \delta\alpha_i, \delta\beta^{a'}, \delta\beta_j,), \tag{1}$$

defined as follows:

$$\begin{aligned}\delta\pi[(\delta\mathbf{u}_i, \delta\mathbf{d}^{a'}, \delta\mathbf{d}_j; \delta\alpha^a, \delta\alpha_i, \delta\beta^{a'}, \delta\beta_j,)] = \int_{\mathcal{T}} \Big(\sum_a h_0^a \delta\alpha^a + \sum_i (\mathbf{t}_i \cdot \delta\mathbf{u}_i + h_i \delta\alpha_i) \\ + \sum_{a'} (\mathbf{z}_0^{a'} \cdot \delta\mathbf{d}^{a'} + h_0^{a'} \delta\beta^{a'}) + \sum_j (\mathbf{z}_j \cdot \delta\mathbf{d}_j + h_j \delta\beta_j) + \sum_l \mathbf{m}_l \cdot \delta\mathbf{u}_l \Big),\end{aligned} \tag{2}$$

where the summations are extended to all the a particles, a' slits, i bonds between two generic elements of $\mathcal{L}$, j bonds between two generic elements of $\mathcal{L}'$ and all the l links between an element of $\mathcal{L}$ and an element of $\mathcal{L}'$; $\mathcal{T}$ is the observation time interval.

Consistently with the molecular theory of elasticity [9], to account for short range interactions, it is sufficient to consider homogeneous deformations by assuming that the lattice descriptors are connected to continuum fields as:

$$\begin{aligned}\mathbf{u}^a &= \mathbf{u}(\mathbf{x}) + \nabla\mathbf{u}(\mathbf{x})(\mathbf{a} - \mathbf{x}), \\ \mathbf{d}^{a'} &= \mathbf{d}(\mathbf{x}) + \nabla\mathbf{d}(\mathbf{x})(\mathbf{a}' - \mathbf{x}), \\ \alpha^a &= \alpha(\mathbf{x}) + \nabla\alpha(\mathbf{x}) \cdot (\mathbf{a} - \mathbf{x}), \\ \beta^{a'} &= \beta(\mathbf{x}) + \nabla\beta(\mathbf{x}) \cdot (\mathbf{a}' - \mathbf{x}),\end{aligned} \tag{3}$$

where: $\mathbf{u}$ is the *mechanical displacement*, whose time derivative $\mathbf{v} := \dot{\mathbf{u}}$ is the *velocity*; $\mathbf{d}$ is the *mechanical micro-displacement* $\mathbf{d}$, whose time derivative is the *micro-velocity* denoted by $\mathbf{l} := \dot{\mathbf{d}}$; α is the *thermal displacement*, whose time derivative $\vartheta := \dot{\alpha}$ represents the *absolute temperature*; β is the *thermal micro-displacement*, whose time derivative is $\theta := \dot{\beta}$, that we call *micro-temperature*. The field $\mathbf{d}$ is the perturbation induced on the regular macroscopic displacement field by the mean displacement jump which it would have over the actual microcrack; it can be interpreted as the difference between the actual displacement field in the body with microcracks and the regular displacement field of the same body without microcracks. The origin of this interpretation is investigated in [18] and [30]. Analogously, β is the perturbation induced on the regular macroscopic temperature field by the mean temperature jump which it would have over the actual microcrack as introduced in [13].

Thus, the following strain measures of the lattice system can be defined:

$$\begin{aligned}&\mathbf{u}_i = \nabla\mathbf{u}(\mathbf{a} - \mathbf{b}), \quad \mathbf{d}^{a'} = \nabla\mathbf{d}(\mathbf{a}' - \mathbf{b}'), \quad \mathbf{u}_l = \nabla\mathbf{u}(\mathbf{a} - \mathbf{a}') - \gamma^{a'}\big(\mathbf{d}(\mathbf{x}) + \nabla\mathbf{d}(\mathbf{a}' - \mathbf{x})\big), \\ &\alpha_i = \nabla\alpha \cdot (\mathbf{a} - \mathbf{b}), \quad \beta_j = \nabla\beta \cdot (\mathbf{a}' - \mathbf{b}'),\end{aligned} \tag{4}$$

where, for sake of simplicity, we omitted the dependence on $\mathbf{x}$. The above assumptions correspond to the hypothesis of regularity on the influence of microcracks on the deformation of

the body and come out from a generalized Cauchy-Born relation used to by-pass the intrinsic limitation of local descriptions avoiding the resort to non-homogeneous deformations [32, 30]. Considering the maps (3) and after some algebra, the mean virtual power of the system can be written as:

$$\delta\Pi[(\delta\mathbf{u},\delta\mathbf{d},\delta\alpha,\delta\beta)] = \frac{1}{V}\delta\pi[(\delta\mathbf{u}_i,\delta\mathbf{d}^{a'}\delta\mathbf{d}_j;\delta\alpha^a,\delta\alpha_i,\delta\beta^{a'},\delta\beta_j,)]$$
$$= \int_{\mathcal{T}} \big(\mathbf{S}\cdot\nabla\delta\mathbf{u} + \mathbf{z}\cdot\delta\mathbf{d} + \mathbf{Z}\cdot\nabla\delta\mathbf{d} + h\delta\alpha + \bar{\mathbf{h}}\cdot\nabla\delta\alpha + k\delta\beta + \bar{\mathbf{k}}\cdot\nabla\delta\beta\big), \tag{5}$$

where:

$$\begin{aligned}
\mathbf{S} &= \frac{1}{V}\Big(\sum_i \mathbf{t}_i\otimes(\mathbf{a}-\mathbf{b}) + \sum_l \mathbf{m}_l\otimes(\mathbf{a}-\mathbf{a}')\Big),\\
\mathbf{z} &= \frac{1}{V}\Big(\sum_{a'} \mathbf{z}_0^{a'} - \sum_{a'}\gamma^{a'}\mathbf{m}_l\Big),\\
\mathbf{Z} &= \frac{1}{V}\left(\sum_{a'}\mathbf{z}_0^{a'}\otimes(\mathbf{a}'-\mathbf{x}) + \sum_j \mathbf{z}_j\otimes(\mathbf{a}'-\mathbf{b}') - \sum_l \gamma^{a'}\mathbf{m}_l\otimes(\mathbf{a}'-\mathbf{a})\right),\\
h &= \frac{1}{V}\sum_a h_0^a,\\
k &= \frac{1}{V}\sum_{a'} h_0^{a'},\\
\bar{\mathbf{h}} &= \frac{1}{V}\sum_i h_i(\mathbf{a}-\mathbf{b}),\\
\bar{\mathbf{k}} &= \frac{1}{V}\left(\sum_{a'} h_0^{a'}(\mathbf{a}'-\mathbf{x}) + \sum_j h_j(\mathbf{a}'-\mathbf{b}')\right),
\end{aligned} \tag{6}$$

are the interaction measures and V is the volume of a representative volume element (RVE), easily to be defined if we assume a periodic microstructure but also detectable for non periodic assemblies [34, 35]. Note that the fabric vectors $\mathbf{a}'-\mathbf{x}$, $\mathbf{a}'-\mathbf{b}'$ and $\mathbf{a}'-\mathbf{a}$ account for the geometry of the microstructure, while the directions of the forces $\mathbf{z}_0^{a'}$ and $\mathbf{z}_j$ account for the orientation of the slits. On assuming the equivalence of the mean virtual power of the lattice system $\delta\Pi$ (5) with the virtual power density of a corresponding multifield continuum:

$$\delta w[(\delta\mathbf{u},\delta\mathbf{d},\delta\alpha,\delta\beta)] = \int_{\mathcal{T}} \big(\mathbf{S}\cdot\nabla\delta\mathbf{u} + \mathbf{z}\cdot\delta\mathbf{d} + \mathbf{Z}\cdot\nabla\delta\mathbf{d} + h\delta\alpha + \bar{\mathbf{h}}\cdot\nabla\delta\alpha + k\delta\beta + \bar{\mathbf{k}}\cdot\nabla\delta\beta\big), \tag{7}$$

by means of the localization theorem, we can recognize that Equations (6) define the stress and the thermodynamic measures of the equivalent continuum. $\mathbf{S}$, $\mathbf{z}$, $\mathbf{Z}$, h, $\bar{\mathbf{h}}$, k, $\bar{\mathbf{k}}$ are measures of the *mechanical* and *thermal interactions* that a point of the continuum body exchanges with its adjacent part, accounting for both macro- and micro-description, dual to the macroscopic kinetic variables.

Then, by assuming constitutive response functions for the mechanical and thermal interactions and in the discrete model we can obtain the effective constitutive relations of the multifield

continuum in the form [13]:

$$\begin{aligned}
&\mathbf{S} = \mathbb{C}\mathbf{E} + \mathbb{D}\nabla\mathbf{d} + (\vartheta - \bar{\vartheta})\mathbf{M} + (\theta - \bar{\theta})\mathbf{\Gamma},\\
&\mathbf{z} = \mathbf{G}\mathbf{d},\\
&\mathbf{Z} = \mathbb{R}\nabla\mathbf{d} + \mathbb{D}\mathbf{E} + (\vartheta - \bar{\vartheta})\mathbf{N} + (\theta - \bar{\theta})\mathbf{\Lambda},\\
&\mathbf{h} = -\chi\nabla\alpha - \nu\nabla\beta,\\
&\mathbf{k} = -\lambda\nabla\beta - \nu\nabla\alpha,
\end{aligned} \tag{8}$$

where $\bar{\vartheta}$ and $\bar{\theta}$ are given reference temperatures and $\mathbb{C}$, $\mathbb{D}$, $\mathbf{M}$, $\mathbf{\Gamma}$, $\mathbf{G}$, $\mathbb{R}$, $\mathbf{N}$, $\mathbf{\Lambda}$ are the constitutive terms, mechanical and thermal. The quantities h and k in 6 have the role of dissipation (see [13] for details on this issue) and then, as it customarily happens in classical thermo-elasticity, we set:

$$h = \mathbf{q}\cdot\nabla\vartheta, \quad k = \widetilde{\mathbf{q}}\cdot\nabla\theta, \tag{9}$$

where $(\mathbf{q}, \widetilde{\mathbf{q}})$ are the *heat influxes*, that we set, as usually done [5, 22], proportional to the entropy influxes through the temperature as follows:

$$\mathbf{q} = \vartheta\mathbf{h}, \quad \widetilde{\mathbf{q}} = \theta\mathbf{k}. \tag{10}$$

3 Balance equations

The structure of the continuum body identified is encoded in the virtual power density formula (7), and it is perceived ad a continuum with local vector structure.

Following the approach described in [13], by requiring the equivalence of the virtual internal and external power of such a continuum (generalized Principle of Virtual Power as in [21]) we derive following local balance equations:

(i) *momenta balances*:

$$\begin{aligned}
&\dot{\mathbf{p}} = \operatorname{div}\mathbf{S} + \mathbf{b},\\
&\dot{\mathbf{r}} = \operatorname{div}\mathbf{Z} - \mathbf{z} + \mathbf{g};
\end{aligned} \tag{11}$$

(ii) *entropies balances*:

$$\begin{aligned}
&\dot{\eta} = -\operatorname{div}\mathbf{h} - h + b,\\
&\dot{\omega} = -\operatorname{div}\mathbf{k} - k + g,
\end{aligned} \tag{12}$$

in $\mathcal{P}\times\mathcal{T}$, $\mathcal{P}$ being an arbitrary part of the body, $\mathbf{h} := -\bar{\mathbf{h}}$ and $\mathbf{k} := -\bar{\mathbf{k}}$, together with the boundary conditions:

$$\mathbf{S}\mathbf{n} = \mathbf{s}, \quad \mathbf{Z}\mathbf{n} = \mathbf{y}, \quad \bar{\mathbf{h}}\cdot\mathbf{n} = s, \quad \bar{\mathbf{k}}\cdot\mathbf{n} = y, \tag{13}$$

on $\partial\mathcal{P}\times\mathcal{T}$ and the initial conditions:

$$\mathbf{p}(x,t_i) = \mathbf{p}_i(x), \quad \mathbf{r}(x,t_i) = \mathbf{r}_i(x), \quad \eta(x,t_i) = \eta_i(x), \quad \omega(x,t_i) = \omega_i(x) \tag{14}$$

for $x\in\mathcal{P}$, where the fields $\mathbf{b}$, $\mathbf{g}$, b and g account for both the macro and the micro external *at-a-distance interactions*, respectively mechanical and thermal; the fields $\mathbf{s}$, $\mathbf{y}$, s and y account for the *contact interactions*; $\mathbf{p}$ is interpreted as the *linear momentum*, $\mathbf{r}$ the *microscopic linear momentum*, η the *entropy*, ω the *micro-entropy*, being all these fields defined per unit referential volume. As Equations (12) reveal, $\mathbf{h}$ and $\mathbf{k}$ can be interpreted as the *entropy* and *micro-entropy influxes*, respectively; the fields h and k are measures of *internal dissipation*, as anticipated in

the previous section. The extended version of the PVL, whence (11) and (12) come out, is based upon the introduction of: (i) a space-time integral, responsible of the appearance of time derivatives, and (ii) those thermal kinetic variables already introduced in the discrete system.

On substituting (8), (9) in (11) and (12), we obtain a system of partial differential equations in the unknowns $\mathbf{u}$, $\mathbf{d}$, α and β. It is worth noting that if, in order to have a linear system of PDEs, we neglect the quadratic quantities [15] and, in order to have a system in terms of the temperature where more meaningful boundary conditions can be setted, by differenting with respect to time the heat equations (12) become:

$$\begin{aligned} c\ddot{\vartheta} &= \chi\Delta\vartheta + \nu\Delta\theta + \mathbf{M}\cdot\nabla\ddot{\mathbf{u}} + \mathbf{N}\cdot\nabla\ddot{\mathbf{d}}, \\ \xi\ddot{\theta} &= \lambda\Delta\theta + \nu\Delta\vartheta + \mathbf{\Gamma}\cdot\nabla\ddot{\mathbf{u}} + \mathbf{\Lambda}\cdot\nabla\ddot{\mathbf{d}}. \end{aligned} \tag{15}$$

4 Numerical simulations

In this Section we present some numerical results confining our attention to a one-dimensional body $\mathcal{C}$ – i.e. a bar – whose typical point x lies on a line of direction $\mathbf{e}$.

We focus the attention to a conductive bar without microcracks versus a conductive bar with microcracks, referring the reader to [37] for underlying the ability of the purely mechanical multifield model to account for the presence of distributed microcracks. In this simple setting, the momenta balances (11) and the heat equations (15) read:

$$\begin{aligned} &\varrho\ddot{u} - Cu'' - Dd'' - M\theta' - \Gamma\theta' = 0, \\ &\mu\ddot{d} - Du'' - Rd'' + Gd - N\vartheta' - \Lambda\theta' = 0, \\ &c\ddot{\vartheta} = \chi\vartheta'' + \nu\theta'' + M\ddot{u}' + N\ddot{d}', \\ &\xi\ddot{\theta} = \lambda\theta'' + \nu\vartheta'' + \Gamma\ddot{u}' + \Lambda\ddot{d}', \end{aligned} \tag{16}$$

where we have set $\mathbf{p} = \varrho\ddot{u}\mathbf{e}$ and $\mathbf{r} = \mu\ddot{d}\mathbf{e}$, with ϱ and μ mass densities per unit length [37].

In order to evaluate the effects of microcracks and microtemperature in heat conduction, we here assume the coupling terms between the purely mechanical fields and the others are null (i.e. $M = D = \Gamma = 0$). As to the initial and boundary conditions for the temperature fields, we set

$$\begin{aligned} &\vartheta(x,0) = 0, \quad \vartheta(a,t) = t^2, \quad \vartheta(b,t) = 0, \quad \dot{\vartheta}(x,0) = 0, \\ &\theta(x,0) = 0, \quad \theta(a,t) = \frac{t^2}{100}, \quad \theta(b,t) = 0, \quad \dot{\theta}(x,0) = 0. \end{aligned} \tag{17}$$

The choice to set for the micro-temperature at the point a a value of $1/100$ of the macro-temperature is suggested by the physical requirement that the effects of the mean jump of the temperature due to microcracks is much smaller than the mean temperature itself. We define as visible temperature the field $\vartheta(x,t) - \theta(x,t)$.

In Fig. 1 we plot the results of our model vs the temperature in a bar without cracks, satisfying the following heat equation, consistent with $(16)_4$

$$c\ddot{\vartheta} = \chi\vartheta''. \tag{18}$$

If we compare the results of (18) with the model with microtemperature, we note that the presence of the microcracks has a double effect: on one side it makes the temperature field wave-like and oscillating; on the other side, it makes the final temperature lower than in the case without microcracks, as one would expect. Moreover, it is possible to show that the wave

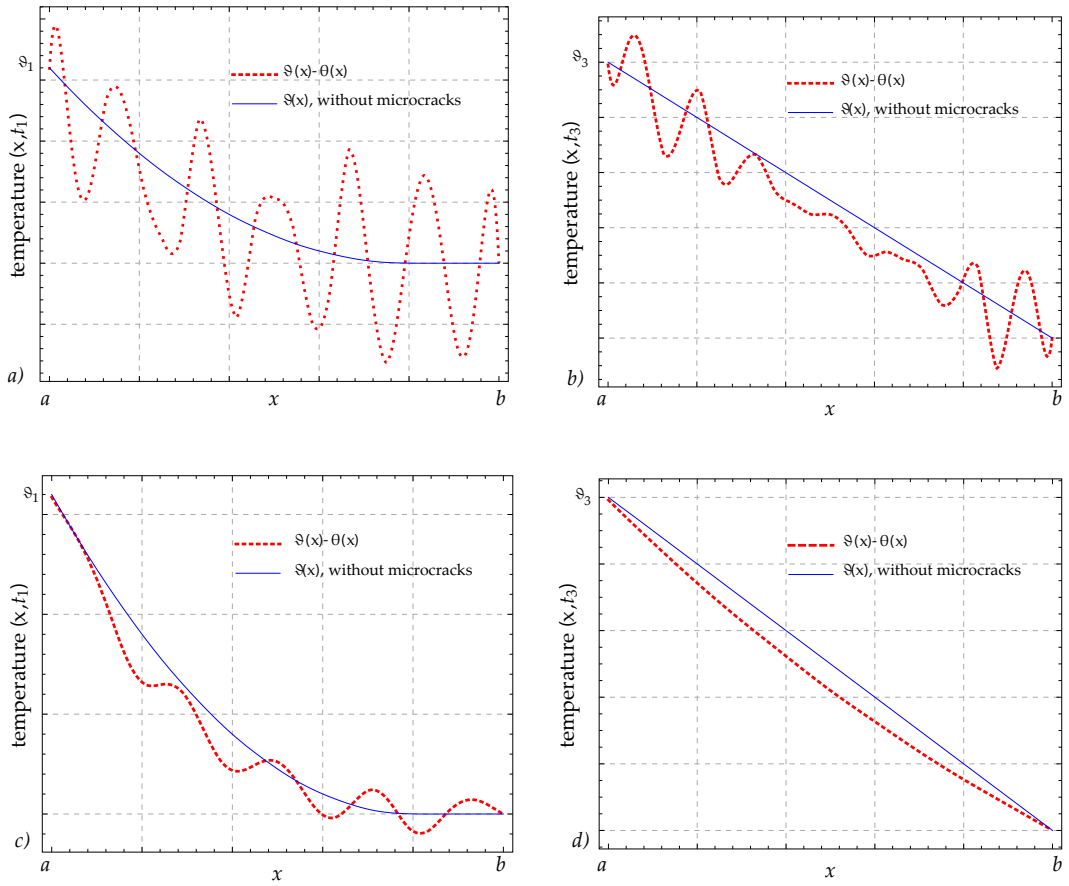


Figure 1: (a) and (b) Temperature in a bar without cracks (ϑ, continue blue line) and visible temperature in a bar with microcracks ($\vartheta - \theta$, dotted red curve), at two different time steps. (c) and (d) Same simulation with higher density of microcracks.

length is related to the microcracks density, what modifies the oscillatory behavior (see Fig. (1) (c)-(d)).

5 Final remarks

We have proposed a multiscale approach, within the framework of the generalized continuum thermo-mechanics, able to investigate the mechanical and thermal properties of a microcracked solid. The model is connected to a discrete description of the matter, with a perspective to a proper constitutive identification. Besides the standard descriptors – displacement and temperature – additional terms are introduced, able to take into account the presence of microcracks. The model proposed is non-local, in that it retains memory of the fine material structure through internal lengths and dispersion properties. The PDEs system obtained has been tested in a simple one-dimensional setting, where we have shown the potentiality of the theory laid down. Focusing in particular on the the thermal properties, we have shown that our model is able to accommodate a non-standard thermal propagation; our numerical tests show that the presence of microcracks has twofold effect: on one side it makes the temperature field wave-like and oscillating, what reveals the presence of the microcracking itself; on the other side, it makes the final temperature lower than in the case without microcracks, as one would expect.

REFERENCES

[1] H. Altenbach, M. Bîrsan, and V.A. Eremeyev. On a thermodynamic theory of rods with two temperature fields. *Acta Mech.*, 223:1583–1596, 2012.

[2] H. Altenbach and V. A. Eremeyev (ed). *Generalized Continua from the Theory to Engineering Applications*, volume 541 of CISM Courses and Lectures. Springer, Berlin, 2013.

[3] A. Bacigalupo and L. Gambarotta. Non-local computational homogenization of periodic masonry. *Int. J. Multisc. Comp. Eng.*, 9(5):796–811, 2011.

[4] C. Bajaj, A. Favata and P. Podio-Guidugli. On a nanoscopically-informed shell theory of carbon nanotubes. *Europ. J. Mech. A/Solids*, 42:137–157, 2013.

[5] S. Bargmann, A. Favata and P. Podio-Guidugli. On energy and entropy influxes in the Green-Naghdi type III theory of heat conduction. *Proc. R. Soc. Lond. A*, 469(2152), 2013.

[6] S. Bargmann, A. Favata, and P. Podio-Guidugli. A revised exposition of the Green-Naghdi theory of heat propagation. *J. of Elast.*, pages 1–12, 2013.

[7] B. Budiansky and R.J. O'Connell. Elastic moduli of a cracked solid. *Int. J. Sol. Struct.*, 12(2):81–97, 1976.

[8] G. Capriz. *Continua with Microstructure*. Springer, 1989.

[9] J. L. Ericksen. Special topics in elastostatics. In C. S. Yih, editor, *Advances in Applied Mechanics*, volume 17, pages 189–244. Academic Press, New York, 1977.

[10] A.C. Eringen. *Microcontinuum Field Theories*. Springer New York, 1999.

[11] A. Favata. A beam theory consistent with three-dimensional thermo-elasticity. *Math. Mech. Sol.*, doi: 10.1177/1081286514524974, 2014.

[12] A. Favata and P. Podio-Guidugli. A shell theory for chiral single-wall carbon nanotubes. *Europ. J. Mech. A/Solids*, 45:198–210, 2014.

[13] A. Favata, P. Trovalusci and R. Masiani, A multiphysics and multiscale approach for modeling microcracked thermo-elastic materials. *Comp. Mat. Sci*, 116, pages 22-31.

[14] S. Forest and D. K. Trinh. Generalised continua and the mechanics of heterogeneous material. *ZAMM – Z. Angew. Math. Mech.*, 91:90–109, 2011.

[15] M. Frémond. The basic laws of mechanics. In *Phase Change in Mechanics*, volume 13 of *Lecture Notes of the Unione Matematica Italiana*, pages 5–29. Springer Berlin Heidelberg, 2012

[16] A.E. Green and P.M. Naghdi. On thermodynamics and the nature of the second law. *Proc. R. Soc. Lond. A*, 4:253–270, 1977.

[17] M. Kachanov. On the effective moduli of solids with cavities and cracks. *Int. J. Frac.*, 59(1):R17–R21, 1993.

[18] P.M. Mariano and P. Trovalusci. Constitutive relations for elastic microcracked bodies: From a lattice model to a multifield continuum description. *Int. J. Dam. Mech.*, 8(2):153–173, 1999.

[19] R.D. Mindlin. Micro-structure in linear elasticity. *Arch. Rat. Mech. Anal.*, 16(1):51–78, 1964.

[20] A. Pau and P. Trovalusci. Block masonry as equivalent micropolar continua: the role of relative rotations. *Acta Mech.*, 223(7):1455–1471, 2012.

[21] P. Podio-Guidugli. A virtual power format for thermomechanics. *Cont. Mech. Thermodyn.*, 20:479–487, 2009.

[22] P. Podio-Guidugli. On energy and entropy inflows in the theory of heat conduction. *Arch. Appl. Mech*, 85, 2015.

[23] T. Sadowski. Gradual degradation in two-phase ceramic composites under compression. *Comp. Mat. Sci.*, 64:209 – 211, 2012.

[24] T. Sadowski and P. Golewski. Heat transfer and stress concentrations in a two-phase polycrystalline composite structure. Part I: Theoretical modelling of heat transfer. *Materwiss. Werksttech.*, 44(5):497–505, 2013.

[25] T. Sadowski and L. Marsavina. Multiscale modelling of two-phase ceramic matrix composites. *Comp. Mat. Sci.*, 50(4):1336 – 1346, 2011.

[26] T. Sadowski and S. Samborski. Prediction of the mechanical behaviour of porous ceramics using mesomechanical modelling. *Comp. Mat. Sci.*, 28(3-4):512–517, 2003.

[27] V. Sansalone, P. Trovalusci and F. Cleri. Multiscale modeling of materials by a multifield approach: Microscopic stress and strain distribution in fiber-matrix composites. *Acta Mater.*, 54(13):3485–3492, 2006.

[28] C. Tekoglu and P.R. Onck. Size effects in two-dimensional Voronoi foams: A comparison between generalized continua and discrete models. *J. Mech. Phys. Sol.*, 56(12):3541–3564, 2008.

[29] P. Trovalusci. Discrete to scale-dependent continua for complex materials. A generalized Voigt approach using the virtual power equivalence. In P. Trovalusci (ed.), *Materials with Internal Structure. Multiscale and Multifield Modelling and Simulation*, Springer Tracts in Mechanical Engineering', Springer, 2015.

[30] P. Trovalusci. Molecular approaches for multifield continua: origins and current developments. In T. Sadowsky and P. Trovalusci (eds), *Multiscale Modelling of Complex Materials: phenomenological, theoretical and computational aspects*, number 556 in CISM Courses and Lectures, pages 211–278. Springer, Berlin, 2014.

[31] P. Trovalusci and G. Augusti. A continuum model with microstructure for materials with flaws and inclusions. *J. Phys. IV*, 8(8):Pr8–383–Pr8–390, 1998.

[32] P. Trovalusci, D. Capecchi, and G. Ruta. Genesis of the multiscale approach for materials with microstructure. *Arch. Appl. Mech.*, 79:981–997, 2009.

[33] P. Trovalusci and R. Masiani. Non-linear micropolar and classical continua for anisotropic discontinuous materials. *Int. J. Sol. Struct.*, 40:1281–1297, 2003.

[34] P. Trovalusci, M. Ostoja-Starzewski, M. L. De Bellis, and A. Murrali. Particulate random composites homogenized as micropolar materialsa. *Meccanica*, 49:2719–2727, 2015.

[35] P. Trovalusci, M. Ostoja-Starzewski, M. L. De Bellis, and A. Murrali. Scale-dependent homogenization of random composites as micropolar continua. *Eur. J. Mech. A/Sol.*, 49:396–407, 2015.

[36] P. Trovalusci and A. Pau. Derivation of microstructured continua from lattice systems via principle of virtual works. The case of masonry-like materials as micropolar, second gradient and classical continua. *Acta Mech.*, 225(1):157–177, 2014.

[37] P. Trovalusci, V. Varano, and G. Rega. A generalized continuum formulation for composite microcracked materials and wave propagation in a bar. *J. Appl. Mech.*, 77(6), 2010.

[38] C. Wózniak. Thermoelasticity of bodies with microstructures. *Archivum Mech. Stos.*, XIX:335, 1967.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

EFFECTIVE ELASTICITY COEFFICIENTS IN DRY POROUS MATERIALS. NUMERICAL AND SEMI-ANALYTICAL APPROACHES

Vladimir Sladek[1], **Bruno Musil**[2], **and Jan Sladek**[1]

[1] Institute of Construction and Architecture, Slovak Academy of Sciences
845 03 Bratislava, Slovakia
e-mail: vladimir.sladek@savba.sk, jan.sladek@savba.sk

[2] Institute of Mechanics, Bundeswehr University Munich
85579 Neubiberg, Germany
bruno.musil@gmail.com

Keywords: Empty pores, Homogenization, Representative volume elements, Micro-structural boundary value problems, Numerical experiments.

Abstract. *The paper deals with homogenization of linear elastic continuum involving empty pores. We concern some aspects of numerical approaches and confront them with the analytical ones in case of 2D elasticity. Of course, there are some restrictions for the analytical concept, because it is applicable only to few shapes of pores. Nevertheless, two or four variants of the analytical approach are developed in order to study the role of incorporation of interaction among the pores as well as the influence of outer boundary of microstructural subdomain and application of various physically admissible correlations between the volume averages for homogenized continuum and micro-structural one. In general, a numerical approach is necessary for solution of micro-structural boundary value problems in case of arbitrary shape and/or distribution of voids. Two approaches are proposed and mathematical models developed for numerical calculation of effective material coefficients for linear elastic continuum involving arbitrary empty pores. Appropriate micro-structural boundary value problems in the RVE are proposed for numerical analyses which are utilized for a posterior evaluation of effective material coefficients. Comparisons of results by the analytical and numerical approaches are discussed for the circular and elliptical pores. Finally, the influence of the shape of vacant pores and porosity on bending of elastic plates is illustrated in numerical simulations.*

1 INTRODUCTION

In practical engineering problems, the material media are not often homogeneous. If the size of material defects is much smaller than the characteristic length in structural design, the concept of homogenization is meaningful and applicable. Since the design is mostly the matter of numerical simulations, it is important to deal well with evaluation of effective material coefficients. Recently also multiscale models are utilized with different physical treatment on different length scales. In this paper, we shall distinguish between the macro-structure and micro-structure with using the linear elasticity as the physical base for description of phenomena on both dimensional scales. The difference consists in modelling micro-heterogeneities in micro-structure and characterization of macro-structure as statistically homogeneous linear elastic composite or defected materials based on the macroscopic or overall or equivalent elastic behavior. Although the subject of homogenization is classical (Voigt and Reuss mixture rules, Hashin and Shtrikman upper and lower bounds [2], self-consistency method [3], Mori-Tanaka method [4-6, 1]), there are still some open questions [7], e.g. the unproved Hill's and Mandel's conjecture statement.

2 HOMOGENIZATION BASED ON MICRO-STRUCTURAL MODELLING. EFFECTIVE MATERIAL COEFFICIENTS

The idea of homogenization is applicable to microscopically inhomogeneous materials only if these materials are macroscopically or statistically homogeneous in considered macrostructure. It means that there exist representative elements (RVEs) of the body under consideration. The RVE must obey the following requirements: (i) it is relatively small sample of the material, i.e. the constraints and loading on the surface of the macrostructure are uniform within the length l which is the linear size of the RVE ($l \ll L$ where L is the characteristic dimension of the macrostructure); (ii) it is sufficiently large as compared with the linear size of micro-inhomogeneity (a) in order the spatial wave-length ($\sim a$) of the stress and strain fluctuations about a mean value be small compared with l ($a \ll l$), and the effects of such fluctuations become insignificant within a few wave-lengths from the boundary of the RVE. Having solved micro-structural boundary value problems in RVE (the micro-constituents are assumed to be homogeneous elastic continua; the shape and distribution of micro-constituents or defects are abstracted from experiment), one can get volume averages of micro-fields over the RVE. The transition from micro- to macro-level (where the macro-structural problems could be solved in effective continuum which is macroscopically homogeneous) depends on finding the connections between suitably defined macro-variables and averages of micro-fields. There are two main questions: (i) how to define macro-variables and boundary data for the RVE in a physically meaningful way; (ii) whether and how the macro-variables (alone or in combinations) are related to the volume averages of their micro-counterparts.

Let us consider a macroscopically homogeneous body and denote by B the regular subregion occupied by a RVE of the same microscopically inhomogeneous 2D body composed of the homogeneous skeleton $\Omega \subset B$ and empty voids $B-\Omega$ with $\partial B = \Gamma$, $\partial\Omega = \Gamma \cup \Gamma_0$, Γ_0 being the boundary of the RVE, skeleton and voids in the RVE, respectively. The volume average or the mean values of the field variables are

$$\left\langle f^{eff} \right\rangle_B = \frac{1}{|B|}\int_B f^{eff}(\mathbf{x})d\Omega, \qquad \left\langle f^{ms} \right\rangle_\Omega = \frac{1}{|\Omega|}\int_\Omega f^{ms}(\mathbf{x})d\Omega \quad . \tag{1}$$

where the former relationship is for the effective field variable $f^{eff}(\mathbf{x})$ defined in the whole B, while the latter is used for the field variable $f^{ms}(\mathbf{x})$ related to the elastic skeleton (matrix mate-

rial) in micro-structural boundary value problems ($ms-bvp$). The mean values without detailed specification will be denoted as $\langle f \rangle_V$.

If $\left[u_i, \varepsilon_{ij}, \sigma_{ij}\right]$ are displacements, corresponding strains and self-equilibrated stress fields corresponding to solution of an elastic boundary value problem in V with boundary ∂V and tractions $t_i = \sigma_{ij} n_j$ ($\mathbf{n}$ is the outward unit normal vector on ∂V), then it can be shown

$$|V|\left\langle \sigma_{ij}\varepsilon_{ij} \right\rangle_V = \int_{\partial V} t_i u_i d\Gamma \tag{2}$$

$$\int_{\partial V} t_i x_j d\Gamma = |V|\left\langle \sigma_{ij} \right\rangle_V = |V|\left\langle \sigma_{ji} \right\rangle_V = \int_{\partial V} t_j x_i d\Gamma, \quad \frac{1}{2}\int_{\partial V}\left(t_i x_j + t_j x_i\right) d\Gamma = |V|\left\langle \sigma_{ij} \right\rangle_V \tag{3}$$

$$\int_{\partial V} u_i n_j d\Gamma = |V|\left\langle u_{i,j} \right\rangle_V, \quad \frac{1}{2}\int_{\partial V}\left(u_i n_j + u_j n_i\right) d\Gamma = |V|\left\langle \varepsilon_{ij} \right\rangle_V \tag{4}$$

$$|V|\left(\left\langle \sigma_{ij}\varepsilon_{ij} \right\rangle_V - \left\langle \sigma_{ij} \right\rangle_V \left\langle \varepsilon_{ij} \right\rangle_V\right) = \int_{\partial V}\left(t_i - n_k \left\langle \sigma_{ik} \right\rangle_V\right)\left(u_i - x_j \left\langle \varepsilon_{ij} \right\rangle_V\right) d\Gamma \tag{5}$$

Thus, the mean value of the deformation energy can be written in the separated view

$$\left\langle \sigma_{ij}\varepsilon_{ij} \right\rangle_V = \left\langle \sigma_{ij} \right\rangle_V \left\langle \varepsilon_{ij} \right\rangle_V \tag{6}$$

if the Hill conditions [ ] are satisfied:

(i) *kinematically uniform boundary conditions* (KUBC): $u_i\big|_{\partial V} = x_j \bar{\varepsilon}_{ij}$, $\bar{\varepsilon}_{ij} = const$, (7)

when $\left\langle \varepsilon_{ij} \right\rangle_V = \bar{\varepsilon}_{ij}$

(ii) *statically uniform boundary conditions* (SUBC): $t_i\big|_{\partial V} = n_k \bar{\sigma}_{ik}$, $\bar{\sigma}_{ik} = const$, (8)

when $\left\langle \sigma_{ik} \right\rangle_V = \bar{\sigma}_{ik}$.

The constitutive relationships in the linear elastic homogenized continuum are given as

$$\sigma_{ij}^{eff}(\mathbf{x}) = c_{ijkl}^{eff}\varepsilon_{kl}^{eff}(\mathbf{x}), \tag{9}$$

where $c_{ijkl}^{eff} = const$. Hence,

$$\left\langle \sigma_{ij}^{eff} \right\rangle_B = c_{ijkl}^{eff}\left\langle \varepsilon_{kl}^{eff} \right\rangle_B \tag{10}$$

and from the symmetry the stress and strain tensors, we have the following symmetry properties

$$c_{ijkl}^{eff} = c_{jikl}^{eff} = c_{ijlk}^{eff} \tag{11}$$

while the symmetry

$$c_{ijkl}^{eff} = c_{klij}^{eff} \tag{12}$$

results from the requirement of linear elasticity. In view of Eqs. (9) and (10), we could calculate the effective material coefficients c_{ijkl}^{eff} if we knew either $\left[\sigma_{ij}^{eff}(\mathbf{x}), \varepsilon_{kl}^{eff}(\mathbf{x})\right]$ at a point $\mathbf{x}$ or the mean values $\left[\left\langle \sigma_{ij}^{eff} \right\rangle_B, \left\langle \varepsilon_{kl}^{eff} \right\rangle_B\right]$. This would lead to an inverse problem. In order to eliminate the solution of the inverse problem, we try to identify the mean values $\left[\left\langle \sigma_{ij}^{eff} \right\rangle_B, \left\langle \varepsilon_{kl}^{eff} \right\rangle_B\right]$ with the some quantities obtained from the solution of $ms-bvp$ in the RVE.

If we consider a micro-structural KUBC in the RVE, it can be found (using the Green's function) that

$$\varepsilon_{ij}^{ms}(\mathbf{x}) = A_{ijkl}(\mathbf{x})\bar{\varepsilon}_{kl} \quad \text{in } B\ , \tag{13}$$

where $A_{ijkl}(\mathbf{x})$ is the influence tensor function corresponding to the KUBC. According to (4) and (7), we have $\left\langle \varepsilon_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \bar{\varepsilon}_{ij}$. Hence and from (13)

$$\left\langle A_{ijkl}(\mathbf{x}) \right\rangle_{\Omega} = \delta_{ik}\delta_{jl}\ . \tag{14}$$

Bearing in mind the constitutive law in the homogeneous skeleton Ω

$$\sigma_{ij}^{ms}(\mathbf{x}) = c_{ijkl}^{s}\varepsilon_{kl}^{ms}(\mathbf{x}) \tag{15}$$

and $\sigma_{ij}^{ms}(\mathbf{x}) = 0$ in $(B-\Omega)$, we obtain from (13) and (15)

$$\frac{|\Omega|}{|B|}\left\langle \sigma_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \left\langle \sigma_{ij}^{ms}(\mathbf{x}) \right\rangle_{B} = c_{ijkl}\left\langle A_{klrs}(\mathbf{x}) \right\rangle_{B}\bar{\varepsilon}_{rs}\ . \tag{16}$$

Since $\bar{\varepsilon}_{rs} = \left\langle \varepsilon_{rs}^{ms}(\mathbf{x}) \right\rangle_{\Omega}$, we can rewrite (16) as

$$\left\langle \sigma_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \frac{1}{1-p}c_{ijkl}^{s}\left\langle A_{klrs}(\mathbf{x}) \right\rangle_{B}\left\langle \varepsilon_{rs}^{ms}(\mathbf{x}) \right\rangle_{\Omega} \tag{17}$$

where we have introduced the porosity p defined as

$$p := \frac{|B|-|\Omega|}{|B|} = 1-\frac{|\Omega|}{|B|}\ . \tag{18}$$

If we adopted (as physically admissible) identification $\left[\left\langle \sigma_{ij}^{eff} \right\rangle_{B}, \left\langle \varepsilon_{kl}^{eff} \right\rangle_{B}\right] = \left[\left\langle \sigma_{ij}^{ms} \right\rangle_{\Omega}, \left\langle \varepsilon_{kl}^{ms} \right\rangle_{\Omega}\right]$, we would have form (17)

$$\left\langle \sigma_{ij}^{eff} \right\rangle_{B} = \frac{1}{1-p}c_{ijkl}^{s}\left\langle A_{klrs}(\mathbf{x}) \right\rangle_{B}\left\langle \varepsilon_{rs}^{eff} \right\rangle_{B} \tag{19}$$

and finally, in view of (10), we would have

$$c_{ijrs}^{eff} = \frac{1}{1-p}c_{ijkl}^{s}\left\langle A_{klrs}(\mathbf{x}) \right\rangle_{B}. \tag{20}$$

Similarly, if we consider a micro-structural SUBC in the RVE, it can be found that

$$\sigma_{ij}^{ms}(\mathbf{x}) = B_{ijkl}(\mathbf{x})\bar{\sigma}_{kl} \quad \text{in } B\ , \tag{21}$$

where $B_{ijkl}(\mathbf{x})$ is the influence tensor function corresponding to the SUBC. According to (3) and (8), we have $\left\langle \sigma_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \bar{\sigma}_{ij}$. Hence and from (21)

$$\left\langle B_{ijkl}(\mathbf{x}) \right\rangle_{\Omega} = \delta_{ik}\delta_{jl}\ . \tag{22}$$

Bearing in mind the constitutive law in the homogeneous skeleton Ω

$$\varepsilon_{ij}^{ms}(\mathbf{x}) = M_{ijkl}^{s}\sigma_{kl}^{ms}(\mathbf{x}) \tag{23}$$

(where M_{ijkl}^{s} is the tensor of skeleton compliances) and $\varepsilon_{ij}^{ms}(\mathbf{x}) = 0$ in $(B-\Omega)$, we obtain from (21) and (23)

$$\frac{|\Omega|}{|B|}\left\langle \varepsilon_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \left\langle \varepsilon_{ij}^{ms}(\mathbf{x}) \right\rangle_{B} = M_{ijkl}^{s}\left\langle B_{klrs}(\mathbf{x}) \right\rangle_{B}\bar{\sigma}_{rs}\ . \tag{24}$$

Since $\bar{\sigma}_{rs} = \left\langle \sigma_{rs}^{ms}(\mathbf{x}) \right\rangle_{\Omega}$, we can rewrite (24) as

$$\left\langle \varepsilon_{ij}^{ms}(\mathbf{x}) \right\rangle_{\Omega} = \frac{1}{1-p} M_{ijkl}^{s} \left\langle B_{klrs}(\mathbf{x}) \right\rangle_{B} \left\langle \sigma_{rs}^{ms}(\mathbf{x}) \right\rangle_{\Omega} \tag{25}$$

If we adopted (as physically admissible) identification $\left[\left\langle \sigma_{ij}^{eff} \right\rangle_{B}, \left\langle \varepsilon_{kl}^{eff} \right\rangle_{B} \right] = \left[\left\langle \sigma_{ij}^{ms} \right\rangle_{\Omega}, \left\langle \varepsilon_{kl}^{ms} \right\rangle_{\Omega} \right]$, we would have form (25)

$$\left\langle \varepsilon_{ij}^{eff}(\mathbf{x}) \right\rangle_{B} = \frac{1}{1-p} M_{ijkl}^{s} \left\langle B_{klrs}(\mathbf{x}) \right\rangle_{B} \left\langle \sigma_{rs}^{eff}(\mathbf{x}) \right\rangle_{B} \tag{26}$$

and finally in view of $\left\langle \varepsilon_{ij}^{eff} \right\rangle_{B} = M_{ijkl}^{eff} \left\langle \sigma_{kl}^{eff} \right\rangle_{B}$, we would have

$$M_{ijrs}^{eff} = \frac{1}{1-p} M_{ijkl}^{s} \left\langle B_{klrs}(\mathbf{x}) \right\rangle_{B}. \tag{27}$$

Finding of the influence functions is not as simple task, in general. In what follows, we shall consider a special case of elliptical pores in 2D elasticity problems, when the analytical solution can be utilized. Another considered case is the numerical approach, which is applicable to empty pores of arbitrary shape and randomly distributed in the RVE. We shall illustrate this approach on elliptical pores uniformly distributed in the body.

2.1 Analytical approaches

It is well known that the analytical solution is available for elasticity problem in infinite plane with an elliptical empty void and applied tension load $\sigma_{ij}^{\infty} = \sigma \delta_{i1} \delta_{j1}$ at infinity [8-10]. Denoting the major and minor semi-axes as a, b and selecting the RVE as square $l \times l$ with $l \gg a$, we can consider the analytical solution in infinite plane as certain approximation for the micro-structural boundary value problem in the RVE with prescribed constant traction load on the outer boundary and traction free boundary of the elliptical hole.

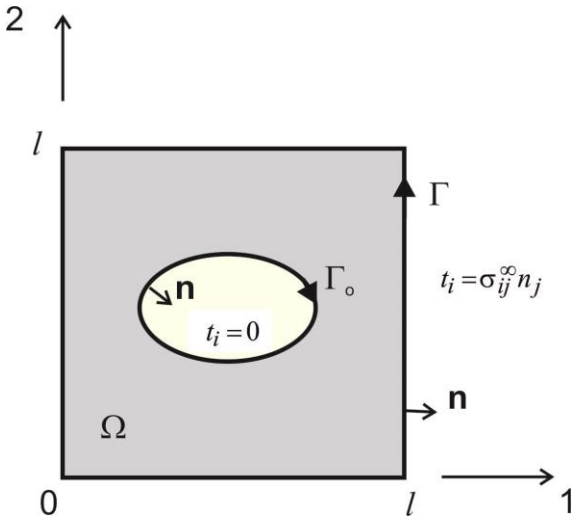


Fig. 1 Geometry of the quadrilateral RVE with one empty elliptical pore

Now, in view of Fig. 1, we can write

$$\left\langle \sigma_{ij}^{ms} \right\rangle_{\Omega} := \frac{1}{|\Omega|} \int_{\Omega} \sigma_{ij}^{ms}(\mathbf{x}) d\Omega = \frac{1}{2|\Omega|} \int_{\Omega} \left(\sigma_{ik}^{ms} x_{j,k} + \sigma_{jk}^{ms} x_{i,k} \right) d\Omega =$$

$$= \frac{1}{2|\Omega|}\int_{\Omega}\left[\left(\sigma_{ik}^{ms}x_j\right)_{,k}+\left(\sigma_{jk}^{ms}x_i\right)_{,k}\right]d\Omega = \frac{1}{2|\Omega|}\int_{\Gamma\cup\Gamma_0}\left(t_i^{ms}x_j+t_j^{ms}x_i\right)d\Gamma =$$

$$= \frac{1}{2|\Omega|}\int_{\Gamma}\left(t_i^{ms}x_j+t_j^{ms}x_i\right)d\Gamma = \frac{|B|}{|\Omega|}\sigma_{ij}^{\infty} = \frac{1}{1-p}\sigma_{ij}^{\infty} \tag{28}$$

$$\left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega} := \frac{1}{|\Omega|}\int_{\Omega}\varepsilon_{ij}^{ms}(\mathbf{x})d\Omega = \frac{1}{2|\Omega|}\int_{\Omega}\left(u_{i,j}^{ms}+u_{j,i}^{ms}\right)d\Omega = \frac{1}{2|\Omega|}\int_{\Gamma\cup\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma =$$

$$= \left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma}+\left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma_0} \tag{29}$$

$$\text{with } \left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma} := \frac{1}{2|\Omega|}\int_{\Gamma}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma, \quad \left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma_0} := \frac{1}{2|\Omega|}\int_{\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma. \tag{30}$$

Both these contributions to the mean value of micro-structural strains can be evaluated by using the exact solution for displacements in infinite plane. Apparently, in the linear theory of elasticity, both these integrals are to proportional to σ_{ij}^{∞}, i.e.

$$\left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma} = \frac{1}{1-p}M_{ijkl}\sigma_{kl}^{\infty}, \quad \left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma_0} = \frac{1}{1-p}H_{ijkl}\sigma_{kl}^{\infty}, \tag{31}$$

where $M_{ijkl}\sigma_{kl}^{\infty} := \frac{1}{2|B|}\int_{\Gamma}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma$, $H_{ijkl}\sigma_{kl}^{\infty} := \frac{1}{2|B|}\int_{\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma$.

What we need is to find certain physically meaningful correlation between the mean values of stresses and strains in effective medium and those found from solution of micro-structural bvp. First, consider the *integral force compatibility* and *average strain compatibility* (IF&ASC):

$$\text{IFC: } |B|\left\langle\sigma_{ij}^{eff}\right\rangle_B = |\Omega|\left\langle\sigma_{ij}^{ms}\right\rangle_{\Omega} \quad \Rightarrow \left\langle\sigma_{ij}^{eff}\right\rangle_B = (1-p)\left\langle\sigma_{ij}^{ms}\right\rangle_{\Omega} = \sigma_{ij}^{\infty} \tag{32}$$

$$\text{ASC: } \left\langle\varepsilon_{ij}^{eff}\right\rangle_B = \left\langle\varepsilon_{ij}^{ms}\right\rangle_{\Omega} = \frac{1}{1-p}\left(M_{ijkl}+H_{ijkl}\right)\sigma_{kl}^{\infty} \quad , \tag{33}$$

where we have utilized Eqs. (28)-(31). Since $\left\langle\varepsilon_{ij}^{eff}\right\rangle_B = M_{ijkl}^{eff}\left\langle\sigma_{ij}^{eff}\right\rangle_B$, we obtain from (32)-(33)

$$M_{ijkl}^{eff} = \frac{1}{1-p}\left(M_{ijkl}+H_{ijkl}\right) \tag{34}$$

If we shifted the outer boundary of the RVE to infinity, i.e. $\Gamma\to\Gamma^{\infty}$, then $M_{ijkl}\to M_{ijkl}^{s}$, where M_{ijkl}^{s} is the tensor of compliances corresponding to the matrix material of the skeleton. In view of such an approximation, we shall distinguish two analytical approaches within the IF&ASC concept:

(i) *analytical approach I*, denoted as **AA(I)**: $M_{ijkl}^{eff} = \frac{1}{1-p}\left(M_{ijkl}^{s}+H_{ijkl}\right)$ (35)

(ii) *analytical approach II*, denoted as **AA(II)**: $M_{ijkl}^{eff} = \frac{1}{1-p}\left(M_{ijkl}+H_{ijkl}\right)$ (36)

Another concept consists in consideration of *integral force compatibility* and *integral energy compatibility* (IF&EC):

$$\text{IEC:}\quad \left\langle \sigma_{ij}^{eff} \right\rangle_B \left\langle \varepsilon_{ij}^{eff} \right\rangle_B = \left\langle \sigma_{ij}^{ms} \right\rangle_\Omega \left\langle \varepsilon_{ij}^{ms} \right\rangle_\Omega \tag{37}$$

hence, in view of (32), we have

$$\left\langle \varepsilon_{ij}^{eff} \right\rangle_B = \frac{1}{1-p} \left\langle \varepsilon_{ij}^{ms} \right\rangle_\Omega$$

and finally in view of (29)-(31) and IFC (Eq.(32))

$$\left\langle \varepsilon_{ij}^{eff} \right\rangle_B = \frac{1}{1-p} \left\langle \varepsilon_{ij}^{ms} \right\rangle_\Omega = \frac{1}{(1-p)^2} \left(M_{ijkl} + H_{ijkl} \right) \sigma_{kl}^{\infty} = \frac{1}{(1-p)^2} \left(M_{ijkl} + H_{ijkl} \right) \left\langle \sigma_{ij}^{eff} \right\rangle_B .$$

Similar to the previous case, we shall distinguish other two analytical approaches:

(i) *analytical approach III*, denoted as **AA(III)**: $M_{ijkl}^{eff} = \frac{1}{(1-p)^2} \left(M_{ijkl}^{s} + H_{ijkl} \right)$ (38)

(ii) *analytical approach IV*, denoted as **AA(IV)**: $M_{ijkl}^{eff} = \frac{1}{(1-p)^2} \left(M_{ijkl} + H_{ijkl} \right)$ (39)

In order to evaluate the integrals M_{ijkl} and H_{ijkl}, we employ the exact values for displacements found from the analytical solution in infinite plane (see Fig. 2)

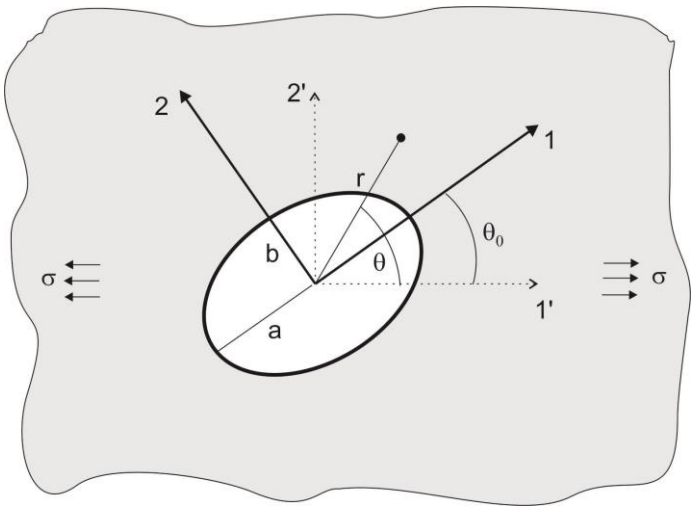


Fig. 2 Elliptical hole in infinite plane with applied uniform stress loading in infinity

$$u_1^{ms}(\rho,\theta) \approx u_1(\rho,\theta) = \frac{\sigma R}{4} \frac{1+\nu}{E_0} \left\{ \cos\theta \left[\Lambda \left(\rho - \frac{2m}{\rho} \right) + \frac{1}{A} \left(2\frac{1+m^2}{\rho} - \rho - 2m\frac{1+m^2}{\rho^3} + 2\frac{m^3}{\rho^5} \right) \right] + \right.$$

$$+ \cos(\theta + 2\theta_0) \left[\frac{2\Lambda}{\rho} + 2\rho + \frac{2}{m\rho} - \frac{1+m^2}{A} \frac{2}{m\rho} \right] + \cos(\theta - 2\theta_0) \frac{2}{A\rho^3} \left(1 + m^2 - \frac{m^2}{\rho^2} \right) +$$

$$\left. + \cos 3\theta \frac{m}{A\rho} \left(\frac{m}{\rho^2} - 2 \right) + \cos(3\theta + 2\theta_0) \frac{2}{A\rho} \right\} \tag{40}$$

$$u_2^{ms}(\rho,\theta)\approx u_2(\rho,\theta)=\frac{\sigma R}{4}\frac{1+\nu}{E_0}\left\{\sin\theta\left[\Lambda\left(\rho+\frac{2m}{\rho}\right)+\frac{1}{A}\left(2\frac{1+m^2}{\rho}-\rho+2m\frac{1+m^2}{\rho^3}-2\frac{m^3}{\rho^5}\right)\right]+\right.$$
$$+\sin(\theta+2\theta_0)\left[\frac{2}{m\rho}-\frac{2\Lambda}{\rho}-2\rho-\frac{2}{A}\frac{1+m^2}{m\rho}\right]-\sin(\theta-2\theta_0)\frac{2}{A\rho^3}\left(1+m^2-\frac{m^2}{\rho^2}\right)-$$
$$\left.-\sin 3\theta\frac{m}{A\rho}\left(2+\frac{m}{\rho^2}\right)+\sin(3\theta+2\theta_0)\frac{2}{A\rho}\right\} \tag{41}$$

Performing the required integrations, we obtain

$$\begin{pmatrix}H_{11kl}\sigma_{kl}^{\infty}\\H_{22kl}\sigma_{kl}^{\infty}\\H_{12kl}\sigma_{kl}^{\infty}\end{pmatrix}=\begin{pmatrix}H_{1111}&H_{1122}&2H_{1112}\\H_{2211}&H_{2222}&2H_{2212}\\H_{1211}&H_{1222}&2H_{1212}\end{pmatrix}\begin{pmatrix}\sigma_{11}^{\infty}\\\sigma_{22}^{\infty}\\\sigma_{12}^{\infty}\end{pmatrix},\quad [H]=p\frac{1+\nu}{E_0}\begin{pmatrix}B_1&B_2&0\\B_3&B_4&0\\0&0&B_5\end{pmatrix} \tag{42}$$

in which

$$H_{1111}=\frac{\pi Rb}{4|B|}\frac{1+\nu}{E_0}\left[\Lambda(3-2m)+3-m-m^2\right]=p\frac{1+\nu}{E_0}B_1\ ,\qquad B_1:=\frac{1+\gamma}{8}\left[\Lambda(3-2m)+3-m-m^2\right]$$

$$H_{1122}=\frac{\pi Rb}{4|B|}\frac{1+\nu}{E_0}\left[-\Lambda(1+2m)-(1+m+m^2)\right]=p\frac{1+\nu}{E_0}B_2\ ,\quad B_2:=\frac{1+\gamma}{8}\left[-\Lambda(1+2m)-(1+m+m^2)\right]$$

$$H_{2211}=\frac{\pi Ra}{4|B|}\frac{1+\nu}{E_0}\left[\Lambda(2m-1)-1+m-m^2\right]=p\frac{1+\nu}{E_0}B_3\ ,\qquad B_3:=\frac{1+\gamma}{8\gamma}\left[\Lambda(2m-1)-1+m-m^2\right]$$

$$H_{2222}=\frac{\pi Ra}{4|B|}\frac{1+\nu}{E_0}\left[\Lambda(3+2m)+3+m-m^2\right]=p\frac{1+\nu}{E_0}B_4\ ,\qquad B_4:=\frac{1+\gamma}{8\gamma}\left[\Lambda(3+2m)+3+m-m^2\right]$$

$$2H_{1212}=\frac{\pi R(a+b)}{2|B|}\frac{1+\nu}{E_0}(\Lambda+1)=p\frac{1+\nu}{E_0}B_5\ ,\qquad B_5:=\frac{1}{2}\left[1+\frac{1}{2}\left(\gamma+\frac{1}{\gamma}\right)\right](\Lambda+1) \tag{43}$$

with

$$p:=\frac{\pi ab}{l^2}\ ,\quad m=\frac{a-b}{a+b}=\frac{1-\gamma}{1+\gamma}\ ,\quad \gamma:=\frac{b}{a},\quad \Lambda=\frac{\lambda+3\mu}{\lambda+\mu}=3-4\tilde{\nu}\ , \tag{44}$$

where λ and μ, are Lame coefficients which are correlated with the Young modulus E_0 and Poisson ratio ν by

$$\mu=\frac{E_0}{2(1+\nu)}\ ,\quad \lambda=\mu\frac{2\tilde{\nu}}{1-2\tilde{\nu}}\ ,\quad \tilde{\nu}=\begin{cases}\nu/(1+\nu)\ , & \text{for plane stress problems}\\ \nu\ , & \text{otherwise}\end{cases} \tag{45}$$

The matrix of compliances of the homogeneous, isotropic and linear elastic skeleton is given as

$$\left[M^s\right]=\begin{pmatrix}M_{1111}^s&M_{1122}^s&2M_{1112}^s\\M_{2211}^s&M_{2222}^s&2M_{2212}^s\\M_{1211}^s&M_{1222}^s&2M_{1212}^s\end{pmatrix}=\frac{1+\nu}{4E_0}\begin{pmatrix}\Lambda+1&\Lambda-3&0\\\Lambda-3&\Lambda+1&0\\0&0&4\end{pmatrix}\ . \tag{46}$$

Note that in the case of elliptic void, we have not the closed form expression for the integrals in the matrix M_{ijkl}, though we have it in the case of circular void

$$M_{1111} = M^s_{1111} + \frac{1+\nu}{2E_0}\left[(\Lambda\pi+4)\frac{p}{\pi} + 2(\pi+2)\left(\frac{p}{\pi}\right)^2\right], \quad M_{1112} = 0,$$

$$M_{1122} = M_{2211} = M^s_{1122} - \frac{1+\nu}{4E_0}\left[((\Lambda-2)\pi+4)\frac{p}{\pi} - 2(\pi+2)\left(\frac{p}{\pi}\right)^2\right], \quad M_{2212} = 0, \tag{47}$$

$$2M_{1212} = 2M^s_{1212} + \frac{1+\nu}{E_0}\left[\Lambda p - 2(\pi+2)\left(\frac{p}{\pi}\right)^2\right], \quad M_{1211} = 0 = M_{1222}$$

Performing the limit $b \to a$, when $m \to 0$ and $\gamma \to 1$ in Eq. (43), we obtain the H_{ijkl} matrix for circular pore

$$B_1 \to 3\frac{\Lambda+1}{4} \text{ and } H_{1111} = 3\frac{Q}{E_0}, \quad B_4 \to 3\frac{\Lambda+1}{4} \text{ and } H_{2222} = 3\frac{Q}{E_0}, \quad Q = p\frac{\Lambda+1}{4}(1+\nu)$$

$$B_2 \to -\frac{\Lambda+1}{4} \text{ and } H_{1122} = -\frac{Q}{E_0}, \quad B_3 \to -\frac{\Lambda+1}{4} \text{ and } H_{2211} = -\frac{Q}{E_0} \tag{48}$$

$$B_5 \to \Lambda+1 \text{ and } 2H_{1212} = 4\frac{Q}{E_0} \quad .$$

Thus, in the case of circular pore, the effective continuum is isotropic, while in the case of oriented elliptical pore it is anisotropic (it is neither isotropic nor orthotropic, since $B_2 \neq B_3$). However, if we consider elliptical pore randomly oriented, we obtain the compliance matrix by angular averaging of the tensor of compliances for elliptical pores with sloped major axis with respect to the applied loading. Making use the transformation properties of tensors, $\left\langle M^{eff}_{ijkl}\right\rangle_\psi := \frac{1}{2\pi}\int_0^{2\pi} \tilde{M}^{eff}_{ijkl} d\psi$, $\tilde{M}^{eff}_{ijkl} = \mathrm{O}^T_{i\alpha}(\psi)\mathrm{O}^T_{j\beta}(\psi)\mathrm{O}^T_{k\gamma}(\psi)\mathrm{O}^T_{l\delta}(\psi)M^{eff}_{ijkl}$

one can perform the angular averaging in closed form

$$\left\langle M^{eff}_{1111}\right\rangle_\psi = \frac{3}{8}\left(M^{eff}_{1111} + M^{eff}_{2222}\right) + \frac{1}{8}\left(M^{eff}_{1122} + M^{eff}_{2211} + 4M^{eff}_{1212}\right) = \left\langle M^{eff}_{2222}\right\rangle_\psi$$

$$\left\langle M^{eff}_{1122}\right\rangle_\psi = \frac{1}{8}\left(M^{eff}_{1111} + M^{eff}_{2222} - 4M^{eff}_{1212}\right) + \frac{3}{8}\left(M^{eff}_{1122} + M^{eff}_{2211}\right) = \left\langle M^{eff}_{2211}\right\rangle_\psi \tag{49}$$

$$\left\langle M^{eff}_{1112}\right\rangle_\psi = 0 = \left\langle M^{eff}_{2212}\right\rangle_\psi = \left\langle M^{eff}_{1211}\right\rangle_\psi = \left\langle M^{eff}_{1222}\right\rangle_\psi$$

$$\left\langle M^{eff}_{1212}\right\rangle_\psi = \frac{1}{8}\left(M^{eff}_{1111} + M^{eff}_{2222} - M^{eff}_{1122} - M^{eff}_{2211}\right) + \frac{1}{2}M^{eff}_{1212} \quad .$$

Thus, the effective continuum in the case of randomly oriented elliptical pore is isotropic.

Having known the matrices M_{ijkl}, M^s_{ijkl}, H_{ijkl} one can find the matrices of compliances for effective continuum in particular approaches AA(I), AA(II), AA(III) and AA(IV). Recall that the matrix M_{ijkl} is known in closed form only for circular pore, but not for elliptical pore. Thus, for the elliptical pore, we have closed form results only by AA(I) and AA(III). The ma-

trix of stiffness coefficients for effective continuum is obtained by inversion of the matrix of compliances.

2.2 Numerical approach

If the shape of pore is arbitrary, one can solve the micro-structural bvp in the RVE numerically and get the micro-structural mean values of strains and stresses. If we consider the boundary of the pore to be traction-free $t_i\big|_{\Gamma_0}=0$, we can express the mean values in terms of boundary integrals as

$$\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega} := \frac{1}{|\Omega|}\int_{\Omega}\sigma_{ij}^{ms}(\mathbf{x})d\Omega = \frac{1}{2|\Omega|}\int_{\Omega}\left(\sigma_{ik}^{ms}x_{j,k}+\sigma_{jk}^{ms}x_{i,k}\right)d\Omega =$$

$$= \frac{1}{2|\Omega|}\int_{\Omega}\left[\left(\sigma_{ik}^{ms}x_j\right)_{,k}+\left(\sigma_{jk}^{ms}x_i\right)_{,k}\right]d\Omega = \frac{1}{2|\Omega|}\int_{\Gamma\cup\Gamma_0}\left(t_i^{ms}x_j+t_j^{ms}x_i\right)d\Gamma =$$

$$= \frac{1}{2|\Omega|}\int_{\Gamma}\left(t_i^{ms}x_j+t_j^{ms}x_i\right)d\Gamma \tag{50}$$

$$\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega} := \frac{1}{|\Omega|}\int_{\Omega}\varepsilon_{ij}^{ms}(\mathbf{x})d\Omega = \frac{1}{2|\Omega|}\int_{\Omega}\left(u_{i,j}^{ms}+u_{j,i}^{ms}\right)d\Omega = \frac{1}{2|\Omega|}\int_{\Gamma\cup\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma =$$

$$= \left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma}+\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma_0} \tag{51}$$

with $\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma} := \frac{1}{2|\Omega|}\int_{\Gamma}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma$,

$$\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma_0} := \frac{1}{2|\Omega|}\int_{\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma = \frac{|B|}{|\Omega|}\Delta\varepsilon_{ij}^{ms},\quad \Delta\varepsilon_{ij}^{ms} := \frac{1}{2|B|}\int_{\Gamma_0}\left(u_i^{ms}n_j+u_j^{ms}n_i\right)d\Gamma$$

Since the mean values of the strains and stresses in effective continuum are not known (we cannot solve any direct bvp in effective medium because of absence of material coefficients), we should find a physically meaningful correlation between $\left[\left\langle \sigma_{ij}^{eff}\right\rangle_B,\left\langle \varepsilon_{kl}^{eff}\right\rangle_B\right]$ and $\left[\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega},\left\langle \varepsilon_{kl}^{ms}\right\rangle_{\Omega}\right]$ in order to get the stiffness coefficients in the homogenized effective continuum

$$\begin{pmatrix}\sigma_{11}^{eff}\\ \sigma_{22}^{eff}\\ \sigma_{12}^{eff}\end{pmatrix}=\begin{pmatrix}c_{11}^{eff} & c_{12}^{eff} & c_{16}^{eff}\\ c_{21}^{eff} & c_{22}^{eff} & c_{26}^{eff}\\ c_{61}^{eff} & c_{62}^{eff} & c_{66}^{eff}\end{pmatrix}\begin{pmatrix}\varepsilon_{11}^{eff}\\ \varepsilon_{22}^{eff}\\ 2\varepsilon_{12}^{eff}\end{pmatrix}\quad\text{or}\quad\begin{pmatrix}\left\langle \sigma_{11}^{eff}\right\rangle_B\\ \left\langle \sigma_{22}^{eff}\right\rangle_B\\ \left\langle \sigma_{12}^{eff}\right\rangle_B\end{pmatrix}=\begin{pmatrix}c_{11}^{eff} & c_{12}^{eff} & c_{16}^{eff}\\ c_{21}^{eff} & c_{22}^{eff} & c_{26}^{eff}\\ c_{61}^{eff} & c_{62}^{eff} & c_{66}^{eff}\end{pmatrix}\begin{pmatrix}\left\langle \varepsilon_{11}^{eff}\right\rangle_B\\ \left\langle \varepsilon_{22}^{eff}\right\rangle_B\\ 2\left\langle \varepsilon_{12}^{eff}\right\rangle_B\end{pmatrix}. \tag{52}$$

For this purpose, we shall use two approaches:

(i) *boundary densities compatibility* (BDC) approach: we assume that the physically relevant boundary densities on the outer boundary of the RVE are the same for both the micro-structural bvp and the bvp in effective continuum, i.e. $u_i^{eff}\big|_{\Gamma}=u_i^{ms}\big|_{\Gamma}$, $t_i^{eff}\big|_{\Gamma}=t_i^{ms}\big|_{\Gamma}$. Then,

$$\left\langle \varepsilon_{ij}^{eff}\right\rangle_B := \frac{1}{|B|}\int_B\varepsilon_{ij}^{eff}(\mathbf{x})d\Omega = \frac{1}{2|B|}\int_B\left(u_{i,j}^{eff}+u_{j,i}^{eff}\right)d\Omega = \frac{1}{2|B|}\int_{\Gamma}\left(u_i^{eff}n_j+u_j^{eff}n_i\right)d\Gamma =$$

$$= \frac{1}{2|B|}\int_{\Gamma}\left(u_i^{ms} n_j + u_j^{ms} n_i\right) d\Omega = (1-p)\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega,\Gamma} \quad , \tag{53}$$

$$\left\langle \sigma_{ij}^{eff}\right\rangle_B := \frac{1}{|B|}\int_B \sigma_{ij}^{eff}(\mathbf{x}) d\Omega = \frac{1}{2|B|}\int_B \left(\sigma_{ik}^{eff} x_{j,k} + \sigma_{jk}^{eff} x_{i,k}\right) d\Omega =$$

$$= \frac{1}{2|B|}\int_B \left[\left(\sigma_{ik}^{eff} x_j\right)_{,k} + \left(\sigma_{jk}^{eff} x_i\right)_{,k}\right] d\Omega = \frac{1}{2|B|}\int_{\Gamma}\left(t_i^{eff} x_j + t_j^{eff} x_i\right) d\Gamma =$$

$$= \frac{1}{2|B|}\int_{\Gamma}\left(t_i^{ms} x_j + t_j^{ms} x_i\right) d\Gamma = (1-p)\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega} . \tag{54}$$

(ii) *integral force & energy compatibilities* (IF&EC) approach: we consider two assumptions:

$$\left.\begin{array}{l} (1)\ \text{integral force equilibrium: } |B|\left\langle \sigma_{ij}^{eff}\right\rangle_B = |\Omega|\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega} \\ (2)\ \text{energy equilibrium: } \left\langle \sigma_{ij}^{eff}\varepsilon_{ij}^{eff}\right\rangle_B = \left\langle \sigma_{ij}^{ms}\varepsilon_{ij}^{ms}\right\rangle_{\Omega} \\ \qquad\qquad\qquad \left\langle \sigma_{ij}^{eff}\varepsilon_{ij}^{eff}\right\rangle_B = \left\langle \sigma_{ij}^{eff}\right\rangle_B \left\langle \varepsilon_{ij}^{eff}\right\rangle_B \end{array}\right\} \Rightarrow$$

$$\Rightarrow \quad \left\{\begin{array}{l} \left\langle \sigma_{ij}^{eff}\right\rangle_B = (1-p)\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega} \\ \left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega}\left\langle \varepsilon_{ij}^{eff}\right\rangle_B = \dfrac{1}{1-p}\left\langle \sigma_{ij}^{ms}\varepsilon_{ij}^{ms}\right\rangle_{\Omega} \end{array}\right.$$

Since

$$\left\langle \sigma_{ij}^{ms}\varepsilon_{ij}^{ms}\right\rangle_{\Omega} = \frac{1}{|\Omega|}\int_{\partial\Omega} t_i^{ms} u_i^{ms} d\Gamma = \frac{1}{|\Omega|}\int_{\Gamma} t_i^{ms} u_i^{ms} d\Gamma$$

$$|\Omega|\left(\left\langle \sigma_{ij}^{ms}\varepsilon_{ij}^{ms}\right\rangle_{\Omega} - \left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega}\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega}\right) = \int_{\Gamma} t_i^{ms} u_i^{ms} d\Gamma - \left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega}\int_{\Gamma} t_i^{ms} x_j d\Gamma -$$

$$-|\Omega|\left\langle \sigma_{ik}^{ms}\right\rangle_{\Omega}\left\langle \varepsilon_{ik}^{ms}\right\rangle_{\Omega} + \left\langle \sigma_{ik}^{ms}\right\rangle_{\Omega}\left\langle \varepsilon_{ik}^{ms}\right\rangle_{\Omega}\int_{\Gamma+\Gamma_0} x_j n_k d\Gamma =$$

$$= \int_{\Gamma}\left(t_i^{ms} - n_k\left\langle \sigma_{ik}^{ms}\right\rangle_{\Omega}\right)\left(u_i^{ms} - x_j\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega}\right) d\Gamma +$$

$$+\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega}\left\langle \varepsilon_{ik}^{ms}\right\rangle_{\Omega}\underbrace{\int_{\Gamma_0} x_j n_k d\Gamma}_{-(|B|-|\Omega|)\delta_{jk}} - \left\langle \sigma_{ik}^{ms}\right\rangle_{\Omega}\underbrace{|\Omega|\left\langle \varepsilon_{ik}^{ms}\right\rangle_{\Omega,\Gamma_0}}_{|B|\Delta\varepsilon_{ik}^{ms}}$$

we have

$$\left\langle \sigma_{ij}^{ms}\varepsilon_{ij}^{ms}\right\rangle_{\Omega} = \frac{1}{1-p}\left\langle \sigma_{ij}^{ms}\right\rangle_{\Omega}\left((1-2p)\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega} - \Delta\varepsilon_{ij}^{ms}\right) + \int_{\Gamma}\left(t_i^{ms} - n_k\left\langle \sigma_{ik}^{ms}\right\rangle_{\Omega}\right)\left(u_i^{ms} - x_j\left\langle \varepsilon_{ij}^{ms}\right\rangle_{\Omega}\right) d\Gamma$$

The last integral is vanishing if the KUBC or SUBC are employed on the outer boundary of the RVE. Then,

$$\left\langle \varepsilon_{ij}^{eff} \right\rangle_B = \frac{1}{(1-p)^2}\left((1-2p)\left\langle \varepsilon_{ij}^{ms} \right\rangle_\Omega - \Delta\varepsilon_{ij}^{ms}\right) \quad . \tag{55}$$

The evaluation of mean values of micro-structural fields is remarkably simplified, if we consider KUBC (kinematic uniform b.c.) on the outer boundary of the RVE, i.e.

$$u_i^{ms}\big|_\Gamma = \tilde{u}_i \text{ with } \tilde{u}_i := \bar{\varepsilon}_{ik} x_k\big|_\Gamma \ , \ \ \bar{\varepsilon}_{ik} = const \, . \tag{56}$$

Then,

$$\left\langle \varepsilon_{ij}^{ms} \right\rangle_{\Omega,\Gamma} := \frac{1}{2|\Omega|}\int_\Gamma \left(u_i^{ms} n_j + u_j^{ms} n_i\right) d\Gamma = \frac{1}{2|\Omega|}\int_\Gamma \left(\bar{\varepsilon}_{ik} x_k n_j + \bar{\varepsilon}_{jk} x_k n_i\right) d\Omega =$$

$$= \frac{1}{2|\Omega|}\int_B \left(\bar{\varepsilon}_{ik} x_{k,j} + \bar{\varepsilon}_{jk} x_{k,i}\right) d\Omega = \frac{|B|}{|\Omega|}\bar{\varepsilon}_{ij} = \frac{1}{1-p}\bar{\varepsilon}_{ij} \tag{57}$$

Moreover, in the case of uniform distribution of pores in the macrostructure, we can create identical rectangular cells around each pore with periodic boundary conditions on the boundaries of cells. Then, if we create the RVE from several cells, we can see that because of the symmetry and consideration of rigid body motion equivalence, the solution in each cell is the same. Therefore it is sufficient to solve the micro-structural bvp in one cell with one pore in the case of uniform distribution of pores in the macrostructure. In what follows, we restrict our consideration to KUBC.

Thus, we can find the mean values of strains and stresses in the effective continuum as

(i) ***BDC-approach***:

$$\left\langle \varepsilon_{ij}^{eff} \right\rangle_B = (1-p)\left\langle \varepsilon_{ij}^{ms} \right\rangle_{\Omega,\Gamma} = \bar{\varepsilon}_{ij} \tag{58}$$

$$\left\langle \sigma_{ij}^{eff} \right\rangle_B = (1-p)\left\langle \sigma_{ij}^{ms} \right\rangle_\Omega = \frac{1}{2|B|}\int_\Gamma \left(t_i^{ms} x_j + t_j^{ms} x_i\right) d\Gamma \tag{59}$$

where t_i^{ms} on Γ should be got from the solution of the micro-structural bvp.

(ii) ***IF&EC-approach***:

$$\left\langle \varepsilon_{ij}^{eff} \right\rangle_B = \frac{1}{(1-p)^2}\left[(1-2p)\left\langle \varepsilon_{ij}^{ms} \right\rangle_\Omega - \Delta\varepsilon_{ij}^{ms}\right] =$$

$$= \frac{1}{(1-p)^2}\left[(1-2p)\left(\left\langle \varepsilon_{ij}^{ms} \right\rangle_{\Omega,\Gamma} + \frac{1}{1-p}\Delta\varepsilon_{ij}^{ms}\right) - \Delta\varepsilon_{ij}^{ms}\right] =$$

$$= \frac{1}{(1-p)^3}\left[(1-2p)\bar{\varepsilon}_{ij} - p\Delta\varepsilon_{ij}^{ms}\right], \qquad \Delta\varepsilon_{ij}^{ms} = \frac{1}{2|B|}\int_{\Gamma_0}\left(u_i^{ms} n_j + u_j^{ms} n_i\right) d\Gamma \tag{60}$$

$$\left\langle \sigma_{ij}^{eff} \right\rangle_B = (1-p)\left\langle \sigma_{ij}^{ms} \right\rangle_\Omega = \frac{1}{2|B|}\int_\Gamma \left(t_i^{ms} x_j + t_j^{ms} x_i\right) d\Gamma \tag{61}$$

where the integrands in both the integrals over Γ_0 and Γ are taken from the solution of the micro-structural bvp.

Now, we show that it is sufficient to solve three micro-structural bvp in order to get all material coefficients in homogenized effective elastic continuum:

b.v.p. (I): $u_i^I(\mathbf{x}) = \bar{\varepsilon}_{ik}^I x_k\big|_\Gamma \;,\quad \bar{\varepsilon}_{ik}^I = \bar{\varepsilon}_{11}^I \delta_{i1}\delta_{k1} \;;\quad t_i^I(\mathbf{x})\big|_{\Gamma_0} = 0$ (62)

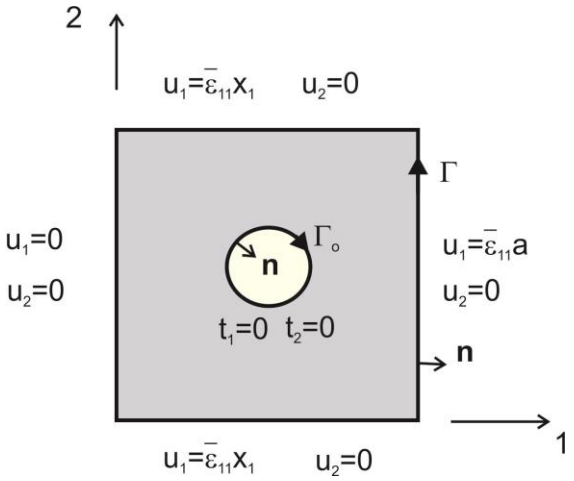


Fig. 3 The first choice of boundary conditions in quadrilateral RVE with one empty pore

Having solved the considered *ms-bvp*, one can evaluate

$$\left\langle \sigma_{ij}^{(I)eff} \right\rangle_B = \frac{1}{2l^2}\int_\Gamma \left(t_i^{(I)ms} x_j + t_j^{(I)ms} x_i \right) d\Gamma \;,\qquad \Delta\varepsilon_{ij}^{(I)ms} = \frac{1}{2l^2}\int_{\Gamma_0} \left(u_i^{(I)ms} n_j + u_j^{(I)ms} n_i \right) d\Gamma \tag{63}$$

b.v.p. (II): $\tilde{u}_i^{II}(\mathbf{x}) = \bar{\varepsilon}_{ik}^{II} x_k\big|_\Gamma \;,\quad \bar{\varepsilon}_{ik}^{II} = \bar{\varepsilon}_{22}^{II}\delta_{i2}\delta_{k2} \;;\quad t_i^{II}(\mathbf{x})\big|_{\Gamma_0} = 0$ (64)

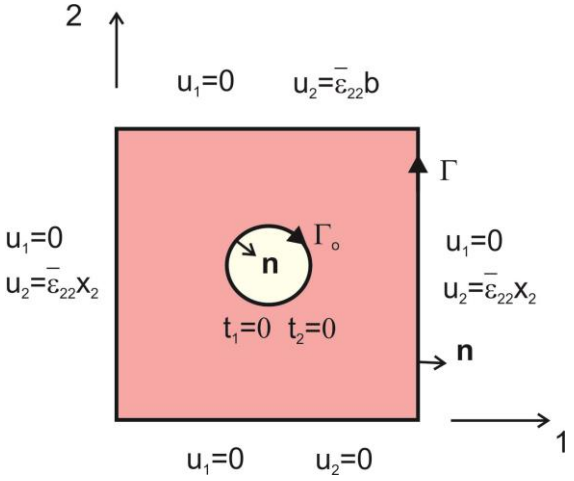


Fig. 4 The second choice of boundary conditions in quadrilateral RVE with one empty pore

Having solved the considered *ms-bvp*, one can evaluate

$$\left\langle \sigma_{ij}^{(II)eff} \right\rangle_B = \frac{1}{2l^2}\int_\Gamma \left(t_i^{(II)ms} x_j + t_j^{(II)ms} x_i \right) d\Gamma \;,\qquad \Delta\varepsilon_{ij}^{(II)ms} = \frac{1}{2l^2}\int_{\Gamma_0} \left(u_i^{(II)ms} n_j + u_j^{(II)ms} n_i \right) d\Gamma \tag{65}$$

b.v.p. (III): $\tilde{u}_i^{III}(\mathbf{x}) = \bar{\varepsilon}_{ik}^{III} x_k\big|_\Gamma \;,\quad \bar{\varepsilon}_{ik}^{III} = \bar{\varepsilon}_{12}^{III}\left(\delta_{i1}\delta_{k2} + \delta_{i2}\delta_{k1}\right);\quad t_i^{III}(\mathbf{x})\big|_{\Gamma_0} = 0$ (66)

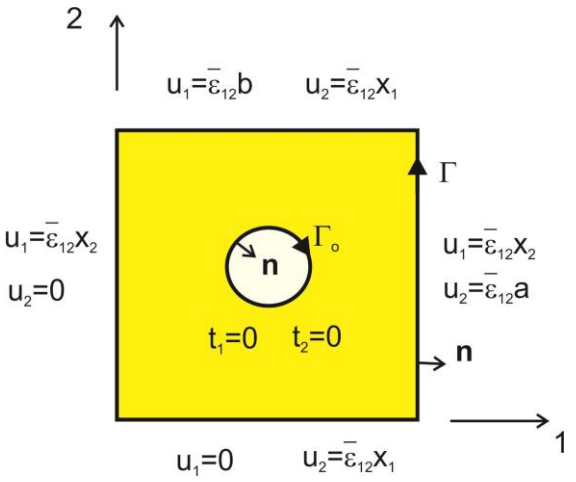


Fig. 5 The third choice of boundary conditions in quadrilateral RVE with one empty pore

Having solved the considered *ms-bvp*, one can evaluate

$$\left\langle \sigma_{ij}^{(III)eff} \right\rangle_B = \frac{1}{2l^2}\int_{\Gamma}\left(t_i^{(III)ms}x_j + t_j^{(III)ms}x_i\right)d\Gamma\ , \qquad \Delta\varepsilon_{ij}^{(III)ms} = \frac{1}{2l^2}\int_{\Gamma_0}\left(u_i^{(III)ms}n_j + u_j^{(III)ms}n_i\right)d\Gamma \tag{67}$$

According to Eqs. (52), (58) and (59) we obtain the effective stiffness coefficients by the ***BDC-approach*** as:

$$c_{11}^{eff} = \frac{\left\langle \sigma_{11}^{(I)eff} \right\rangle_B}{\bar{\varepsilon}_{11}^{I}}\ , \quad c_{21}^{eff} = \frac{\left\langle \sigma_{22}^{(I)eff} \right\rangle_B}{\bar{\varepsilon}_{11}^{I}}\ , \quad c_{61}^{eff} = \frac{\left\langle \sigma_{12}^{(I)eff} \right\rangle_B}{\bar{\varepsilon}_{11}^{I}} \tag{68}$$

$$c_{12}^{eff} = \frac{\left\langle \sigma_{11}^{(II)eff} \right\rangle_B}{\bar{\varepsilon}_{22}^{II}}\ , \quad c_{22}^{eff} = \frac{\left\langle \sigma_{22}^{(II)eff} \right\rangle_B}{\bar{\varepsilon}_{22}^{II}}\ , \quad c_{62}^{eff} = \frac{\left\langle \sigma_{12}^{(II)eff} \right\rangle_B}{\bar{\varepsilon}_{22}^{II}} \tag{69}$$

$$c_{16}^{eff} = \frac{\left\langle \sigma_{11}^{(III)eff} \right\rangle_B}{2\bar{\varepsilon}_{12}^{III}}\ , \quad c_{26}^{eff} = \frac{\left\langle \sigma_{22}^{(III)eff} \right\rangle_B}{2\bar{\varepsilon}_{12}^{III}}\ , \quad c_{66}^{eff} = \frac{\left\langle \sigma_{12}^{(III)eff} \right\rangle_B}{2\bar{\varepsilon}_{12}^{III}}\ . \tag{70}$$

In the case of ***IF&EC-approach*** we obtain the following systems of algebraic equations:

b.v.p. (I):

$$\left[(1-2p)\bar{\varepsilon}_{11} - p\Delta\varepsilon_{11}^{ms}\right]c_{11}^{eff} - p\Delta\varepsilon_{22}^{(I)ms}c_{12}^{eff} - 2p\Delta\varepsilon_{12}^{(I)ms}c_{16}^{eff} = (1-p)^3\left\langle \sigma_{11}^{(I)eff} \right\rangle_B \tag{71a}$$

$$\left[(1-2p)\bar{\varepsilon}_{11} - p\Delta\varepsilon_{11}^{ms}\right]c_{21}^{eff} - p\Delta\varepsilon_{22}^{(I)ms}c_{22}^{eff} - 2p\Delta\varepsilon_{12}^{(I)ms}c_{26}^{eff} = (1-p)^3\left\langle \sigma_{22}^{(I)eff} \right\rangle_B \tag{71b}$$

$$\left[(1-2p)\bar{\varepsilon}_{11} - p\Delta\varepsilon_{11}^{ms}\right]c_{61}^{eff} - p\Delta\varepsilon_{22}^{(I)ms}c_{62}^{eff} - 2p\Delta\varepsilon_{12}^{(I)ms}c_{66}^{eff} = (1-p)^3\left\langle \sigma_{12}^{(I)eff} \right\rangle_B \tag{71c}$$

b.v.p. (II):

$$-p\Delta\varepsilon_{11}^{(II)ms}c_{11}^{eff} + \left[(1-2p)\bar{\varepsilon}_{22} - p\Delta\varepsilon_{22}^{(II)ms}\right]c_{12}^{eff} - 2p\Delta\varepsilon_{12}^{(II)ms}c_{16}^{eff} = (1-p)^3\left\langle \sigma_{11}^{(II)eff} \right\rangle_B \tag{72a}$$

$$-p\Delta\varepsilon_{11}^{(II)ms}c_{21}^{eff} + \left[(1-2p)\bar{\varepsilon}_{22} - p\Delta\varepsilon_{22}^{(II)ms}\right]c_{22}^{eff} - 2p\Delta\varepsilon_{12}^{(II)ms}c_{26}^{eff} = (1-p)^3\left\langle \sigma_{22}^{(II)eff} \right\rangle_B \tag{72b}$$

$$-p\Delta\varepsilon_{11}^{(II)ms}c_{61}^{eff}+\left[(1-2p)\bar{\varepsilon}_{22}-p\Delta\varepsilon_{22}^{(II)ms}\right]c_{62}^{eff}-2p\Delta\varepsilon_{12}^{(II)ms}c_{66}^{eff}=(1-p)^3\left\langle\sigma_{12}^{(II)eff}\right\rangle_B \tag{72c}$$

b.v.p. (III):

$$-p\Delta\varepsilon_{11}^{(III)ms}c_{11}^{eff}-p\Delta\varepsilon_{22}^{(III)ms}c_{12}^{eff}+2\left[(1-2p)\bar{\varepsilon}_{12}-p\Delta\varepsilon_{12}^{(III)ms}\right]c_{16}^{eff}=(1-p)^3\left\langle\sigma_{11}^{(III)eff}\right\rangle_B \tag{73a}$$

$$-p\Delta\varepsilon_{11}^{(III)ms}c_{21}^{eff}-p\Delta\varepsilon_{22}^{(III)ms}c_{22}^{eff}+2\left[(1-2p)\bar{\varepsilon}_{12}-p\Delta\varepsilon_{12}^{(III)ms}\right]c_{26}^{eff}=(1-p)^3\left\langle\sigma_{22}^{(III)eff}\right\rangle_B \tag{73b}$$

$$-p\Delta\varepsilon_{11}^{(III)ms}c_{61}^{eff}-p\Delta\varepsilon_{22}^{(III)ms}c_{62}^{eff}+2\left[(1-2p)\bar{\varepsilon}_{12}-p\Delta\varepsilon_{12}^{(III)ms}\right]c_{66}^{eff}=(1-p)^3\left\langle\sigma_{12}^{(III)eff}\right\rangle_B \tag{73c}$$

Gathering Eqs. (71a), (72a) and (73a), we obtain the following system of equations for unknowns $\left(c_{11}^{eff}\quad c_{12}^{eff}\quad c_{16}^{eff}\right)^T$:

$$[A]\begin{pmatrix}c_{11}^{eff}\\c_{12}^{eff}\\c_{16}^{eff}\end{pmatrix}=(1-p)^3\begin{pmatrix}\left\langle\sigma_{11}^{(I)eff}\right\rangle_B\\\left\langle\sigma_{11}^{(II)eff}\right\rangle_B\\\left\langle\sigma_{11}^{(III)eff}\right\rangle_B\end{pmatrix} \tag{74}$$

$$\text{with } [A]=\begin{pmatrix}\left[(1-2p)\bar{\varepsilon}_{11}-p\Delta\varepsilon_{11}^{ms}\right] & -p\Delta\varepsilon_{22}^{(I)ms} & -2p\Delta\varepsilon_{12}^{(I)ms}\\ -p\Delta\varepsilon_{11}^{(II)ms} & \left[(1-2p)\bar{\varepsilon}_{22}-p\Delta\varepsilon_{22}^{(II)ms}\right] & -2p\Delta\varepsilon_{12}^{(II)ms}\\ -p\Delta\varepsilon_{11}^{(III)ms} & -p\Delta\varepsilon_{22}^{(III)ms} & 2\left[(1-2p)\bar{\varepsilon}_{12}-p\Delta\varepsilon_{12}^{(III)ms}\right]\end{pmatrix} \tag{75}$$

Similarly, from Eqs. (71b), (72b) and (73b), we obtain the system of algebraic equations for unknowns $\left(c_{21}^{eff}\quad c_{22}^{eff}\quad c_{26}^{eff}\right)^T$:

$$[A]\begin{pmatrix}c_{21}^{eff}\\c_{22}^{eff}\\c_{26}^{eff}\end{pmatrix}=(1-p)^3\begin{pmatrix}\left\langle\sigma_{22}^{(I)eff}\right\rangle_B\\\left\langle\sigma_{22}^{(II)eff}\right\rangle_B\\\left\langle\sigma_{22}^{(III)eff}\right\rangle_B\end{pmatrix} \tag{76}$$

Finally, from Eqs. (71c), (72c) and (73c), we obtain the system of algebraic equations for unknowns $\left(c_{61}^{eff}\quad c_{62}^{eff}\quad c_{66}^{eff}\right)^T$:

$$[A]\begin{pmatrix}c_{61}^{eff}\\c_{62}^{eff}\\c_{66}^{eff}\end{pmatrix}=(1-p)^3\begin{pmatrix}\left\langle\sigma_{12}^{(I)eff}\right\rangle_B\\\left\langle\sigma_{12}^{(II)eff}\right\rangle_B\\\left\langle\sigma_{12}^{(III)eff}\right\rangle_B\end{pmatrix}. \tag{77}$$

3 NUMERICAL EXPERIMENTS

3.1 Circular pores

In the case of circular pore we have results by 4 analytical and 2 numerical approaches. Comparison of such results for c_{11}^{eff} is shown if Fig.6. One pore is considered in the RVE and porosity is increased by increasing the radius of the pore.

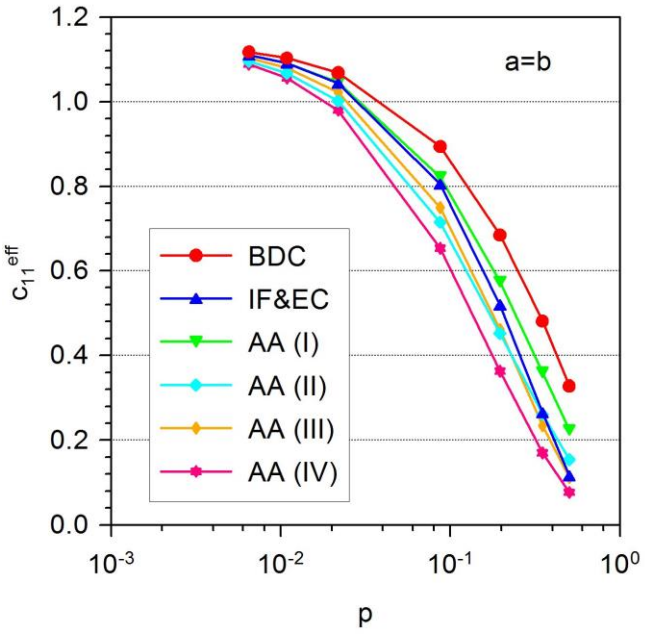
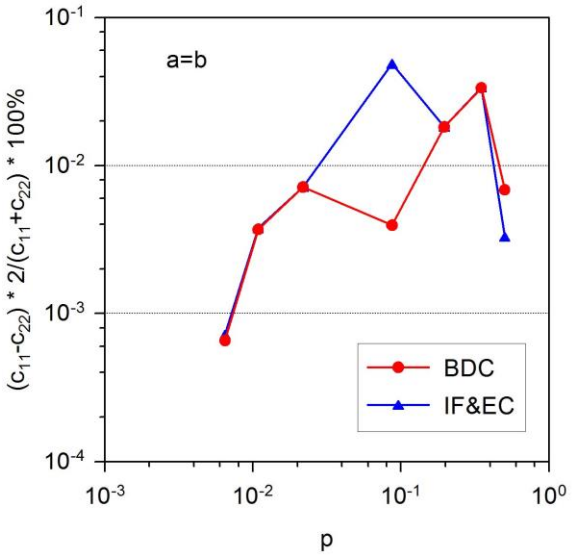
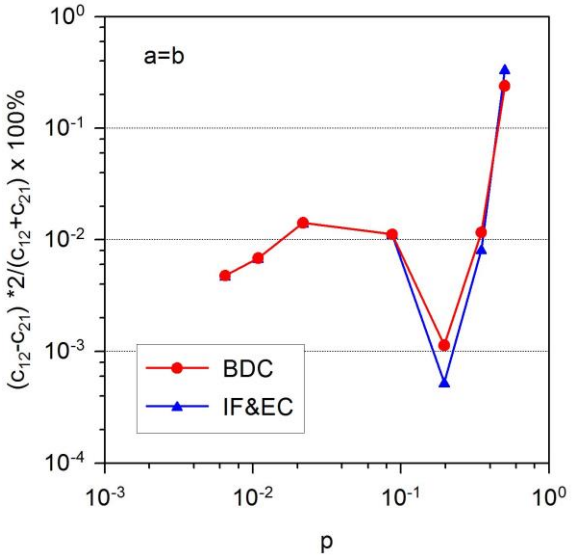


Fig. 6 $c_{11}^{eff}(p)$ by 6 various approaches

The effective material should be isotropic in the case of circular pores. The measure of anisotropy is shown in Fig.7





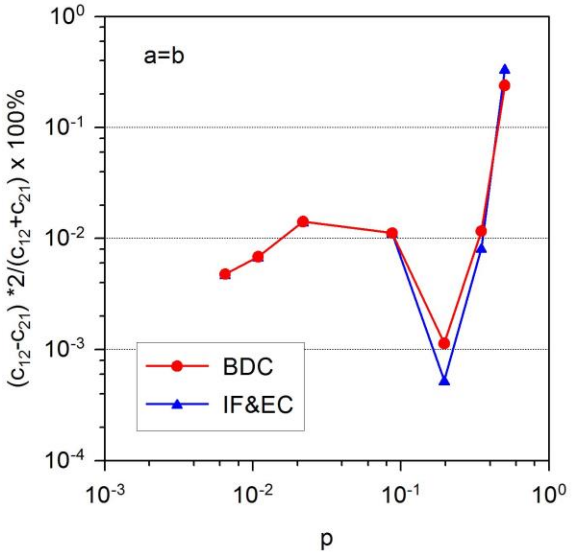


Fig.7 Percentage deviations from isotropy for stiffness coefficients of homogenized continuum with circular pores

Only numerical results are compared in this isotropy study because the analytical approaches exhibit perfect isotropy.

In numerical approaches, we have increased the porosity also by consideration of more pores in the RVE. Comparisons of numerical results obtained by modelling one and more pores are presented in Fig.8.

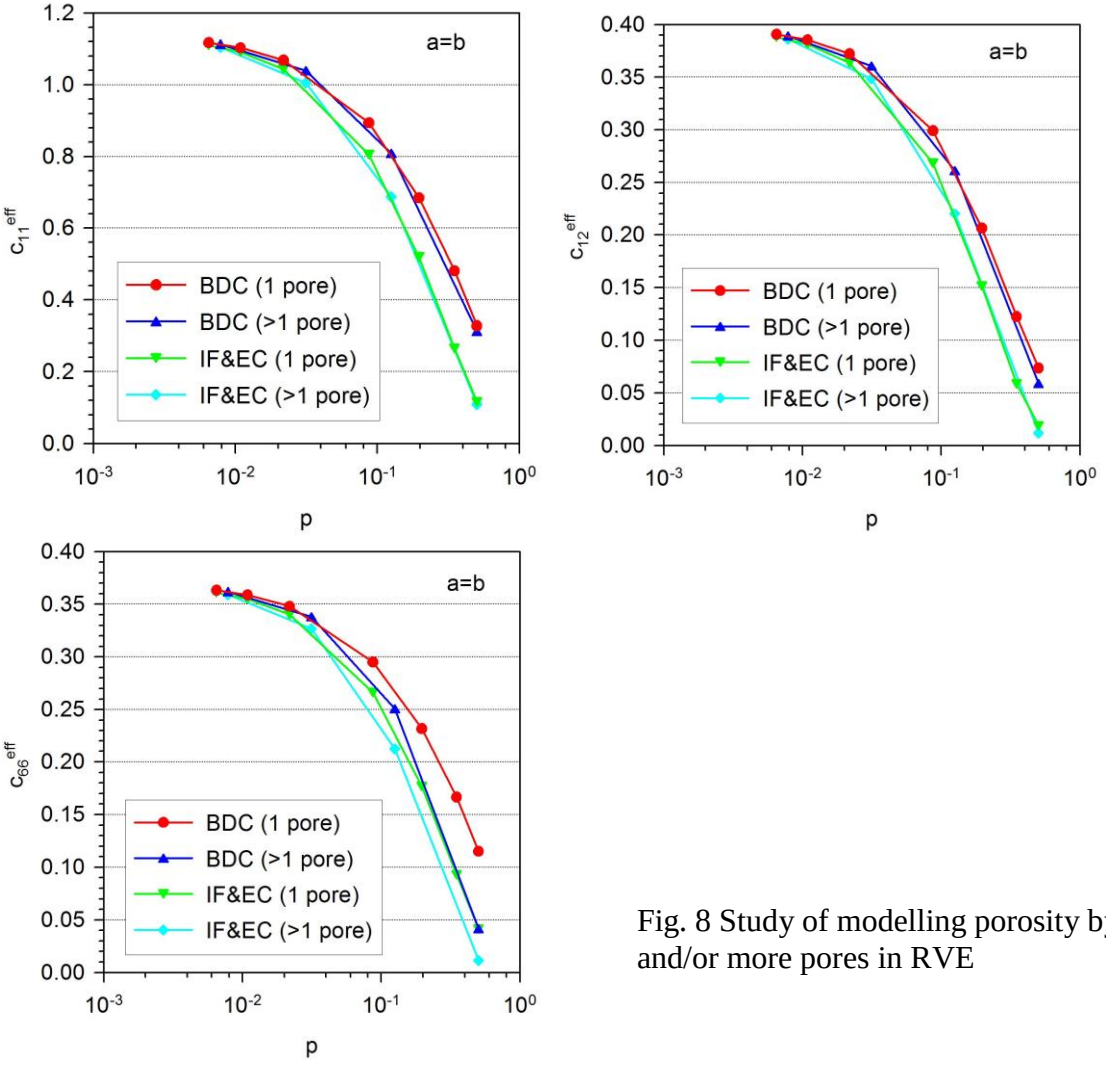


Fig. 8 Study of modelling porosity by one and/or more pores in RVE

The independence of c_{11}^{eff} and c_{12}^{eff} on the kind of modelling the increase of porosity demonstrates the theoretical expectation in the case of uniform distribution of pores. The interaction among the pores is involved into the model via the periodic boundary conditions even if only one pore is considered in the RVE. The deviations for c_{66}^{eff} can be explained by very short distance of pores from the boundary of the RVE for higher values of porosity and rather small value of c_{66}^{eff} .

3.2 Elliptical pores

Now we have numerical results by two approaches and analytical results only by AA(I) approach. Fig.9 shows comparison of results for c_{11}^{eff} coefficient in the case elliptical pores with $a = 2b$

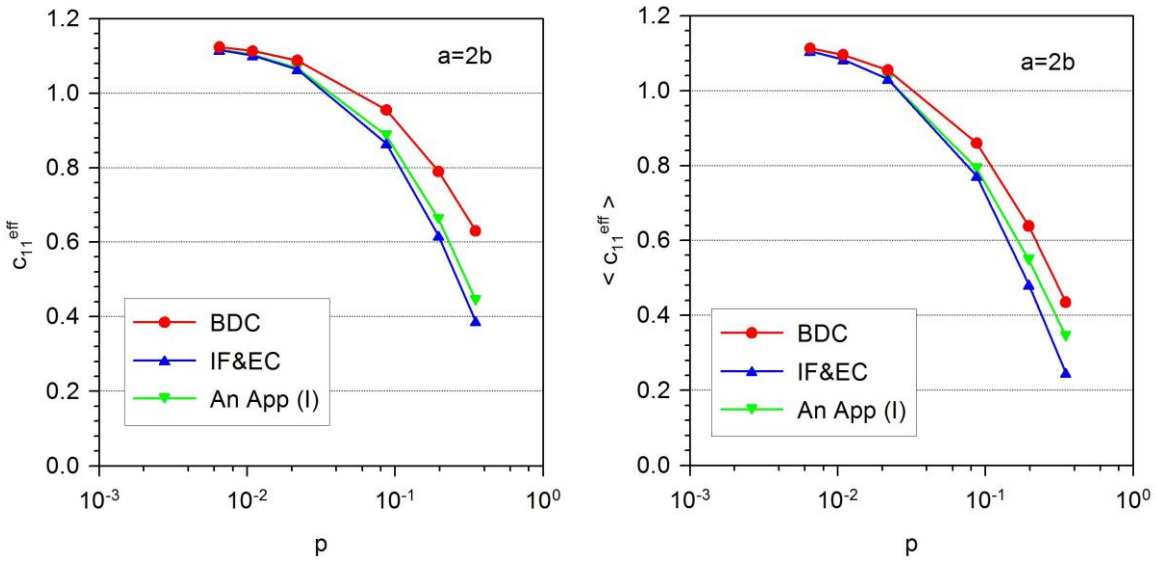
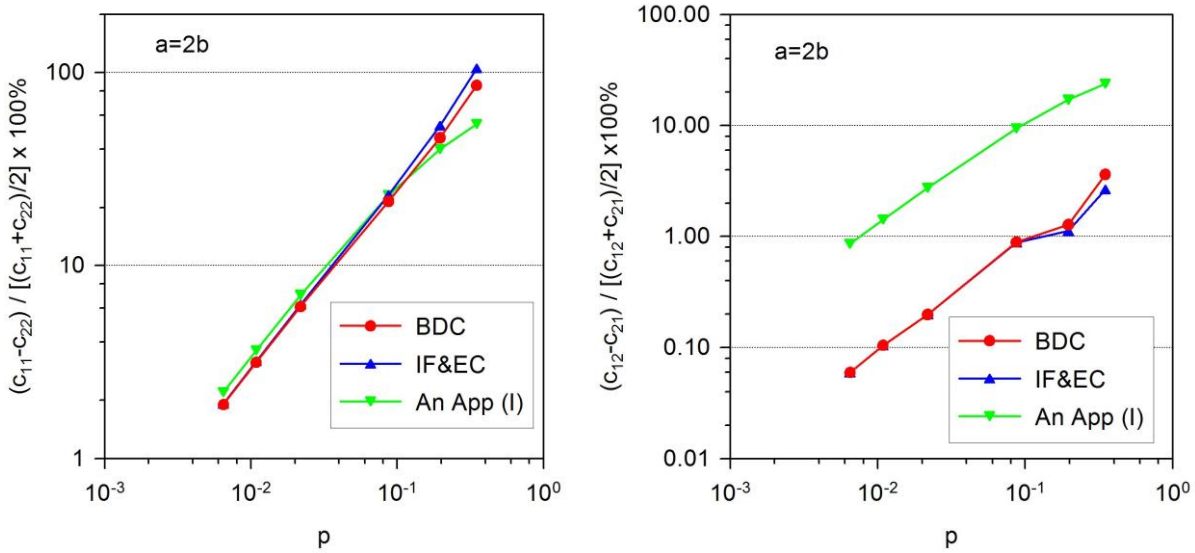


Fig. 9 $c_{11}^{eff}(p)$ for oriented and randomly oriented elliptical pores with $a = 2b$

The anisotropy in this case is verified remarkably for higher values of porosity as can be seen from Fig. 10



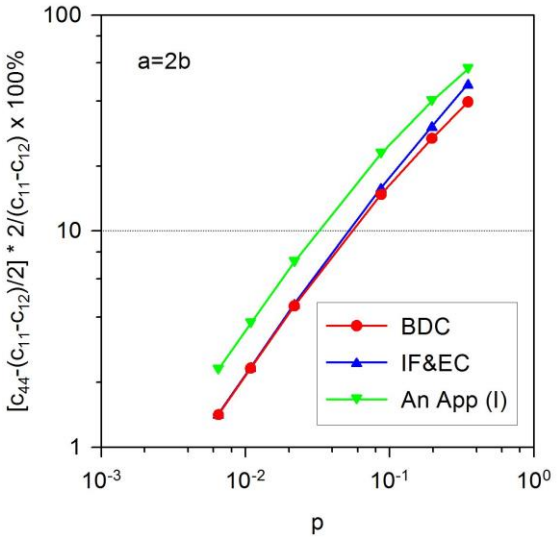


Fig. 10 Demostration of anisotropy in the case of oriented elliptic pores

The disturbance of orthotropic properties ($c_{12}^{eff} \neq c_{21}^{eff}$) is rather weak even for high values of porosity.

In the case of IF&EC approach the periodicity of boundary conditions is not guaranteed for homogenized bvp in contrast to the BDC approach. Therefore we can observe certain deviations between the results by IF&EC approaches when the one and/or more pores are used for modelling porosity (see Fig. 11).

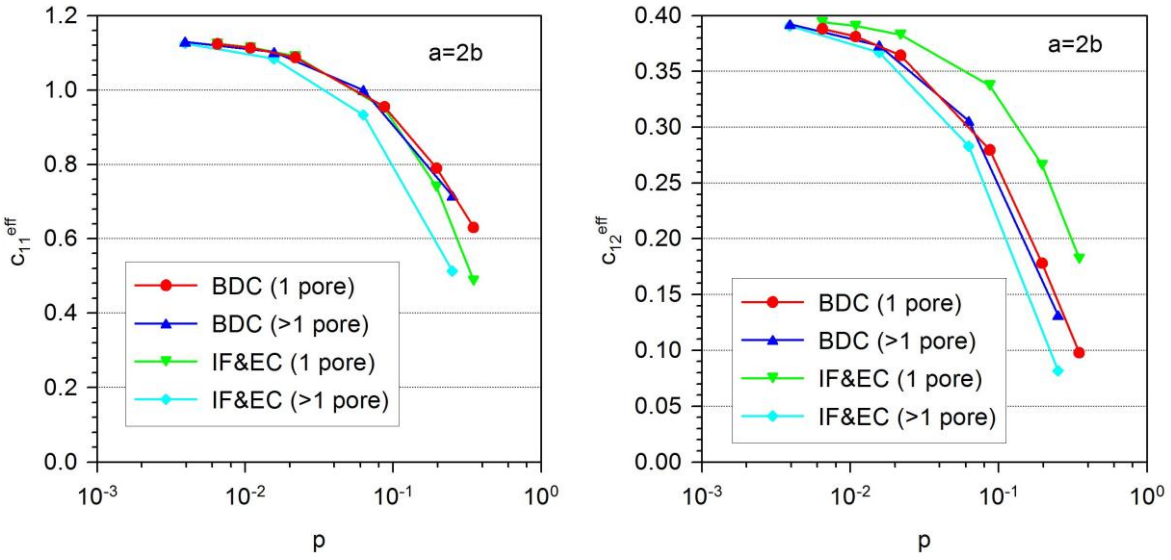


Fig. 11 Demonstration of dependence of results by IF&EC approach on modelling of porosity

It can be seen from Fig. 12 that the solution of the micro-structural boundary value problem with 16 pores in the RVE exhibits the periodic symmetry. This symmetry is confirmed also by the BDC results, but it is not the case for homogenized problem of the IF&EC approach

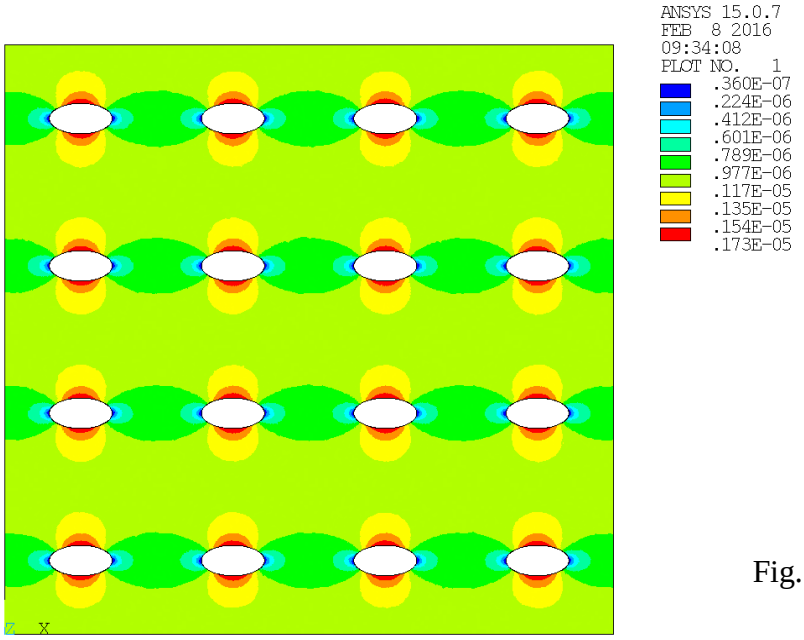


Fig. 12 Distribution of $\sigma_{11}^{(I)ms}$

3.3 Application to plate bending problems

For illustration, we present numerical results for bending of circular plate with a central circular hole subjected to uniform transversal loading, for which the exact solution is also available within the Kirchhoff-Love theory [11]. Fig. 13 shows the dependence of the effective Young modulus and Poisson ratio on the porosity with consideration of two shapes of pores.

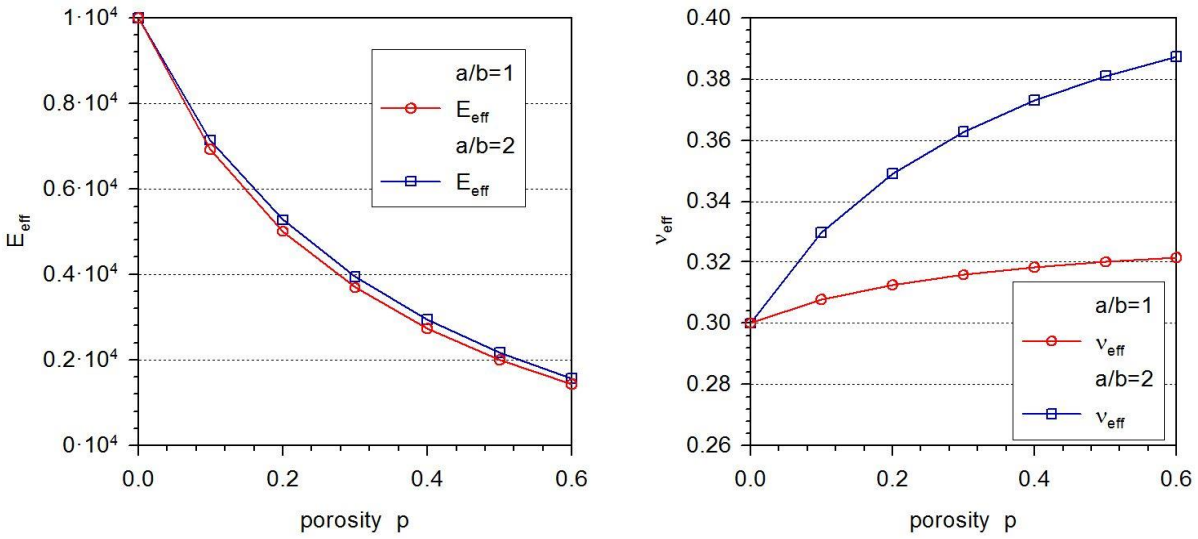


Fig. 13 The effective Young modulus and Poisson ratio vs porosity for circular pores ($a/b=1$) and elliptic pores ($a/b=2$)

From Fig. 14, we can see a strong influence of the porosity on deflections of the plate.

The influence of the shape of pores on deflection can be seen from Fig. 15.

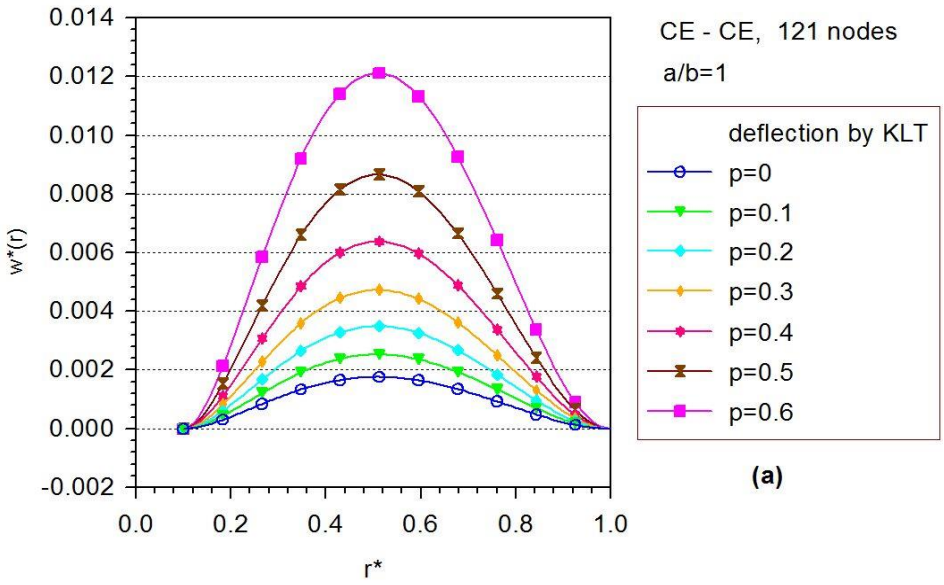


Fig. 14 The radial distribution of deflections in circular plates with various values of porosity

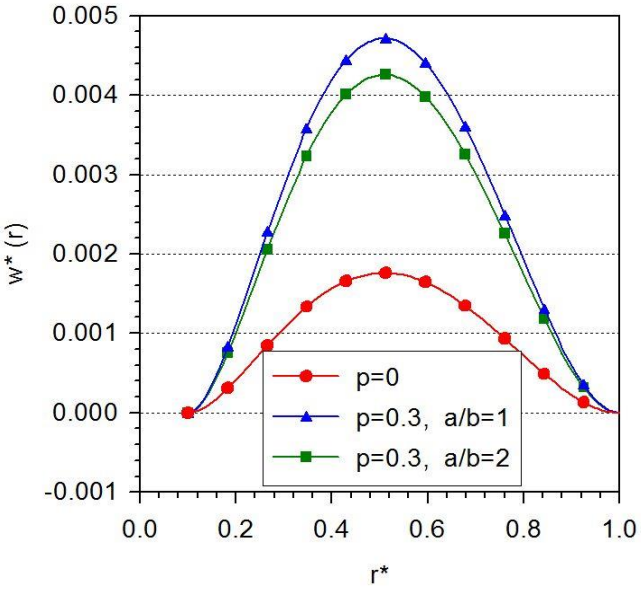


Fig. 15 Comparison of deflections in porous plates with two shapes of pores

It can be seen from Fig. 16 that the dependence of the maximal deflections on the porosity is nonlinear. It can be seen that the dependence is more expressive for thick plates. For thin plate ($r_1 = 50h$) all three theories give almost the same results, for the medium thickness ($r_1 = 10h$) there is coincidence only between the FSDPT (1st order shear deformation plate theory) and the TSDPT (3rd order shear deformation plate theory), and finally for the thick plate ($r_1 = 5h$) a deviation can be seen even between the FSDPT and TSDPT results. The differences between the results by various theories are increasing with increasing the porosity.

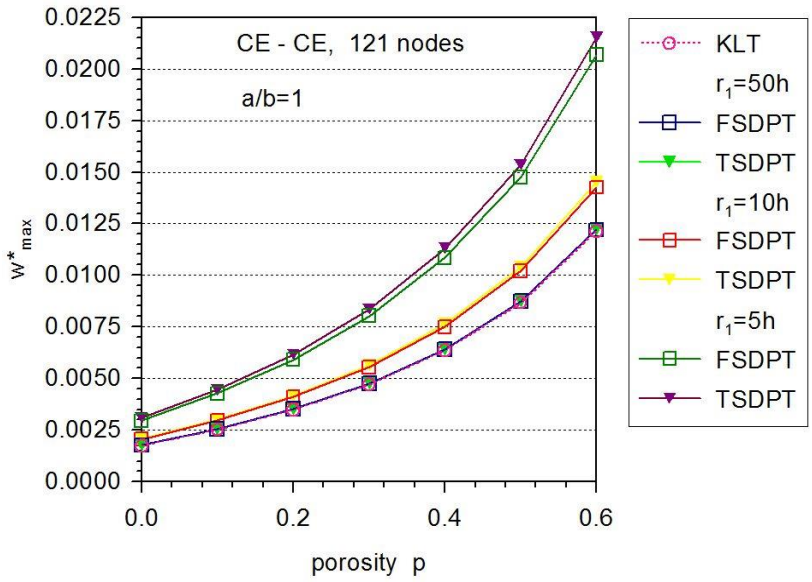


Fig. 16 Dependence of maximal deflections on the porosity

4 CONCLUSIONS

The homogenization of linear elastic media with empty pores is developed on the base of solution of micro-structural boundary value problems. In the case of uniformly distributed simple shape pores, such as circular and elliptic pores, we derived analytical formulae for compliances and/or stiffness coefficients of homogenized effective continuum. For the case of pores of arbitrary shape, we proposed two approaches for evaluation of material coefficients for effective homogenized continuum utilizing the solution of 3 appropriately selected boundary value problems. If the pores are uniformly distributed in the macrostructure, it is sufficient to select the RVE with one pore only. Various aspects of evaluation of effective coefficients are discussed in section devoted to numerical experiments. Finally, the obtained results for homogenization of porous media have been applied to study the influence of porosity on bending of elastic porous plates.

REFERENCES

[1] J.D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion and related problems. *Proc. Roy. Soc. London*, **A241**, 376-396, 1957.

[2] Z. Hashin, S. Shtrikman, A variational approach to the theory of the elastic behavior of multiphase materials. *J. Mech. Phys. Solids*, **11**, 127-140, 1963.

[3] R. Hill, A self-consistent mechanics of composite materials. *J. Mech. Phys. Solids*, **13**, 213-222, 1965.

[4] T. Mori, K. Tanaka, Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metal.*, **21**, 571, 1973.

[5] Y. Benveniste, A new approach to the application of Mori-Tanaka's theory in composite materials. *Mechanics of Materials*, 6, 147-157, 1987.

[6] T. Mura, *Micromechanics of Defects in Solids*. Martinus Nijhoff, 1982.

[7] N.D. Cristescu, E.-M. Craciun, E.Soos, *Mechanics of Elastic Composites*. Chapman & Hall/CRC, 2004.

[8] N.I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity*. P. Noordhoff, 1953.

[9] L.I. Sedov, *Mechanics of Continuous Media, Vol.2.* World Scientific Publ. Co., 1997.

[10] A.I. Lurie, *Theory of Elasticity*. Springer, 2005.

[11] V. Sladek, J. Sladek, L. Sator, Physical decomposition of thin plate bending problems and their solution by mesh-free methods. *Engineering Analysis with Boundary Elements*, **37**, 348-365, 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

INFLUENCE OF MICRO CRACKS ON EFFECTIVE MATERIAL PROPERTIES IN FIBER REINFORCED SMART COMPOSITE MATERIALS

M. Wünsche[1], **J. Sladek**[1], **and V. Sladek**[1]

[1]Institute of Construction and Architecture, Slovak Academy of Sciences,
84503 Bratislava, Slovakia
e-mail: wuensche@bauwesen.uni-siegen.de
jan.sladek@savba.sk
vladimir.sladek@savba.sk

Keywords: piezoelectric fiber reinforced composites, micro cracks, effective material properties, Galerkin BEM

Abstract. *In this paper, the analysis of boundary value problems with fiber reinforced piezoelectric representative volume elements (RVEs) containing micro cracks of arbitrary shape is presented. Interface cracks between fiber and matrix as well as cracks inside the matrix and fibers are investigated. For this purpose, the resulting boundary value problem is formulated as boundary integral equations (BIEs). The symmetric Galerkin method (SGBEM) is applied for the spatial discretization of the boundary to solve the BIEs numerically. Numerical examples will be presented and discussed to show the efficiency of the present SGBEM, the influences of the fiber variation, the micro cracks and the crack-face boundary condition on the effective material properties.*

1 INTRODUCTION

Smart composite materials offer certain important performance advantages over conventionally used metals. Due to the capability of converting electric energy into mechanical energy and vice versa, piezoelectric materials play an important role in smart structures. In recent years a new class of piezoelectric composites has been developed by combining passive elastic materials as matrix and piezoelectric ceramics or polymers as active fibers. These smart materials can be optimized by taking advantages of the most beneficial properties of each constituent and to satisfy the high-performance requirements according to different in-service conditions. Piezoelectric ceramics are very brittle with low fracture toughness and micro as well as macro cracks may be induced during the manufacturing and under the in-service condition [2]. Beside cracks inside the homogeneous matrix and fibers, interface cracks play an important role for the design and safety of real structures. The investigation of effective piezoelectric properties of fiber reinforced material containing micro-cracks is of special scientific significance and engineering importance. Due to the high mathematical complexity, the analytical solution of coupled multi-field problems is only possible to very simple geometrical configurations and loading conditions. Therefore, the solution of general boundary value problems requires advanced numerical methods. By applying a homogenization approach the representative volume element (RVE) is formulated and the Galerkin BEM is developed for a numerical evaluation of the effective material properties.

2 PROBLEM FORMULATION

Let us consider a linear piezoelectric fiber matrix composite with regularly distributed fibers. Micro-cracks of arbitrary shape can exist in the interface between fiber and matrix as well as inside the matrix and fibers. Without body forces, free electrical charges and by applying the generalized notation the constitutive equations are defined by

$$\sigma_{iJ}(\mathbf{x}) = C^{\zeta}_{iJKl}\varepsilon_{Kl}(\mathbf{x}) = C^{\zeta}_{iJKl}u_{K,l}(\mathbf{x}), \tag{1}$$

where u_I, σ_{iJ}, ε_{iJ} and C_{iJKl} are the generalized displacements, stresses, straines and elasticity tensor

$$u_I = \begin{cases} u_i, & I = i, \quad \text{(displacements)} \\ \varphi, & I = 4, \quad \text{(electric potential)}, \end{cases} \tag{2}$$

$$\sigma_{iJ} = \begin{cases} \sigma_{ij}, & J = j, \quad \text{(stresses)} \\ D_i, & J = 4, \quad \text{(electric displacements)}, \end{cases} \tag{3}$$

$$\varepsilon_{iJ} = \begin{cases} \varepsilon_{ij}, & J = j, \quad \text{(strains)} \\ E_i, & J = 4, \quad \text{(electric field)}, \end{cases} \tag{4}$$

$$C_{iJKl} = \begin{cases} c_{ijkl}, & J = j;\ K = k, \quad \text{(elasticity tensor)} \\ e_{lij}, & J = j;\ K = 4, \quad \text{(piezoelectric tensor)} \\ e_{ikl}, & J = 4;\ K = k, \quad \text{(piezoelectric tensor)} \\ -\kappa_{il}, & J = K = 4, \quad \text{(electric permittivity tensor)}. \end{cases} \tag{5}$$

Lower case Latin indices take the values 1 and 2 (elastic), while capital Latin indices take the values 1, 2 (elastic) and 4 (electric). A comma after a quantity designates spatial derivatives and unless otherwise stated, the conventional summation rule over repeated indices is implied. The boundary conditions

$$u_I(\mathbf{x}) = \bar{u}_I(\mathbf{x}), \quad \mathbf{x} \in \Gamma_u, \tag{6}$$

$$t_I(\mathbf{x},t) = \bar{t}_I(\mathbf{x},t), \quad \mathbf{x} \in \Gamma_t, \tag{7}$$

the continuity and equilibrium conditions on the interface

$$u_I^I(\mathbf{x}) = u_I^{II}(\mathbf{x}), \quad \mathbf{x} \in \Gamma_{if}, \tag{8}$$

$$t_I^I(\mathbf{x}) = -t_I^{II}(\mathbf{x}), \quad \mathbf{x} \in \Gamma_{if}, \tag{9}$$

are satisfied, with t_I being the generalized tractions

$$t_I(\mathbf{x}) = \sigma_{jI}(\mathbf{x}) e_j(\mathbf{x}). \tag{10}$$

In the Eqs. (6)-(10), Γ_{if} is the interface between the homogeneous domains Ω^ζ ($\zeta = 1, 2, ..., N$), Γ_t is the external boundary where the generalized tractions are prescribed, Γ_u is the external boundary where the generalized displacements are given and e_j is the outward unit normal vector. In the present work the crack is considered either as electrical impermeable

$$D_i(\mathbf{x} \in \Gamma_{c^+}) = D_i(\mathbf{x} \in \Gamma_{c^-}) = 0 \tag{11}$$

or as electrical permeable

$$D_i(\mathbf{x} \in \Gamma_{c^+}) = D_i(\mathbf{x} \in \Gamma_{c^-}), \quad \varphi(\mathbf{x} \in \Gamma_{c^+}) - \varphi(\mathbf{x} \in \Gamma_{c^-}) = 0. \tag{12}$$

Γ_{c^+} and Γ_{c^-} denote the upper and lower crack-face. The generalized crack-opening-displacements (CODs) are defined by

$$\Delta u_I(\mathbf{x}) = u_I(\mathbf{x} \in \Gamma_{c^+}) - u_I(\mathbf{x} \in \Gamma_{c^-}). \tag{13}$$

3 BOUNDARY INTEGRAL EQUATIONS AND FUNDAMENTAL SOLUTIONS

To solve the corresponding initial boundary value problem with a BEM it is formulated as boundary integral equations (BIEs). In the sense of a weighted residual formulation, the Galerkin BIEs are defined by

$$\begin{aligned} \int_\Gamma \psi(\mathbf{x}) u_J(\mathbf{x}) \mathrm{d}\Gamma_x &= \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_b} \Big[u_{IJ}^G(\mathbf{x},\mathbf{y})\, t_I(\mathbf{y}) - t_{IJ}^G(\mathbf{x},\mathbf{y})\, u_I(\mathbf{y}) \Big] \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x \\ &\quad + \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_{c^+}} t_{IJ}^G(\mathbf{x},\mathbf{y})\, \Delta u_I(\mathbf{y}) \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x, \end{aligned} \tag{14}$$

where $\psi(\mathbf{x})$ is a weighting function, Γ_b is the external boundary, $\Gamma = \Gamma_b + \Gamma_{if} + \Gamma_{c^+}$ and $u_{IJ}^G(\mathbf{x},\mathbf{y})$ are the generalized displacement fundamental solutions. The generalized traction fundamental solutions $t_{IJ}^G(\mathbf{x},\mathbf{y})$ are obtained by substitution of the displacement fundamental solutions into the constitutive equation (1) and Eq. (10) as

$$t_{IJ}^G(\mathbf{x},\mathbf{y}) = C_{qIKr} e_q(\mathbf{y}) u_{KJ,r}^G(\mathbf{x},\mathbf{y}). \tag{15}$$

The traction BIEs are derived by substituting Eq. (14) into Eqs. (1) and (10) as

$$\begin{aligned} \int_\Gamma \psi(\mathbf{x}) t_J(\mathbf{x}) \mathrm{d}\Gamma_x &= \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_b} \Big[v_{IJ}^G(\mathbf{x},\mathbf{y})\, t_I(\mathbf{y}) - w_{IJ}^G(\mathbf{x},\mathbf{y})\, u_I(\mathbf{y}) \Big] \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x \\ &\quad + \int_\Gamma \psi(\mathbf{x}) \int_{\Gamma_{c^+}} w_{IJ}^G(\mathbf{x},\mathbf{y})\, \Delta u_I(\mathbf{y}) \mathrm{d}\Gamma_y \mathrm{d}\Gamma_x, \end{aligned} \tag{16}$$

with the traction and the higher-order traction fundamental solutions

$$v_{IJ}^{G}(\mathbf{x},\mathbf{y}) = -C_{pIKs}e_{p}(\mathbf{x})u_{KJ,s}^{G}(\mathbf{x},\mathbf{y}), \tag{17}$$

$$w_{IJ}^{G}(\mathbf{x},\mathbf{y}) = C_{pIKs}e_{p}(\mathbf{x})C_{qJLr}e_{q}(\mathbf{y})u_{KL,sr}^{G}(\mathbf{x},\mathbf{y}). \tag{18}$$

The fundamental solutions for homogeneous, generally anisotropic and linear piezoelectric solids are applied [5].

4 SOLUTION ALGORITHM

For the spatial discretization, the external boundary Γ_b, the interface Γ_{if} and the crack-face Γ_{c^+} of the piezoelectric solid are discretized by quadratic elements

$$\Gamma = \Gamma_b + \Gamma_{if} + \Gamma_{c^+} = \sum_{e=1}^{E} \Gamma_e, \tag{19}$$

with $E = E_b + E_{if} + E_{c^+}$. The strongly singular and hypersingular boundary integrals are computed directly with special analytical and numerical techniques [1]. The symmetry properties of the fundamental solutions are utilized to enhance the efficiency of the present Galerkin-BEM.

After spatial discretizations the BIEs (14) and (16) lead to the following systems of linear algebraic equations for each subdomain Ω^{ζ} $(\zeta = 1, 2, ..., N)$

$$\mathbf{C}_{\zeta}\mathbf{u}_{\zeta} = \mathbf{U}_{\zeta}^{S}\mathbf{t}_{\zeta} - \mathbf{T}_{\zeta}^{S}\mathbf{u}_{\zeta} + \mathbf{T}_{\zeta}^{S}\Delta\mathbf{u}_{\zeta}, \tag{20}$$

$$\mathbf{D}_{\zeta}\mathbf{t}_{\zeta} = \mathbf{V}_{\zeta}^{S}\mathbf{t}_{\zeta} - \mathbf{W}_{\zeta}^{S}\mathbf{u}_{\zeta} + \mathbf{W}_{\zeta}^{S}\Delta\mathbf{u}_{\zeta}. \tag{21}$$

By invoking the boundary conditions (6),(7), the interface conditions (8) and (9) and the crack-face conditions (11), (12) all equations can be summarized to the system

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{22}$$

where $\mathbf{A}$ is the system matrix, $\mathbf{x}$ is the vector of the unknown boundary data and the vector $\mathbf{b}$ contains the prescribed boundary data

5 COMPUTATION OF EFFECTIVE MATERIAL PROPERTIES

To compute the effective material properties of a fiber reinforced piezoelectric composite a homogenization procedure is applied. A uniform distribution of piezoelectric fibers is assumed in the present case and the corresponding representative volume element is formulated. The relation between the generalized average stresses and the boundary tractions is expressed as

$$\langle \sigma_{iJ} \rangle = \frac{1}{V}\int_{\Omega} \sigma_{iJ}\mathrm{d}\Omega = \begin{cases} \dfrac{1}{2V}\displaystyle\int_{\Gamma} (t_i x_j + t_j x_i)\,\mathrm{d}\Gamma, & i,j = 1,2 \\ \dfrac{1}{V}\displaystyle\int_{\Gamma} t_J x_i\,\mathrm{d}\Gamma, & i = 1,2;\, J = 4, \end{cases} \tag{23}$$

where Ω is the domain of the RVE and Γ is the whole boundary of the RVE containing the external boundary Γ_b, the crack-faces Γ_c, and the interfaces Γ_{if}. In the same sense the volume

integral of the generalized average straines can be transformed to the boundary by

$$\langle \varepsilon_{iJ} \rangle = \frac{1}{V} \int_{\Omega} \varepsilon_{iJ} \mathrm{d}\Omega = \begin{cases} \dfrac{1}{2V} \displaystyle\int_{\Gamma} (u_i e_j + u_j e_i)\, \mathrm{d}\Gamma, & i,j = 1,2 \\ \dfrac{1}{V} \displaystyle\int_{\Gamma} u_J e_i \, \mathrm{d}\Gamma, & i = 1,2; J = 4. \end{cases} \tag{24}$$

Taking into account the constitutive relationships

$$\langle \sigma_{iJ} \rangle = C^{eff}_{iJKl} \langle \varepsilon_{Kl} \rangle \tag{25}$$

the effective material coefficients C^{eff}_{iJKl} can be computed using the generalized average stresses (23) and the generalized average strains obtained from the numerical solution of properly selected boundary value problems in the RVE sample [3, 4].

6 NUMERICAL RESULTS

As a numerical example, we consider a square RVE of length l with a central circular fiber of radius r. Between the fiber and the matrix a interface crack of the arc length 2α exist.

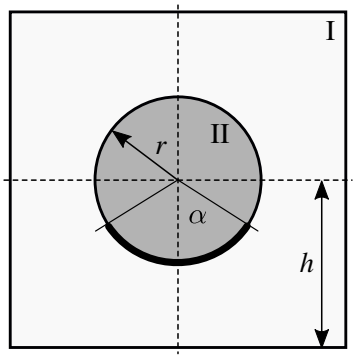


Figure 1: A square RVE with a interface crack between the central circular fiber and the matrix

As non-piezoelectric matrix (domain I) Epoxy is chosen, which has the material parameters

$$\begin{aligned} &C_{11} = 8.0\text{GPa}, \quad C_{12} = 4.4\text{GPa}, \quad C_{22} = 8.0\text{GPa}, \quad C_{66} = 1.8\text{GPa}, \\ &\kappa_{11} = 0.0372\text{C/(GVm)}, \quad \kappa_{22} = 0.0372\text{C/(GVm)}, \end{aligned} \tag{26}$$

while for the fiber (domain II) PZT7 with the following material constants

$$\begin{aligned} &C_{11} = 154.8\text{GPa}, \quad C_{12} = 82.7\text{GPa}, \quad C_{22} = 131.4\text{GPa}, \quad C_{66} = 25.7\text{GPa}, \\ &e_{21} = -2.121\text{C/m}^2, \quad e_{22} = 9.522\text{C/m}^2, \quad e_{16} = 9.349\text{C/m}^2, \\ &\kappa_{11} = 4.065\text{C/(GVm)}, \quad \kappa_{22} = 2.079\text{C/(GVm)} \end{aligned} \tag{27}$$

is applied. For the numerical computations the external boundary of the RVE is discretized by using a uniform mesh with 16 quadratic elements and the circular interface including the crack is divided into 20 elements. The crack is assumed as impermeable for the electric field. The normalized effective material properties for a volume fraction of 20% of the piezoelectric fiber and different crack angles α are shown in Fig. 2.

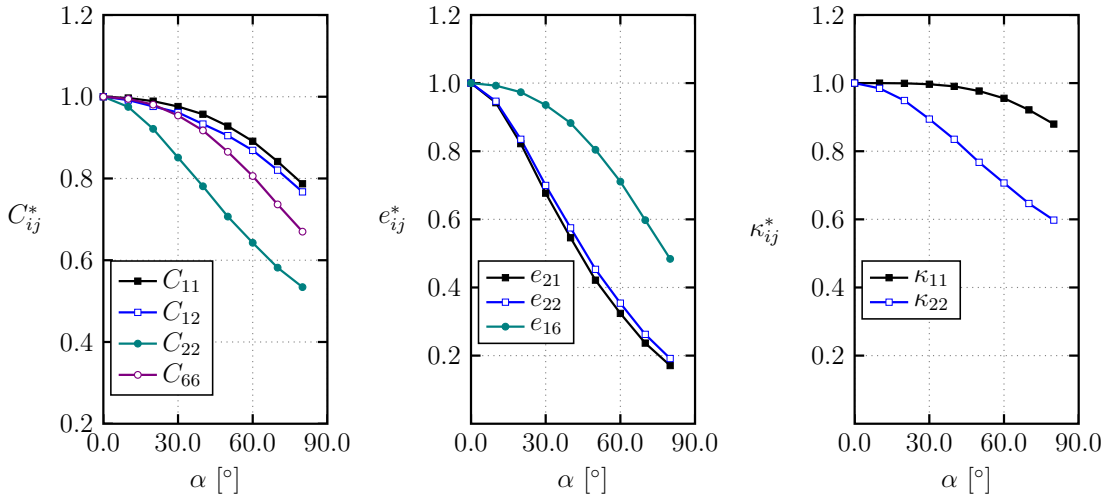


Figure 2: Effective material properties for different crack lengths

As depicted in Fig. 2 the interface crack has a significant influence on the effective material properties of the composite. The elastic constants decrease drastically with increasing crack angles α. For a crack of the arc length of $90°$ the elastic constants are only 55-80% of the uncracked composite. In the same sense the electric permittivities κ_{11} and κ_{22} reduce significantly with increasing interface crack lengths. It should be mentioned that this is the consequence of the applied electric impermeable crack-face boundary condition. This implies that the crack is free of electric charges and the electric field lines circumvent the crack. In contrast if the crack is considered as fully permeable for the electric field both crack-faces have equal electric potentials and therefore the crack is fully penetrated by the electric field. In other words the crack does not exist for the electric field and the application of the permeable crack modell would lead to identical electric permittivities for the RVE with and without a crack. As well expected the piezoelectric constants show a similar tendency and decrease with a increasing interface crack arc length.

7 CONCLUSIONS

The analysis of effective material properties for fiber reinforced piezoelectric representative volume elements (RVEs) containing micro cracks of arbitrary shape is presented in this paper. Both, interface cracks between fiber and matrix as well as cracks inside the matrix and fibers are investigated. In order to solve the resulting boundary value problem with the Galerkin BEM it is formulated as boundary integral equations (BIEs). Analytical and numerical techniques are used for a direct computation of the strongly singular and hypersingular boundary integrals. The developed BEM is general without limitations of the geometry, poling direction and material anisotropy. Numerical examples will be presented and discussed to show the efficiency of the present Galerkin BEM, the influences of the fiber variation, the micro cracks and the crack-face boundary condition on the effective material properties.

ACKNOWLEDGEMENT

This work is supported by the Slovak Academy of Sciences Project (SASPRO) 0106/01/01. The financial support is gratefully acknowledged.

REFERENCES

[1] L.J. Gray, Evaluation of singular and hypersingular Galerkin boundary integrals: direct limits and symbolic computation. In: Sladek V, Sladek J, editors. Advances in boundary elements Southampton, UK: Computational Mechanics Publishers; 1998. p. 3384.

[2] M. Kuna, Fracture mechanics of piezoelectric materials - where are we right now ?. *Engineering Fracture Mechanics*, **77**, 309–326, 2010.

[3] Z. Li, C. Wang, C. Chen, Effective electromechanical properties of transversely isotropic piezoelectric ceramics with microvoids. *Computational Materials Science*, *27*, 381–392, 2003.

[4] J. Sladek, V. Sladek, S. Krahulec, C. Song, Crack analyses in porous piezoelectric brittle materials by the SBFEM. *xxx*, **116**, 103–112, 2016.

[5] M. Wünsche, F. García-Sánchez, A. Sáez, Ch. Zhang, A 2D time-domain collocation-Galerkin BEM for dynamic crack analysis in piezoelectric solids. *Engineering Analysis with Boundary Elements*, **34**, 377–387, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DEFECT ACCUMULATION IN NANOPOROUS WEAR-RESISTANT COATINGS UNDER COLLECTIVE RECRYSTALLIZATION. SIMULATION BY HYBRID CELLULAR AUTOMATON METHOD

Dmitry D. Moiseenko[1], Pavel V. Maksimov[2], Sergey V. Panin[3] and Viktor E. Panin[4]

[1,2,3,4] Institute of Strength Physics and Materials Science SB RAS
634055, Akademichesky av. 2/4, Tomsk, Russia
e-mail: moisey@rocketmail.com[1], mpv@ispms.tsc.ru[2]

[3,4] National Research Tomsk Polytechnic University
634050, Lenin av. 30, Tomsk, Russia
e-mail: svp@ispms.tsc.ru[3], paninve@ispms.tsc.ru[4]

Keywords: wear-resistant coatings, cellular automata, recrystallization, thermal loading.

Abstract. *A modification of a multiscale hybrid discrete-continual approach of excitable cellular automata is developed. The new version of the method is completed by taking into account porosity and nanocrystalline structure of a material and the algorithms of calculation of local force moments and angular velocities of microrotations. The excitable cellular automata method was used to carry out numerical experiment (NE) for heating of continuous and nanoporous specimens consisting of nanocrystalline TiAlC coatings. The numerical experiments have shown that nanoporosity allows to substantially reducing the rate of collective crystallization. Nanoporosity slowed down propagation of the heat front in specimens. This fact can play both positive and negative roles in deposition of coating and its further use. On the one hand, by slowing the heat front propagation one can significantly reduce the level of thermal stresses in deeper layers of the material. On the other hand, such deceleration in case of the high value of the coefficient of thermal expansion can give rise to the formation of large gradients of thermal stress, which initiate nucleation and rapid growth of the main crack.*

1 INTRODUCTION

Currently, composite coatings on the basis of MAX phases with composite properties (tribological and heat-proof at the same time) become more and more relevant [1–7]. This new class of materials is characterized by unique properties combining the benefits of both metallic alloys and ceramic materials. On the one hand, such compounds demonstrate high heat conductivity and thermal resistance. On the other hand, they feature high elastic moduli, low thermal expansion coefficient, high heat resistance and extreme high-temperature strength. Such materials are very resistant to cycle loads and temperatures higher than 1000 °C, and in this temperature range their resistance to cycle loads leaves behind the majority of known heat-proof and heat-stable materials, including the intermetallide-based alloys [3].

The practical interest to such new materials is primarily connected with new requirements to critical units of aircraft. In combination with low density of the composite the unique characteristics of MAX-phases unveil wide range of their application in rocket and aircraft engineering. For instance, the implementation of heat-stable Ti-Al-based material with protective coating allows increasing up to 50% the relation of lift force to weight in comparison with the best Ni-based analogues [6].

However, the creation of new classes of wear resistant materials operating under extremely high temperature gradients (~1000 K) requires detailed study of both initial microstructure of created coatings and regulations of its evolution in the course of thermal and mechanical loads. For example, the melting temperature of γ-TiAl phase is about 1750 K, which limits the application of such intermetallides under extreme thermocycling.

Recent years have witnessed the works on modeling of either behavior of wear resistant and heat-proof coatings under mechanical loading [8, 9], or the process of their creation [10–11]. The majority of them is dedicated to the investigation of titanium-aluminum alloys. Unfortunately, the authors of present work are not aware of works that aim at detailed study of the change in the mesostructure of Ti-Al-C coatings under thermocycling at temperatures close to melting temperatures of separate structural elements. The capability of collective crystallization as well as its impact on redistribution of internal thermal stresses in a mesostructure are also far from being studied thoroughly.

Consequently, the authors of this work tried to implement computer modeling to study collective crystallization of nanostructured highly porous Ti-Al-C coating, as well as the effect of dynamic rearrangement of the structure on distribution of thermal stresses.

In the present work, the authors refer to the principles of physical mesomechanics on scale invariability assuming that a real solid body is an hierarchically organized system [12]. Such 3D-system contains the combination of interfaces, which are independent 2D-systems, in which rotary wave fluxes of mass and energy propagate [13]. Fig. 1 depicts the scheme of interaction of 2D planar subsystem of interface and 3D subsystem of the crystal lattice.

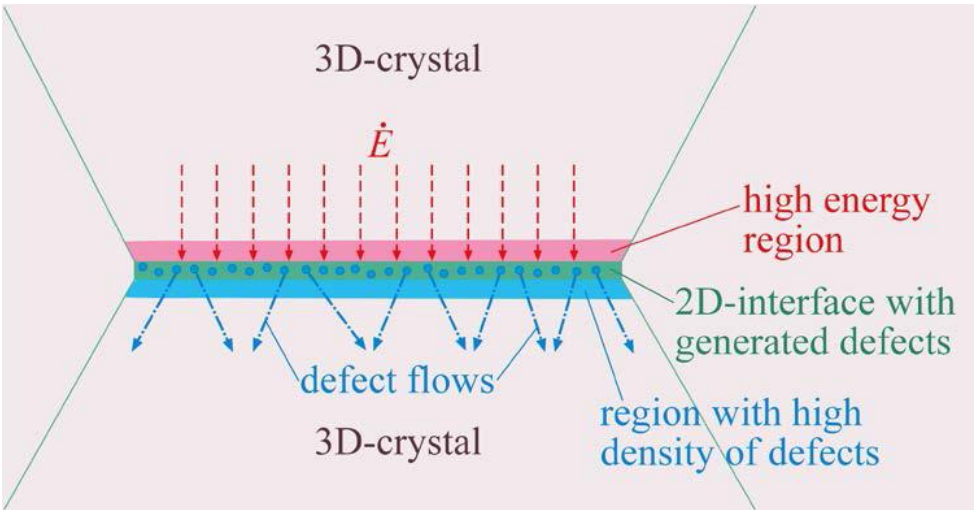


Fig. 1. Scheme of interaction of 2D planar subsystem of interface and 3D subsystem of the crystal lattice.

Rotary wave fluxes are particularly distinguished under dynamic loading, for instance under friction with run-out. Most distinctively the microrotations are demonstrated in experiments on fragmentation during propagation of a shock wave [14]. Rotation modes of deformation occur after the action of local moment of force. For computer modeling of the initiation and consequent evolution of fields of local force moments under intense dynamic loads, a discrete continuum method of stochastic excitable cellular automata was developed.

The authors were mostly referring to existing works on modeling of recrystallization processes [15–17] and to their own previous studies [18–20].

2 SIMULATION METHOD

Within the framework of present study, a new modification of the multiscale hybrid discrete continuum approach of stochastic excitable cellular automata was developed. The new method explicitly accounts for the porosity and nanocrystalline structure of material, the algorithm for computing local force moments and angular velocities of microrotations occurring in structurally nonuniform medium.

In the frame of SECA method (Stochastic Excitable Cellular Automata), the modeled specimen is represented as a cellular automaton, i.e. the combination of ordered active elements, each both imitating in part the material contained in particular volume of space and characterized by such numerical parameters as heat capacity and heat conductivity. The network of cellular automaton elements is divided into *clusters*, each being the model of a separate grain having its own orientation of the crystal lattice characterized by Euler angles ψ, φ, η.

To determine the temperature of the elements it was necessary to introduce a discrete model of heat transfer of heterogeneous medium. At zero step of the numerical experiment, for each element, the initial values of temperature, deformation, heat conductivity, heat capacity and coefficient of linear thermal expansion were assigned. Then, at each n-th time step, a new temperature value is calculated for an element with due consideration of heat fluxes from each adjacent element.

$$T_i^n = T_i^{n-1} + \frac{1}{c_i \rho_i V} \sum_{k=1}^{N} Q_{ik}^n , \tag{1}$$

where T_i^{n-1}, T_i^n are temperatures of the *i*-th element at *n*-1 and *n*-th time steps, c_i is heat capacity of the *i*-th element, ρ_i is the density of the *i*-th element, *V* is element volume, Q_{ik}^n is the flux of heat energy from neighboring *k*-th element into considered *i*-th at the *n*-th time step, *N* is the number of neighbors.

The change of heat energy Q_{ik}^n is calculated using Fourier law:

$$Q_{ik}^n = \frac{\lambda_{ik}\,\Omega}{l}\left(T_k^{n-1} - T_i^{n-1}\right)\Delta\ . \tag{2}$$

Here λ_{ik} is he coefficient of mutual heat conductivity, *l* is the distance between the centers of considered elements, Ω is the area of adjacent plane, Δτ is time step. The coefficient of mutual heat conductivity is calculated as follows:

$$\lambda_{ik}=(\lambda_i+\lambda_k)/2. \tag{3}$$

Here λ_i, λ_k are heat conductivity coefficients of the *i*-th and *k*-th elements, with

$$\lambda_j = \tilde{\lambda}_j e^{-\phi_j \theta_{ik}}\ \ (j = i, k), \tag{4}$$

where $\tilde{\alpha}_j$ is the coefficient of heat conductivity of the *j*-th automaton, ϕ_j is the coefficient determining the fraction of impact of phonons on the material heat conductivity, u_{ik} is angle (in degrees) determined by the difference of crystal lattice orientations of *i*-th and *k*-th automata (in current program version $\theta_{ik} = |\phi_i - \phi_k|$).

After the calculation of new temperature values T_i^n using equation (1), for each *i*-th element the following values are calculated:

a. change in deformation due to thermal expansion:

$$\Delta_Q \varepsilon_i^n = \alpha_i \cdot (T_i^n - T_i^{n-1}); \tag{5}$$

b. corresponding changes in stress and mechanical energy A_i^n:

$$\Delta_Q \sigma_i^n = \alpha_i \cdot (T_i^n - T_i^{n-1}) \cdot Y_i^n, \tag{6}$$

$$\Delta_Q A_i^n = \frac{Y_i^n \cdot V_{CA}}{2} \cdot \left(sign(\varepsilon_i^{n-1} + \Delta_Q \varepsilon_i^n) \cdot (\varepsilon_i^{n-1} + \Delta_Q \varepsilon_i^n)^2 - sign(\varepsilon_i^{n-1}) \cdot (\varepsilon_i^{n-1})^2\right); \tag{7}$$

c. new values of heat energy Q_i^n:

$$Q_i^n = Q_i^{n-1} + c_i \cdot (T_i^n - T_i^{n-1}) - \Delta_Q A_i^n; \tag{8}$$

d. according to (8), a correction is made for a new temperature value:

$$T_i^n = T_i^n - \frac{\Delta_Q A_i^n}{c_i}. \tag{9}$$

Thus, the fraction of element heat energy change goes for the change in its mechanical energy due to the process of thermal expansion. As a result, internal thermal stresses accumulate, which transform into fluxes of mechanical energy distributed long the network of elements.

The modeling of heat transfer in porous body should account the transfer of heat energy by radiation. In the frame of the considered method, a state of element is introduced that indicates its belonging to a pore. If the considered *i*-th automaton element on the first coordination sphere has an element in *pore* state, then the heat inflow into the *i*-th element is calculated from the assumption of cubic packing of elements, and that the pore boundaries are reflective, i.e. radiated and reflected rays propagate perpendicularly to the boundary. In this case, the neighboring element is the elements in *material* state that is the closest on the way of propagation of rays from the *i*-th element. Let this element have index *k*. The change in heat energy of the *i*-th element after the interaction with the *k*-th element at current *n*-th time step Q_{ik}^n is determined as follows.

In the case of radiant heat exchange, the *i*-th element, in addition to its generated radiation E_i^n, partially reflects the incident energy from the *k*-th element. The sum of energy of generated and reflected radiation is the effective radiation of the *i*-th element $\left(E_i^n\right)_{eff}$. For the calculation of the radiant heat exchange between elements, it is important to determine resulting radiation, which is the difference between the radiant flux received by an element and radiant flux which it radiates. To determine the flux density of the resulting radiation q_{ik}^n, an energy balance equation is composed:

$$q_{ik}^n = \left(E_i^n\right)_{eff} - \left(E_k^n\right)_{eff}. \tag{10}$$

On the other hand, the resulting radiation is the difference between the flux radiated by an element and the fraction of absorbed energy:

$$q_{ik}^n = E_i^n - \mu_i \left(E_k^n\right)_{eff}, \tag{11}$$

where $\mu_i \in [0;1]$ - is absorption coefficient of the *i*-th element. Taking into account (10) and (11) yields:

$$\left(E_i^n\right)_{eff} = q_{ik}^n \left(1 - \frac{1}{\mu_i}\right) + \frac{E_i^n}{\mu_i}. \tag{12}$$

To calculate the heat energy Q_{ik}^n transferred from the *k*-th element to the *i*-th one, a model of heat flux is used in a system of paralleled reflecting plates. In the frame of the model, using the principles of the law of radiant heat exchange, the equation for calculating the radiant heat exchange between elements can be derived. Let us derive the equation for determining heat flux q_{ik}^n from the *k*-th element to the *i*-th one. This value is the difference between effective fluxes from each of the elements:

$$q_{ik}^n = E_k^n - E_i^n. \tag{13}$$

Thus, according to (12) and taking into account that under stabilized regime, the resulting fluxes for each of the elements have equal value and opposite signs ($q_{ik}^n = -q_{ki}^n$):

$$\begin{aligned} \left(E_k^n\right)_{eff} &= -q_{ik}^n\left(1-\frac{1}{\mu_k}\right)+\frac{E_k^n}{\mu_k}, \\ \left(E_i^n\right)_{eff} &= q_{ik}^n\left(1-\frac{1}{\mu_i}\right)+\frac{E_i^n}{\mu_i}, \end{aligned} \tag{14}$$

where μ_i is the absorption coefficient of material of the *i*-th element. Substituting (14) into (13) yields the expression for heat flux q_{ik}^n :

$$q_{ik}^n = \frac{\mu_i\mu_k}{\mu_i+\mu_k-\mu_i\mu_k}\left[\frac{E_k^n}{\mu_k}-\frac{E_i^n}{\mu_i}\right]. \tag{15}$$

According to the laws of Stefan-Boltzmann and Kirchhoff, the relation of the flux density of heat energy E_i^n —radiated by the surface of heated *i*-th element containing material—to its absorption coefficient is calculated as follows:

$$\frac{E_i^n}{\mu_i} = \eta\left(\frac{T_i^n}{100}\right)^4. \tag{16}$$

Here η = 5.67 W/(m^2·K^4) is the radiation coefficient of gray body, μ_i is absorption coefficient of material of the *i*-th element, T_i^n - temperature of the *i*-th element at previous time step.

Obviously, the heat flux Q_{ik}^n over the time step $\Delta\tau$ through the element plane area Ω is expressed as:

$$Q_{ik}^n = q_{ik}^n\Omega\Delta\tau. \tag{17}$$

Taking into account (15) and (16) yields the expression for heat flux Q_{ik}^n :

$$Q_{ik}^n = \frac{\mu_i\mu_k}{\mu_i+\mu_k-\mu_i\mu_k}\eta\left[\left(\frac{T_k^n}{100}\right)^4-\left(\frac{T_i^n}{100}\right)^4\right]\Omega\Delta\tau. \tag{18}$$

Here, $\mu_i, \mu_k \in [0;1]$ - are absorption coefficients of the *i*-th and *k*-th elements.

Thus, the modeling of radiant heat exchange using (2) requires adding expression (18):

$$Q_{ik}^n = \frac{\lambda_{ik}\,\Omega}{l}\left(T_k^{n-1}-T_i^{n-1}\right)+\frac{\mu_i\mu_k}{\mu_i+\mu_k-\mu_i\mu_k}\eta\left[\left(\frac{T_k^n}{100}\right)^4-\left(\frac{T_i^n}{100}\right)^4\right]\Omega\Delta\tau. \tag{19}$$

Taking into account that the amount of radiated energy is known to be larger than the amount of energy absorbed by the walls of pores due to dissipation (connected with the absorption and dissipation coefficients), the energy dissipation during the heat transfer through

pores is expressed using optical thickness of gaseous filler of pores. Hence, coefficients μ_i and μ_k should be selected with the regard for the heat energy dissipation in the filler gas in pores. The chemical composition of the gas can be determined experimentally.

Further, due to heterogeneous thermal expansion, in different structural elements, internal thermal stresses accumulate that transform into fluxes of mechanical energy redistributed along the network of a cellular automaton. The method explicitly accounted grain boundaries, their curvature and grain-boundary angles of the crystal lattice. The created algorithms allowed calculating the value of local force moments, vorticity tensor and dissipation of torsion energy conditioned by the formation of new defect structures.

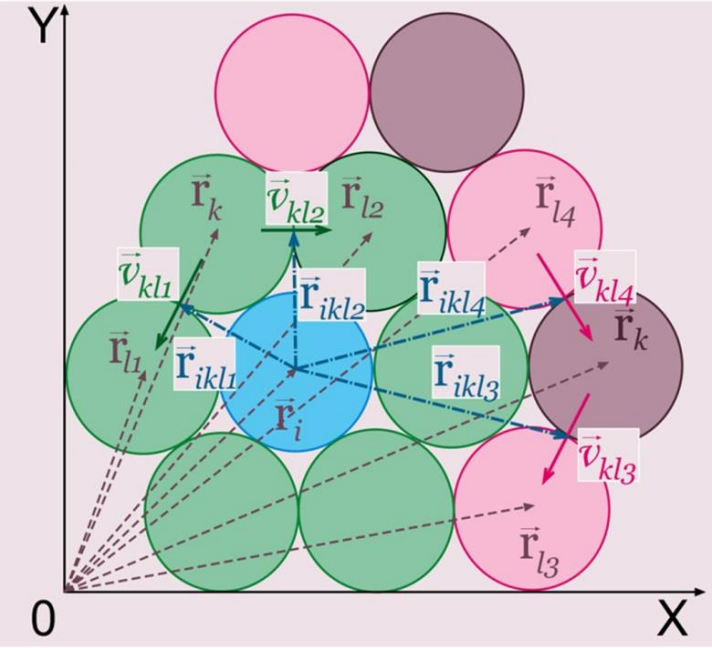


Fig. 2. Scheme for calculation of angular velocity of material rotation in cellular automaton element

Angular velocity of the i-th element in Fig. 2 under the flux of matter acting through the boundary of the *k*-th and *l*-th elements (each *k*-th element lies on the 1st coordination sphere of the *i*-th element; each *l*-th element lies on the intersection of 1st coordination spheres of the *i*-th and corresponding *k*-th elements) is calculated as follows:

$$\vec{\omega}_{ikl} = \frac{\vec{r}_{ikl} \times \vec{v}_{kl}}{\left|\vec{r}_{ikl}\right|^2}. \tag{20}$$

The velocity of matter flux at the boundary of the k-th and l-th elements is calculated using Newton's viscous flow equation:

$$\vec{v}_{kl} = m_{kl}(p_l - p_k)\vec{n}_{kl}, \tag{21}$$

where p_l, p_k are pressures (all-way stresses) in the *k*-th and *l*-th elements, $\vec{n}_{kl}$ is vector of a normal to the boundary of the *k*-th and *l*-th elements, m_{kl} is movability of the boundary between the *k*-th and *l*-th elements.

The total angular velocity of the *i*-th element is determined as the following sum:

$$\vec{\omega}_i = \sum_{k=1}^{K} \sum_{l=1}^{L} \vec{\omega}_{ikl} \,. \tag{22}$$

Here K is the number of elements on 1st coordination sphere of the i-th element, L is the number of elements at the intersection of the 1st coordination spheres of the i-th element and each k-th neighbor.

Three-dimensional rotation angle of the i-th element over time τ ($\Delta\vec{\gamma}_i$) is proportional to the total angular velocity:

$$\Delta\vec{\gamma}_i = \vec{\omega}_i \tau \,. \tag{23}$$

Change of the force moment of the i-th element over time τ ($\Delta\vec{M}_i$) is calculated as follows:

$$\Delta\vec{M}_i = \frac{G\pi r_c^3 \Delta\vec{\gamma}_i}{2} = \frac{G\pi r_c^3 \vec{\omega}_i \tau}{2} \,. \tag{24}$$

Here G is shear modulus of material, contained in the i-th element, r_c is element radius.

Vortex vector is determined as follows:

$$\vec{\omega}_i = rot\ \vec{v} \,. \tag{25}$$

Components of vortex vector are expressed as follows:

$$\begin{aligned} \omega_1 &= 2\Omega_{23}, \\ \omega_2 &= 2\Omega_{31}, \\ \omega_3 &= 2\Omega_{12}, \end{aligned} \tag{26}$$

where Ω is vorticity tensor,

$$\Omega_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i} \right). \tag{27}$$

The meaning of angular velocity calculated using expressions 20–22 is similar to that of vortex vector. Taking into account equation 24, the following expressions for vortex vector components are yielded:

$$\begin{aligned} \omega_1 &= \frac{2M_1}{G\pi r_c^3 \tau}, \\ \omega_2 &= \frac{2M_2}{G\pi r_c^3 \tau}, \\ \omega_3 &= \frac{2M_3}{G\pi r_c^3 \tau}. \end{aligned} \tag{28}$$

Thus, vorticity tensor Ω can be expressed as follows:

$$\Omega = \begin{pmatrix} 0 & \frac{M_3}{G\pi r_c^3 \tau} & -\frac{M_2}{G\pi r_c^3 \tau} \\ -\frac{M_3}{G\pi r_c^3 \tau} & 0 & \frac{M_1}{G\pi r_c^3 \tau} \\ \frac{M_2}{G\pi r_c^3 \tau} & -\frac{M_1}{G\pi r_c^3 \tau} & 0 \end{pmatrix}. \quad (29)$$

The member responsible for accumulation of latent energy of defects is expressed in the following way:

$$\left[\Delta \vec{A}_d\right]_i^n = \frac{k_{tors} G_i \left|\Delta \vec{\gamma}_i^n\right| \pi r_c^3}{2} \Delta \vec{\gamma}_i^n . \quad (30)$$

Here, k_{tors} is the rate of accumulation of defects (can be determined experimentally).

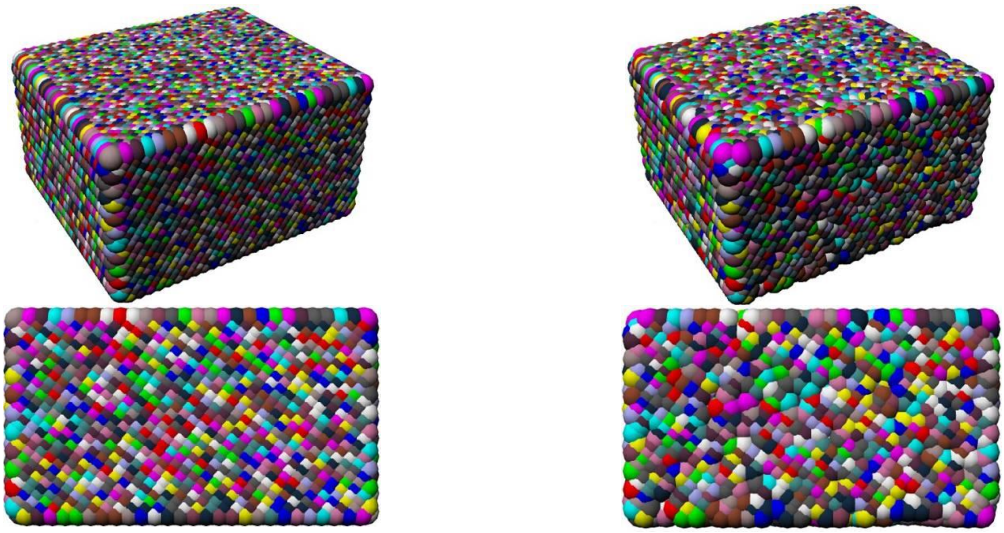
It should be noted that expression (30) has general sense and, depending on formulated rules of switching of cellular automaton, can have the sense of energy required for reversible structure and phase transformation. In such a case, the coefficient k_{tors} can be calculated using atomic configurations of initial and final states.

3 NUMERICAL EXPERIMENT

In the frame of this work, theoretical study was performed of the character of structure and phase transformations in nanocrystalline and nanoporous specimens under thermocycling. The hybrid method of excitable and bistable cellular automata was used to simulate heat transfer with due account for the process of collective recrystallization.

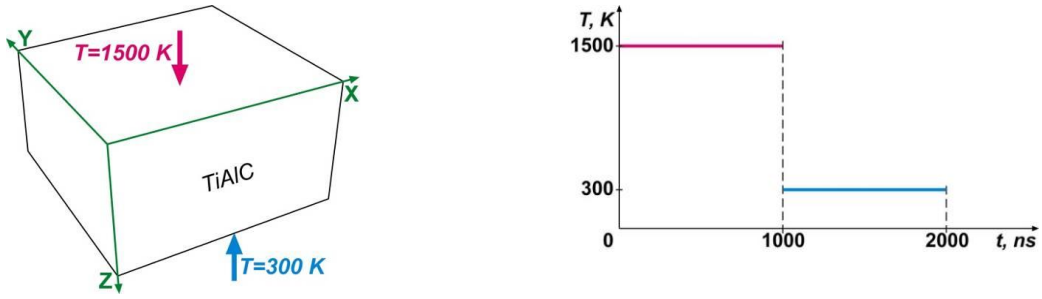
Using the method of stochastic excitable cellular automata a numerical experiment (NE) was performed including the heat-up of solid and nanoporous specimens consisting of nanocrystalline TiAlC with high ($12 \cdot 10^{-6}$ K^{-1}) and low ($7 \cdot 10^{-6}$ K^{-1}) thermal expansion coefficient (TEC). The NE included the simulation of collective crystallization and thermal expansion of materials, as well as the explicit account for thermal transition of mechanical energy into its latent fraction during local rotations: at each time step, the mechanical energy of each element of a cellular automaton was decreasing for a value proportional to the square of absolute value of accumulated three-dimensional angle of rotation of this element.

Each specimen was simulated by means of a cellular automaton with fcc arrangement of elements with the dimensions of 60 nm. The dimensions of each element were 2.88 × 2.4 × 1.6 μm; the dimensions of grain and pore were set to 60 nm (Fig. 3). Initial temperature of each element was 300 K; initial values of deformation and stress were zero. The value of time step was 0.02 ns. Thermal loading consisted of two time steps of 1000 μs: at the first stage, heating up to upper limit of 1500 K was simulated; at the second stage, cooling to 300 K (Fig. 4).



a **b**

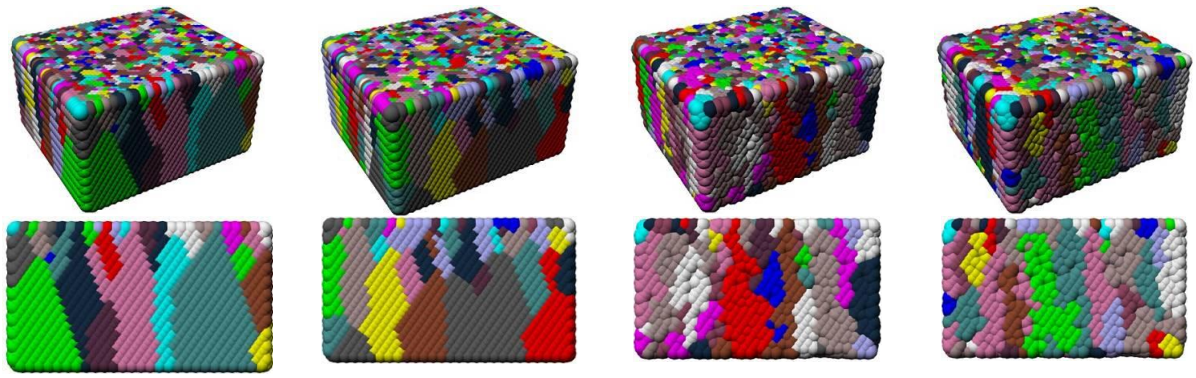
Fig. 3. Grain structure of specimens (in three-dimensional form and at the front face) at the initial moment of time: a) without pores with high and low TEC; b) porous with high and low TEC.



a **b**

Fig. 4. Scheme of thermal load application (a) and time dependence of temperature at upper face (b)

Fig. 5 illustrates the images of grain structure of specimens (in three-dimensional form and at the front face), formed after collective crystallization by specific moment of time.



300 ns 300 ns 600 ns 600 ns

a **b** **c** **d**

Fig. 5. Images of grain structure of specimens (in three-dimensional form and at the front face), formed after collective crystallization by specific moment of time: a) without pores with high TEC; b) without pores with low TEC; c) porous with high TEC; d) porous with low TEC.

Figs. 6–9 depict the time dependencies of mean values of effective stress, specific latent mechanical energy, relative latent mechanical energy and modulus of force moment at the stage of heating of the specimens.

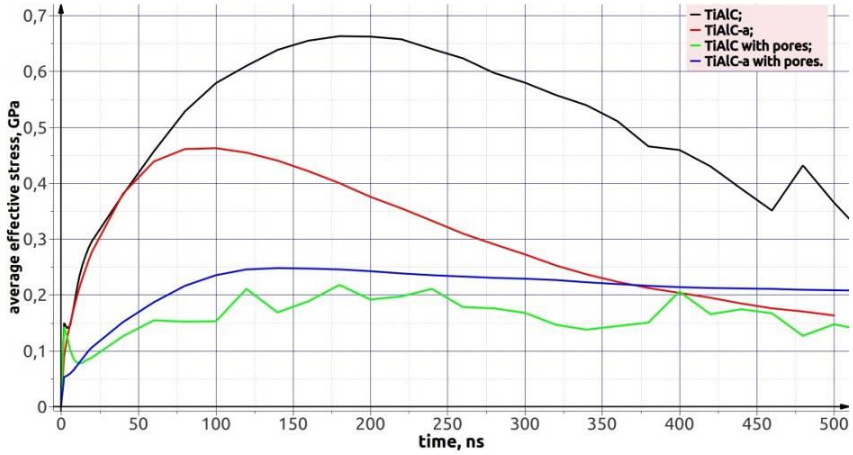


Fig. 6. Time plots of mean **effective stress** for specimens without pores with high TEC (black), without pores with low TEC (red), porous with high TEC (green), porous with low TEC (blue).

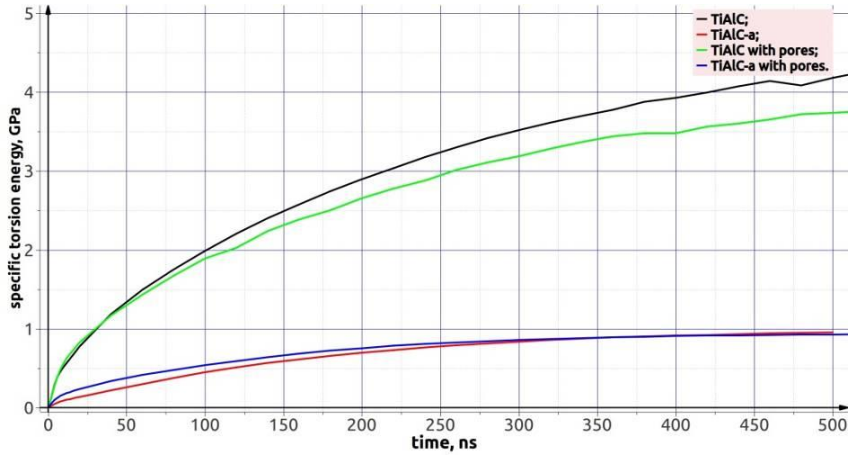


Fig. 7. Time plots of mean **specific latent mechanical energy** for specimens without pores with high TEC (black), without pores with low TEC (red), porous with high TEC (green), porous with low TEC (blue).

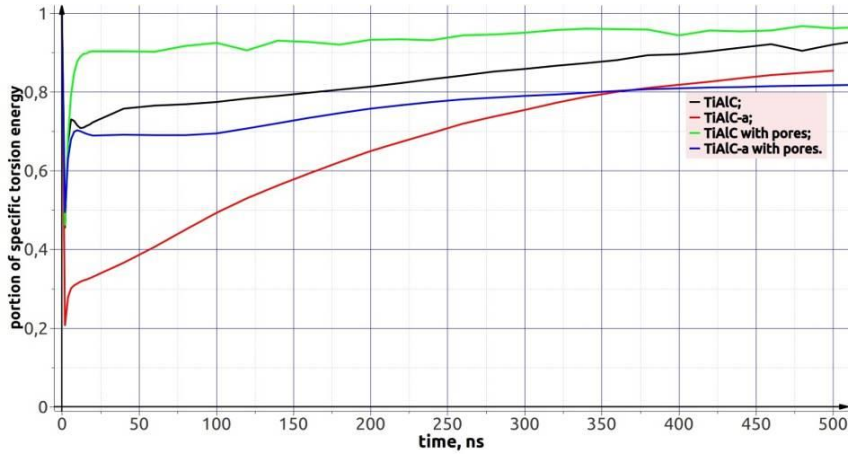


Fig. 8. Time plots of mean **relative latent mechanical energy** for specimens without pores with high TEC (black), without pores with low TEC (red), porous with high TEC (green), porous with low TEC (blue).

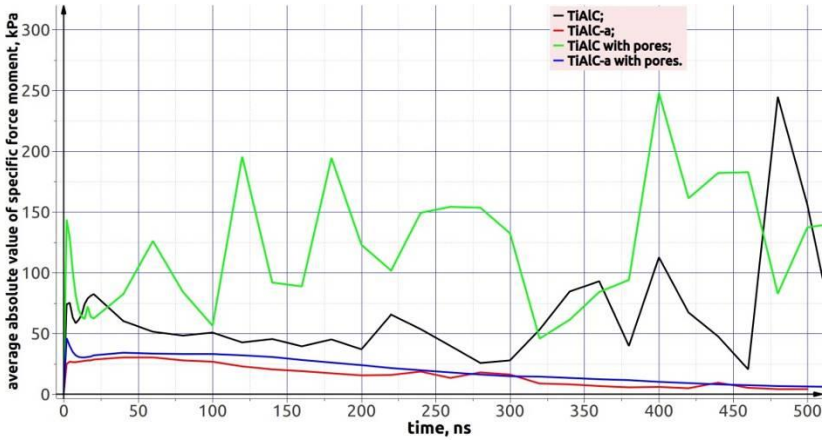


Fig. 9. Time plots of mean **modulus of force moment** for specimens without pores with high TEC (black), without pores with low TEC (red), porous with high TEC (green), porous with low TEC (blue).

The study performed by direct numerical experiment has elucidated that the introduction of the nanoporosity allows appreciably reducing the rate of collective crystallization. It should be noted that the nanoporosity was slowing down the propagation of the heat front in studied modeled specimens. This fact can play both positive and negative role in the process of deposition and further use of a coating. On the one hand, the slowdown of heat front can considerably reduce the level of thermal stresses in deep layers of a material. On the other hand, the slowdown of heat propagation front in the case of high TEC of a material can result in the formation of sharp gradients of thermal stresses that initiate the origination and rapid propagation of a main crack.

In addition, it can be noted that in porous specimens, the rate of collective crystallization was considerably lower than that in nanocrystalline specimens. The study results demonstrate that the value of thermal expansion coefficient has no impact on the rate of recrystallization in nonporous specimens. This means that the driving force in the recrystallization process in this case is primarily the temperature gradient, rather than mechanical stresses.

Figs. 10–13 represent the images of grain structure, distribution of effective stress, modulus of force moment and relative latent mechanical energy in studied specimens at different moments of time.

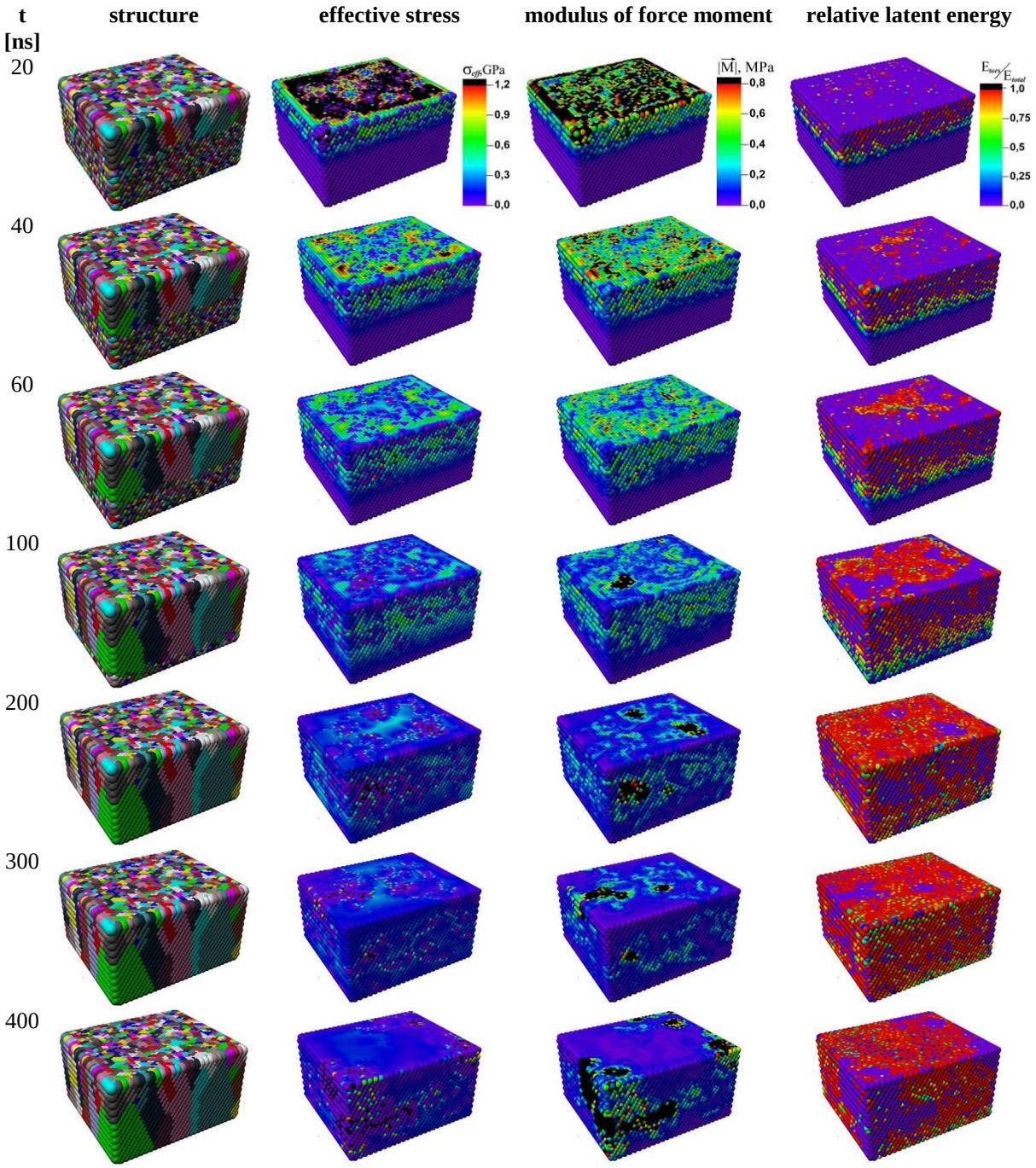


Fig. 10. Images of grain structure, distribution of effective stress, modulus of force moment and relative latent mechanical energy in specimen **without pores and with high TEC** at different moments of time.

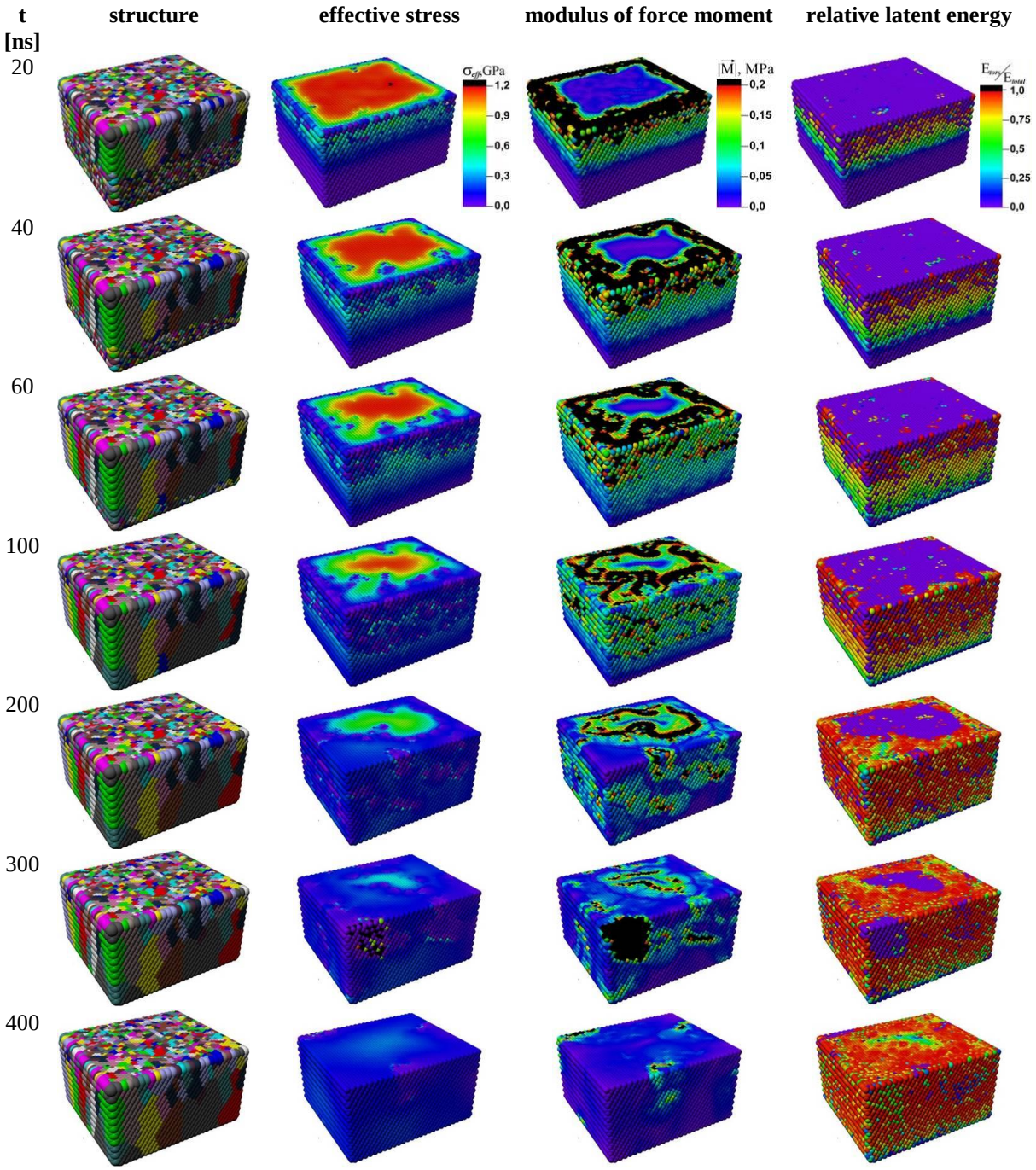


Fig. 11. Images of grain structure, distribution of effective stress, modulus of force moment and relative latent mechanical energy in specimen **without pores and with low TEC** at different moments of time.

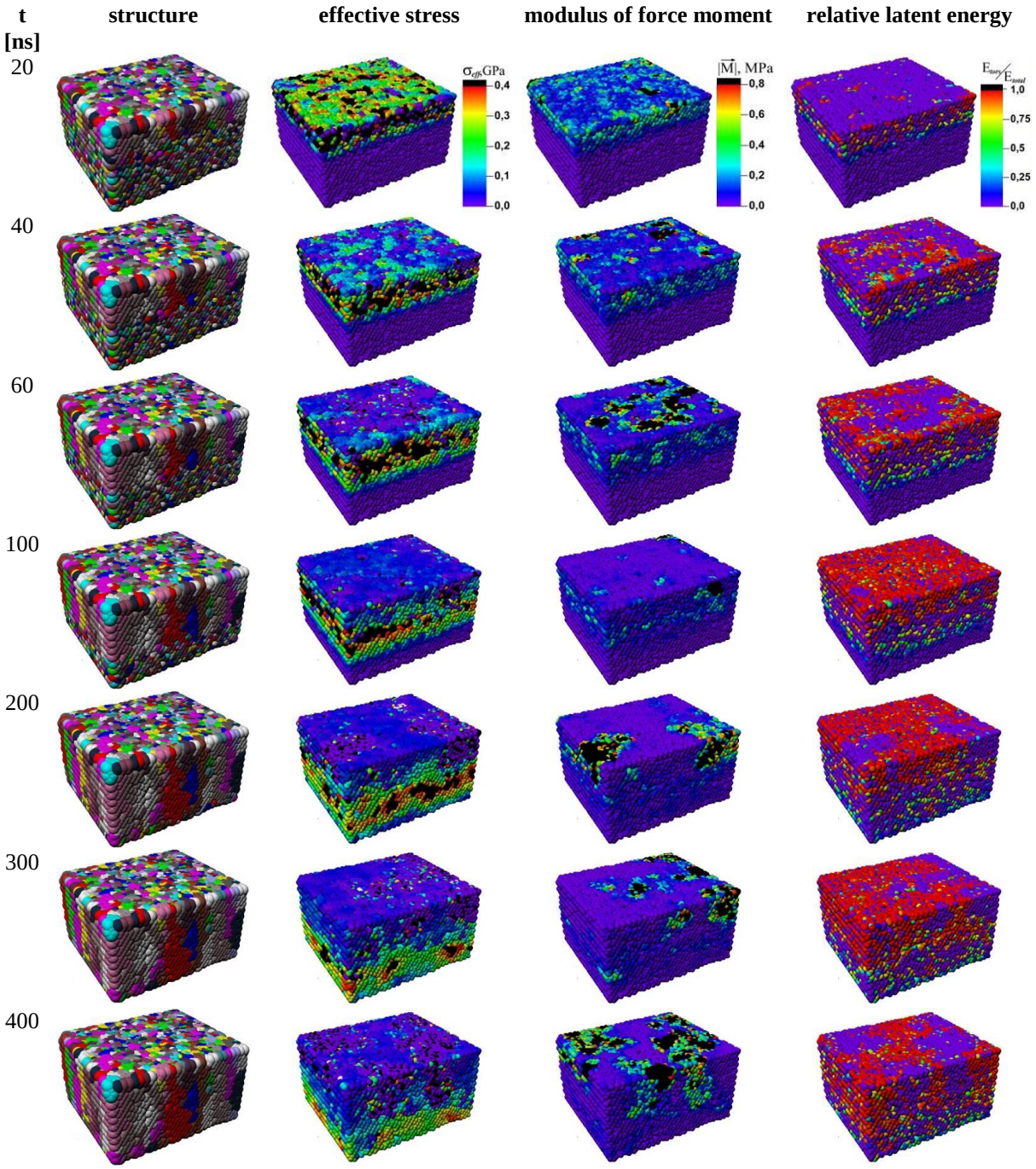


Fig. 12. Images of grain structure, distribution of effective stress, modulus of force moment and relative latent mechanical energy in **porous specimen with high TEC** at different moments of time.

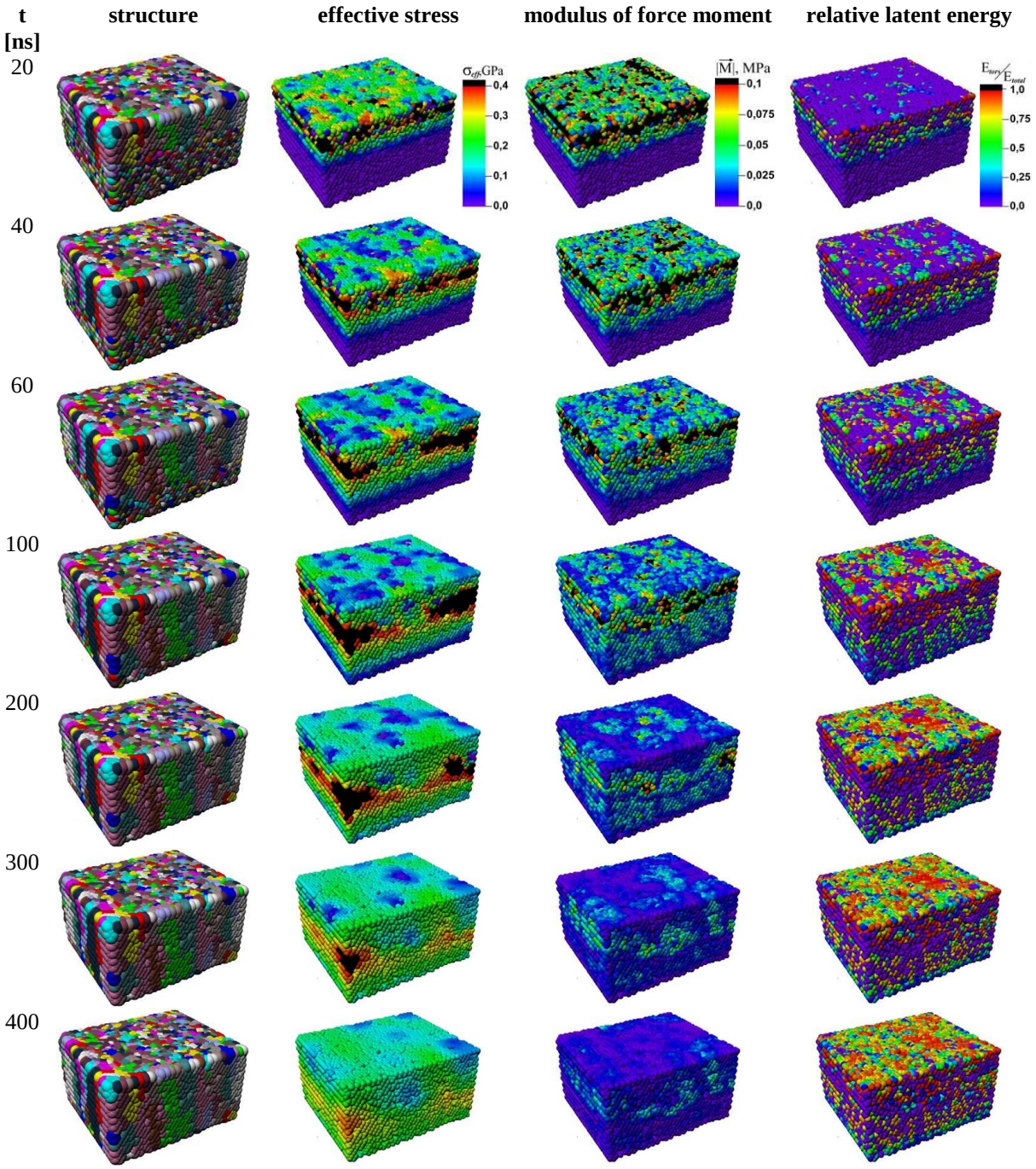


Fig. 13. Images of grain structure, distribution of effective stress, modulus of force moment and relative latent mechanical energy in **porous specimen with low TEC** at different moments of time.

Comparison of the dynamics of defect structure accumulation (Figs. 10–11) and the analysis of plots of relative specific torsion energies (Fig. 8) allows concluding that in nonporous nanocrystalline material, with the decrease of TEC, the rate of defect structure accumulation increases. This fact is explained by the fact that for high values of thermal expansion coefficient, material has higher possibility to relax thermal stresses via three-dimensional deformation without disturbing the translational invariability of the crystal lattice, but increasing the interatomic distance of the latter. In the case of low TEC, the major amount of thermal

stresses transit into local force moments, initiating curvature and torsion of crystal lattice, which unavoidably is accompanied by origination of defect structures at microscale level.

This makes particularly indicative the plots of time dependencies of total modulus of force moment in Fig. 9. Evidently, averaging the modulus of force moment along whole specimen leads at mesoscale level to considerable oscillations of force moment in specimens with high TEC, which testifies elastic rotations of whole conglomerates of grains, which, nevertheless, does not lead to appreciable formation of defect structures in them. The plots of time dependencies of force moments in specimens with low TEC, conversely, are smooth and have 2–4 times lower values.

The comparison of images of defect structure energy distribution both in porous and nonporous specimens shows its more dense and uniform distribution in material with low TEC, while in material with high values of thermal expansion coefficient, the torsion energy is distributed as large areas of its high and low concentration (Figs. 10–13).

This confirms the idea about material with low thermal expansion coefficient achieving larger curvature of crystal lattice at microscale, accompanied by bigger rate of defect structure formation. Such systems rapidly drop stresses into moment forces at microscale impeding the self-organization of rotation moment in large conglomerates of grains.

4 CONCLUSIONS

A new modification of the multiscale hybrid discrete continuum approach of stochastic excitable cellular automata was developed. The new method explicitly accounts for the porosity and nanocrystalline structure of material, the algorithm for computing local force moments and angular velocities of microrotations occurring in structurally nonuniform medium with due account for energy dissipation.

Using the direct numerical experiment it was demonstrated that in porous specimens the rate of collective crystallization was considerably lower than that in nanocrystalline specimens. The study results demonstrate that the value of thermal expansion coefficient has no impact on the rate of recrystallization in nonporous specimens.

On the other hand, the value of thermal expansion coefficient of nanoporous and nanocrystalline specimens significantly impacts the distribution character and rate of defect structure accumulation. For instance, it was shown that the material with low thermal expansion coefficient reaches higher curvature of crystal lattice at microscale. The formation of curvature and torsion in the lattice is accompanied by the formation of defect structures. Such systems are more capable of relaxing moment stresses at microscale, which interferes the self-organization of rotation moment in conglomerates of grains at mesoscale.

Thus, in addition to accounting the local rotational modes of structural elements and thermal radiation by pore walls, it is necessary to study the ability of material to reversible structural transformations in the material of pore walls. The reversible structural transformations will accumulate peak values of energy in elastic rotational stress fields that are knowingly inertial, which will allow preventing catastrophic growth of main cracks.

ACKNOWLEDGEMENTS

The study was conducted under support of fundamental Research Program of Russian State Academies of Sciences for 2013-2020, RFBR projects Nos. 14-01-00789, 14-01-92005 and 15-08-05818 as well as grant of the President of the Russian Federation for support of leading scientific schools No. NSh-2817.2014.1.

REFERENCES

[1] M. Radovic, M. Barsoum, MAX phases: Bridging the gap between metals and ceramics. *American Ceramic Society Bulletin*, **92**, 20-27, 2013

[2] E. Levashov, A. Merzhanov, D. Shtansky, Advanced Technologies, Materials and Coatings Developed in Scientific-Educational Center of SHS. *Galvanotechnik*, **9**, 1-13, 2009

[3] J.P. Lin, L.L. Zhao, G.Y. Li et al., Effect of Nb on oxidation behavior of high Nb containing TiAl alloys. *Intermetallics*, **19**, 131-136, 2001

[4] A. A. Voevodin, J.S. Zabinski, Nanocomposite and nanostructured tribological materials for space applications. *Composites Science and Technology*, **65**, 741–748, 2006

[5] A.V. Kartavykh, S.D. Kaloshkin, V.V. Cherdyntsev et. al., Application of microstructured intermetallides in turbine manufacture. Part 1: Present state and prospects (a review). *Inorganic Materials: Applied Research*, **4**, 12-20, 2013

[6] A. A. Voevodin, J. S. Zabinski, C. Muratore, Recent Advances in Hard, Tough, and Low Friction Nanocomposite Coatings. *Tsinghua Science and Technology*, **10**, 665-679, 2005

[7] D. Shtansky, Ph. Kiryukhantsev-Korneev, A. Sheveyko et. al., Comparative investigation of TiAlC(N), TiCrAlC(N), and CrAlC(N) coatings deposited by sputtering of MAX-phase Ti2−xCrxAlC targets. *Surface & Coatings Technology*, **203**, 3595–3609, 2009

[8] J. Pearson, M. Zikry, K. Wahl, Microstructural modeling of adaptive nanocomposite coatings for durability and wear. *Wear*, **266**, 1003–1012, 2009

[9] F. Appel, J.-D. Heaton, Gamma Titanium Aluminide Alloys: Science and Technology. Wiley-VCH, 2011

[10] G. Ying, X. Wang, Numerical Simulation of the SHS Temperature Fields of Al-Ti-C System Based on Plane Propagating Pattern. *Int. J. Mod. Phys. C*, **20**, 1087, 2009

[11] G. Ying, X. He, S.-Y. Du et al., Kinetics and numerical simulation of self-propagating high-temperature synthesis in Ti–Cr–Al–C systems. *Rare Metals*, **33**, 527-533, 2014

[12] V.E. Panin, Physical Mesomechanics of Heterogeneous Media and Computer-Aided Design of Materials, Camb. Internat. Science Publ., 1998.

[13] V.E. Panin, V.E. Egorushkin, Curvature solitons as generalized structural wave carriers of plastic deformation and fracture. *Phys. Mesomech.*, **16,** 267-286, 2013

[14] Yu.I. Meshcheryakov, T.A. Khantuleva, Nonequilibrium processes in condensed media: Part 1. Experimental studies in light of nonlocal transport theory. *Phys. Mesomech.*, **18**, 228-243, 2015

[15] J.Kroc, Application of cellular automata simulations to modelling of dynamic recrystallization. *Lecture Notes in Computer Science*, **2329,** 773-782, 2002.

[16] A. Godara, D. Raabe, Mesoscale simulation of the kinetics and topology of spherulite growth during crystallization of isotactic polypropylen (iPP) by using a cellular automaton. *Modelling Simulation Materials Science Engineering*, **13**, 733-751, 2005.

[17] F.J. Humphreys, M. Hatherly, *Recrystallization and Related Annealing Phenomena*, Pergamon, New York, 1995.

[18] D.D. Moiseenko, V.E. Panin, P.V. Maksimov, S.V. Panin, F. Berto, Material fragmentation as dissipative process of micro rotation sequence formation: hybrid model of excitable cellular automata. *AIP Conference Proceedings*, **1623,** 427–430, 2014.

[19] D.D. Moiseenko, V.E. Panin, T.F. Elsukova, Role of Local Curvature in Grain Boundary Sliding in a Deformed Polycrystal. *Phys. Mesomech.*, **16**, 335-347 , 2013.

[20] V.E. Panin, V.E. Egorushkin, D.D. Moiseenko et al., Functional role of polycrystal grain boundaries and interfaces in micromechanics of metal ceramic composites under loading. *Computational Materials Science*, **116**, 74-81, 2016

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

NONLINEAR COMPOSITE VOXELS AND FFT-BASED HOMOGENIZATION

Matthias Kabel[1], **Andreas Fink**[1], **Felix Ospald**[2], **and Matti Schneider**[1]

[1]Department of Flow and Material Simulation
Fraunhofer ITWM
e-mail: {matthias.kabel,andreas.fink,matti.schneider}@itwm.fraunhofer.de

[2] Faculty of Mathematics
Chemnitz University of Technology
e-mail: felix.ospald@mathematik.tu-chemnitz.de

Keywords: Composite materials, Homogenization, Lippmann-Schwinger equation, FFT.

Abstract. *The FFT-based homogenization method of Moulinec-Suquet [14] has reached a degree of sophistication and maturity, where it can be applied to microstructures of industrial size and realistic scope. However, for non-linear or load-path-dependent problems the method reaches its limits, in particular if variations of the geometry are considered or the determination of the full material law on the macro-scale is required.*

Time and memory considerations are primarily responsible for these limitations. This work focuses on the composite voxel technique, where sub-voxels are merged into bigger voxels to which an effective material law based on laminates is assigned. Due to the down-sampled grid, both the memory requirements and the computational effort are severely reduced, while retaining the original accuracy. We discuss the extensions of linear elastic ideas [6, 9] to incremental problems at small strains. In contrast to conventional model order reduction methods, our approach does neither rely upon a "offline phase" nor on preselected "modes". We demonstrate our ideas with several numerical experiments, comparing to full-resolution computations heavily relying upon our MPI-parallel implementation FeelMath [1].

1 Introduction

In spite of the superior computational speed of FFT-based homogenization [13], direct nonlinear elastic computations [14, 11, 3, 8] on state-of-the-art CT images consisting of appr. 8000^3 voxels [2] cannot be considered as currently technically feasible. Therefore, the size of the structure has to be limited to a reasonable value by algorithmic improvements, such as the composite voxel technique [6, 9, 15].

In this paper, we first extend the composite voxel technique to physically nonlinear problems by affine linear approximation of the material laws in Section 2. Then, in Section 3 we apply the laminate and the Voigt mixing rule to a coarsely resolved long fiber reinforced plastic (LFT) microstructure (Fig. 2), compare the effective behavior to full resolution results and asses the possible speedup for linear and nonlinear simulations. Finally, in in Section 4 this algorithmic speedup is compared to the speedup due to an MPI parallelization

2 Composite voxels

The numerical execution of the so-called basic scheme of FFT-based homogenization [13] requires the evaluation of the stress tensor $\sigma(\epsilon, \epsilon_{\text{old}}, \ldots)$ in every voxel of the unit cell. If one is working directly on a CT image or, more generally, a voxelized geometry where every voxel is occupied by precisely one phase, the corresponding material law is applied. However, sometimes CT images are too large, or results are needed quickly, so that working on a cruder resolution seems inevitable. Then, the question arises how voxels containing more than one phase, so-called composite voxels, should be treated. In a preliminary work [9] that covered linearized elasticity the accuracy of various mixing rules were compared. In the work at hand we will concentrate on the Voigt and the laminate mixing rules.

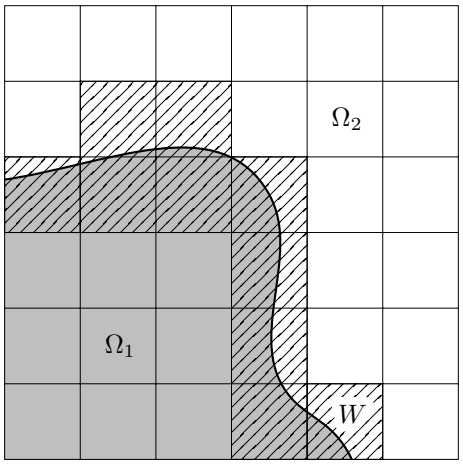


Figure 1: Voxelized geometry with composite voxels.

Before we go to the nonlinear case, we first recall these mixing rules. The Voigt mixing rule for a composite voxel W (see Fig. 1) reads $\mathbb{C}_W^{\text{Voigt}} = \langle \mathbb{C} \rangle$, where $\langle \cdot \rangle$ denotes the spatial average according. The Voigt average $\mathbb{C}_W^{\text{Voigt}}$ constitutes an upper bound on the effective elastic tensor. For a composite voxel W containing precisely two constituents the laminate mixing rule assumes that the interface between the phases is approximately linear with normal vector n. Inspired by the formula for a general laminate, cf. Milton [12, Section 9.5], the elastic tensor of the laminate mixing rule is obtained by solving the system

$$\left(\mathbb{P} + \lambda(\mathbb{C}_W^{\text{laminate}} - \lambda\,\mathrm{Id})^{-1}\right)^{-1} = \left\langle \left(\mathbb{P} + \lambda(\mathbb{C} - \lambda\,\mathrm{Id})^{-1}\right)^{-1} \right\rangle,$$

where the four-tensor

$$\mathbb{P}_{ijlm} = \frac{1}{2}\left(n_i\delta_{jl}n_m + n_i\delta_{jm}n_l + n_j\delta_{il}n_m + n_j\delta_{im}n_l\right) - n_in_jn_ln_m$$

encodes the interface direction n and $\lambda > 0$ is chosen sufficiently large. More precisely, λ should be larger than the largest eigenvalue of the stiffness matrices $\mathbb{C}(x)$ for all x.

We extend these mixing rules to the physical nonlinear case in three steps

1. Linearization: In every loading step the (nonlinear) material laws are linearized at the deformed state ϵ^n of the composite voxel

$$\mathbb{C}(x) = \frac{\partial \sigma}{\partial \epsilon}(x, \epsilon^n, \epsilon^{n-1}, \ldots)$$

2. Mixing: The effective linearized stiffness $\mathbb{C}_W$ is calculated according to the consistent mixing rules for the linear elastic case. The nonlinear behavior of the composite voxel is approximated by the affine linear function

$$\sigma(\epsilon^{n+1}, \epsilon^n, \ldots) = \sigma^n + \mathbb{C}_W : (\epsilon^{n+1} - \epsilon^n)$$

3. Internal variables: The internal variables of the constituents are updated due to the deformation ϵ^{n+1} of the composite voxel.

3 Long Fiber Reinforced Thermoplastic (LFT)

Long-fiber reinforced thermoplastics (LFT) bridges the price-to-performance gap between short fiber thermoplastic (SFT) materials and continuous fiber composite materials while still being processable via efficient injection or compression molding methods [18, 16, 7]. The investigated material, produced by compression molding, features an extremely high maximum fiber length of approx. 50 mm and a corresponding aspect ratio of approximately 3000. To incorporate the complete fiber length spectrum and thereby conserve the scale separation, the size of the representative volume elements of the LFT structure would be extremely large and outside the limits of computational resources, if equal edge lengths of the RVE would be considered, as pointed out in the following: Glass fibers have a diameter ranging between 5 μm and 24 μm. This enforces a minimal resolution of 5 μm. Considering a value of 50 mm for the maximum fiber length, the RVE needs to be discretized with 10000^3 voxels. To limit the size of the structure to a reasonable value, the maximum fiber length is therefore exclusively implemented in flow direction (the longest RVE edge), whereas the transverse directions are cropped, resulting in differing edge lengths, see Figure 2.

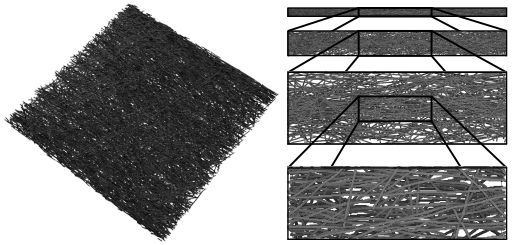


Figure 2: Virtually generated LFT microstructure [5] (left). Part of the LFT microstructure used for simulations (right). Sample size: $40 \times 1.5 \times 0.134$ mm^3. Discretization: $8000 \times 300 \times 27$ voxels.

3.1 Material parameters

The investigated LFT material consists of a polypropylene matrix (DOW® C711-70RNA) and 30 wt% glass fibers (TufRov® 4575). As specified by the manufacturer, in the linear elastic regime the glass fibers are characterized by a Young's modulus of $E_{glass} = 72$ GPa and Poisson's ratio of $\nu_{glass} = 0.2$, whereas for the polypropylene (PP) matrix we have a Young's modulus $E_{PP} = 1.25$ GPa and a Poisson's ratio of $\nu_{PP} = 0.35$.

For uniaxial tension of polypropylene Fig. 3 contains the measured stress-strain curve as determined by Fliegener [4]. The behavior is linear initially, followed by a plastic phase, roughly up to 5% strain. For higher strains, a strain softening effect is observed, which is commonly attributed to fracture or damage effects. However, as the regime of 5% strain and higher usually demands a description by large deformations, we neglect this strain-softening effect, and rely upon an incremental elastic-plastic material law at small strains governed by a J_2-flow theory with isotropic hardening. To approximate the nonlinear curve of Fig. 3 we rely upon a piecewise linear hardening curve. The fitting is reasonably accurate up to a relative strain of 5%.

ϵ^p_{eq} [%]	0	1.33697	2.37197	4.08397
σ_{eq} [MPa]	12	16	18	19

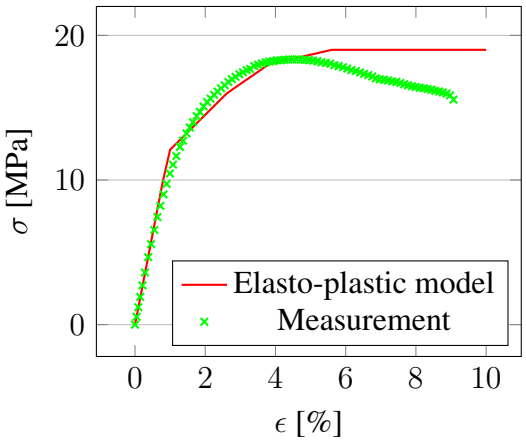


Figure 3: Calibrated hardening parameters for PP (top). Tensile tests of PP and calibrated elasto-plastic model [4] (bottom).

3.2 Numerical results

All simulation were performed on the part of the virtual microstructure shown in Fig. 2 (right) consisting of $8000 \times 300 \times 27$ voxels, which is discretized on a staggered grid [17]. Therefore, the resulting discretized problems have 194.4 million (displacement) degrees of freedom. By downsampling the geometry by a factor of 2 resp. 4 we obtain two alternative discretizations with composite voxels, see Fig. 4. Technical details concerning the downsampling procedure with composite voxels can be found in [9].

In the linear elastic case 6 loading cases were performed which allow to calculate the effective stiffness $\mathbb{C}_{\text{eff}}$ of the LFT. To compare the effective stiffnesses for different numerical methods, the direction-dependent Young's moduli in the fiber plane are evaluated, see Fig. 5. The direction-dependent Young's

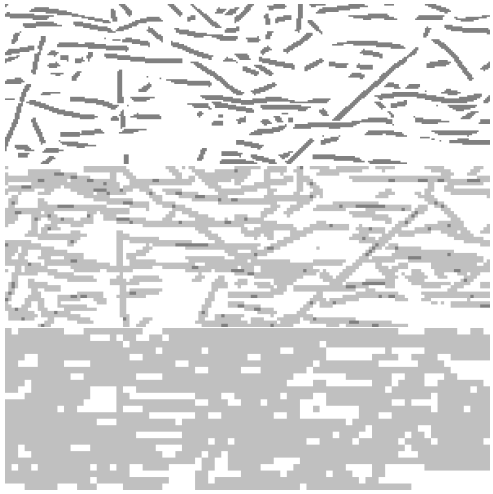


Figure 4: Slice of the LFT microstructure at full resolution (top), half resolution (middle) and quarter resolution (bottom) which contain PP (white), Glass (dark gray) and composite voxels (light gray)

modulus is defined by

$$\frac{1}{E(d)} = (d \otimes d) \cdot \mathbb{C}_{\text{eff}}^{-1} : (d \otimes d)$$

with

$$d = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \\ 0 \end{pmatrix}$$

The relative error of the numerical solutions using composite voxels on downsampled geometries exhibits the following behavior

- Laminate mixing: The effective stiffness is underestimated. At both half and quarter resolution the relative error has its maximum at appr. $40°$. The relative error stays below 7% and 24% respectively. Orthogonal to the main fiber direction the relative error is below 1% for both resolutions. In main fiber direction the relative error is appr. 4% and 17% respectively

- Voigt mixing: The effective stiffness is overestimated. At both half and quarter resolution the relative error has its maximum at appr. $10°$. The relative error stays below 8% and 20% respectively. Orthogonal to the main fiber direction the relative error is below 3% and 8% respectively. In main fiber direction the relative error is appr. 8% and 19% respectively

In light of to the fluctuations of the three measurements in $0°$ and $90°$-direction [4] it is reasonable to perform linear elastic simulations on the half resolution even if the original geometry resolves the fibers only with 3 voxels over the diameter.

Due to the results for the linear elastic case we decided to perform the nonlinear simulations only at full and half resolutions. In Fig. 5 we compare the results of simulated uniaxial tensile tests with measurements and investigate the evolution of the relative error for increasing loading. At the beginning both mixing rules start with a relative error slightly below 20%. While the relative error for the Voigt mixing rapidly increases and reaches 37% at the end of the loading, the error decreases monotonously for laminate mixing. Taking into account the variance of the experimental results for the uniaxial tensile tests in main fiber direction, the error due to downsampling by a factor of 2 is acceptable if laminate composite voxels are used.

3.3 Computational effort

After assessing the quality of the numerical solution we study the speedup on for downsampled geometries with composite voxels. For the linear elastic case we apply the conjugate gradient method (CG) to accelerate the convergence speed of FFT-based homogenization [19]. Since composite voxels lead to a smoothing of the material contrast at the interfaces, one could expect less iterations for the problems on the downsampled geometries. As observed by Kabel et al. [9] this is not the case, see Tab. 1. For linear elastic problems composite voxels do not influence the convergence speed. According to Tab. 3 using half resolution with composite voxels reduces

Resolution	Mixing rule	# Iterations
Full	-	1049
Half	Laminate	1007
	Voigt	1054
Quarter	Laminate	954
	Voigt	1076

Table 1: Number of iterations for 6 loadcases.

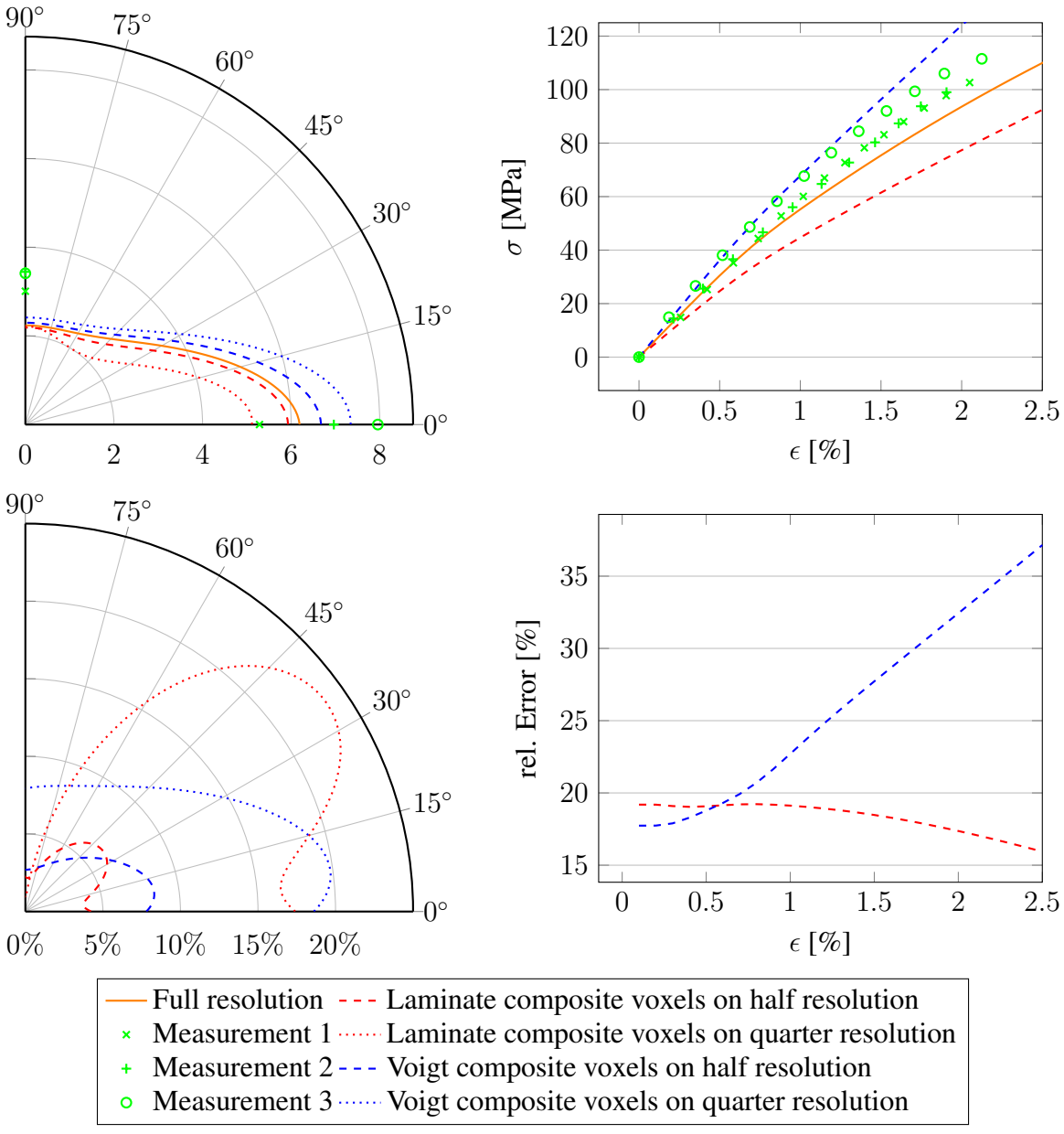


Figure 5: Young's modulus body [GPa] in x-y-plane (top left) and relative error (bottom left). Simulated tensile test in main fiber direction (top right) and relative error of the results with using composite voxels (bottom right).

Resolution	Laminate Total [s]	Laminate FFT [s]	Laminate Material [s]	Voigt Total [s]	Voigt FFT [s]	Voigt Material [s]
Full	8423	6706	1643	8423	6706	1643
Half	4691	2316	2316	3540	2421	1080
Quarter	3511	1034	2433	1974	1165	787
Full	59452	17593	38519	59452	17593	38519
Half	8416	682	7409	6493	699	5568

Table 3: Computational effort of 10 cores for composite voxels on half (24.3 million DOFs) and quarter (3.0375 million DOFs) resolution in comparison to the effort of full-resolution (194.4 million DOFs) computations. Linear elasticity (top). Von Mises plasticity (bottom)

the computational time by 45% for laminate mixing and 58% for Voigt mixing compared to full resolution simulations.

Since the volume fraction of the composite voxels increases dramatically at quarter resolution, see Tab. 2, and the evaluation of the stress response for composite voxels is computationally more expensive than evaluating linear elastic material laws, no further significant speedup can be obtained on coarser geometries. To be more precise, using quarter resolution with laminate composite voxels reduces the runtime by 58% and is approximately as fast as Voigt mixing at half resolution. Quarter resolution with Voigt mixing reduces the runtime by 76%.

Resolution	PP	Glass	Composite
Full	86.77	13.23	0.00
Half	57.78	2.61	39.61
Quarter	27.51	0.00	72.49

Table 2: Volume fractions [vol%] for the LFT microstructure

The physically nonlinear case we use the basic scheme developed by Moulinec-Suquet [14]. In contrast to the linear case, the number of iterations needed for convergence is significantly reduced by using composite voxels, see Fig. 6. This effect is due to the linearization of the composite voxels. As a side effect also the repeated application of the material law in the composite voxels is less expensive than applying the original nonlinear material law, see Tab. 3. At half resolution using laminate mixing decreases the total runtime by 85 % and using Voigt mixing even decreases it by 90 %.

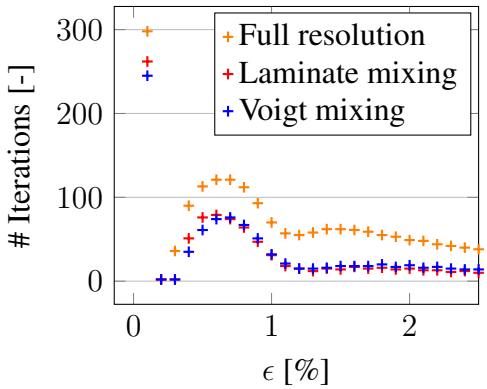


Figure 6: Number of iterations needed for convergence for the tensile tests in main fiber direction at full and half resolution.

4 Parallel efficiency

If highly accurate results are needed it is possible to accelerate the simulations by parallelization of the algorithm. In this section we will study the possible speedup for an MPI parallelization of the basic scheme. Since all operations of the basic scheme except for the fast Fourier transform (FFT) are local operations, the parallel efficiency depends crucially on the implementation of FFT. For computing the discrete Fourier transform we are using the C

Cores	Total [s]	Efficiency [%]	FFT [s]	Efficiency [%]	Material [s]	Efficiency [%]
10	8423	100	6706	100	1643	100
20	4243	99	3332	101	871	94
40	2653	79	2134	79	494	83
80	1485	71	1214	69	242	85
160	876	60	727	58	121	84
320	584	45	455	46	65	78
10	59451	100	17593	100	38519	100
20	32332	92	12351	71	18408	105
40	14785	100	4876	90	9129	105
80	7481	99	4609	48	4609	104
160	3931	95	2307	48	2307	104
320	2280	81	1124	49	1124	107

Table 4: Parallel efficiency for full-resolution (194.4 million DOFs) simulations. Linear elasticity (top). Von Mises plasticity (bottom).

(subroutine) library FFTW[1].

All simulations have been performed on the Beehive cluster at Fraunhofer ITWM, which consists of 192 nodes (Dell Power Edge M620) each having a dual Intel Xeon E5-2670 (”Sandy Bridge”) processor, i.e 16 CPU Cores per node. The nodes are connected by FDR Infiniband.

4.1 Strong scaling

In the linear elastic case FFT runtime dominates the numerical effort, see Tab. 4 and Fig. 7. Therefore, also the parallel effiency of the FFT-based scheme almost coincides with the parallel efficiency of FFT. In comparison to composite voxels the runtime on 2 nodes is similar to solving the problem at half resolution on 1 node and the runtime on 4 nodes is similar to that on 1 node at quarter resolution.

In the physically nonlinear case the material law dominates the numerical effort. Since the parallel efficiency of applying the material law is perfect in our example, the total parallel efficiency is greater than 82 % for up to 32 nodes and 320 MPI processes, see Tab. 4. Nonetheless using the composite voxel technique with half resolution, 8 times more nodes are required to have comparable runtimes.

5 Conclusion

Previous studies on the composite voxel technique always assumed highly resolved objects. In many application it is currently not possible to obtain CT-images (or virtually generated microstructures) that are representative and additionally resolve all objects well. For LFT it is a rule of thumb to make CT-Images with a resolution that correspond to 3 - 4 voxels over the diameter of the fibers. Therefore, at first glance the application of the composite voxels technique to LFT microstructures seemed to be questionable. Nonetheless it tuned out, that for both linear and nonlinear material modelling of LFT, simulations at half resolution with laminate composite voxels give reasonable results from an engineering point of view and lead to a significant speedup.

[1] www.fftw.org

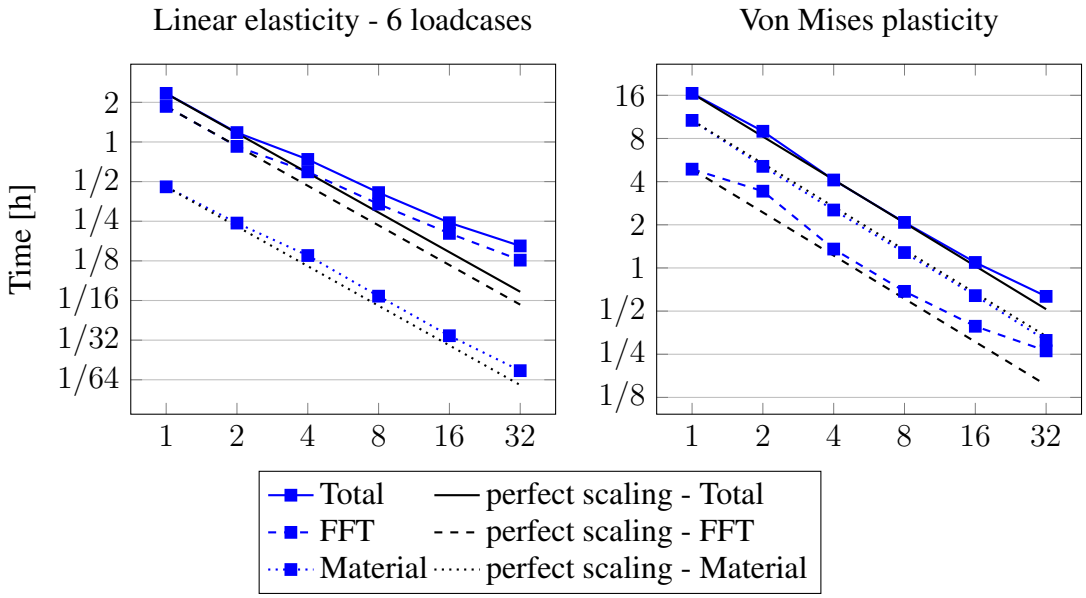


Figure 7: Strong scaling for 10 MPI processes per node.

Concerning the relative error of the effective linear elastic behaviour, the laminate mixing rule consistently outperforms the Voigt mixing rule at half resolution but fails at quarter resolution when the interface directions calculated in the downsampling process are no longer meaningful (the complete cross section of a fiber can be contained in a composite voxel). More strikingly, the relative error of the effective stress-strain curve considered as a function of the applied load decreases when using composite voxels with laminate mixing whereas it grows linearly for Voigt mixing. Thus, working on a coarser resolution for physically nonlinear problems is only acceptable in combination with laminate mixing.

To further increase the precision of laminate mixing in the physically nonlinear case we will consider direct laminate mixing of nonlinear material laws in a future work [10].

Acknowledgements

While writing this article, Felix Ospald received financial support from the German Research Foundation (DFG), Federal Cluster of Excellence EXC 1075 “MERGE Technologies for Multifunctional Lightweight Structures”.

REFERENCES

[1] FeelMath. Fraunhofer ITWM `http://www.itwm.fraunhofer.de/feelmath`. Accessed: 2015-11-20.

[2] High-Resolution Computed Tomography flyer. NEXT ENERGY EWE-Forschungszentrum für Energietechnologien e.V. Accessed 12.02.2016, `http://www.next-energy.de/tl_files/pdf/NE_Flyer_CT_2014_EN.pdf`.

[3] P. Eisenlohr, M. Diehl, R.A. Lebensohn, and F. Roters. A spectral method solution to crystal elasto-viscoplasticity at finite strains. *International Journal of Plasticity*, 46(0):37–53, 2013.

[4] S. Fliegener. *Micromechanical finite element modeling of long fiber reinforced thermoplastics*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2015.

[5] S. Fliegener, M. Luke, and P. Gumbsch. 3d microstructure modeling of long fiber reinforced thermoplastics. *Composites Science and Technology*, 104(0):136 – 145, 2014.

[6] L. Gélébart and F. Ouaki. Filtering material properties to improve fft-based methods for numerical homogenization. *J. Comput. Phys.*, 294(C):90–95, August 2015.

[7] Henning F, Ernst H, Brüssel R, Geiger O, Krause W. LFTs for automotive applications. *Reinforced Plastics*, 49:24–33, 2005.

[8] M. Kabel, T. Böhlke, and M. Schneider. Efficient fixed point and Newton-Krylov solvers for FFT-based homogenization of elasticity at large deformations. *Computational Mechanics*, 54(6):1497–1514, 2014.

[9] M. Kabel, D. Merkert, and M. Schneider. Use of composite voxels in FFT-based homogenization. *Computer Methods in Applied Mechanics and Engineering*, 294(0):168–188, 2015.

[10] M. Kabel and M. Schneider. Physically nonlinear composite voxel. In preparation.

[11] N. Lahellec, J.C. Michel, H. Moulinec, and P. Suquet. Analysis of inhomogeneous materials at large strains using fast fourier transforms. In Christian Miehe, editor, *IUTAM Symposium on Computational Mechanics of Solid Materials at Large Strains*, volume 108 of *Solid Mechanics and Its Applications*, pages 247–258. Springer Netherlands, 2003.

[12] G. W. Milton. *The Theory of Composites*. Cambridge University Press, 2002. Cambridge Books Online.

[13] H. Moulinec and P. Suquet. A fast numerical method for computing the linear and nonlinear mechanical properties of composites. *Comptes rendus de l'Académie des sciences. Série II, Mécanique, physique, chimie, astronomie*, 318(11):1417–1423, 1994.

[14] H. Moulinec and P. Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering*, 157(1-2):69–94, 1998.

[15] F. Ospald, M. Schneider, and M. Kabel. Hyperelastic laminates, composite voxels and FFT-based homogenization at finite strains. *Computer Methods in Applied Mechanics and Engineering*, submitted.

[16] C.W. Peterson, G. Ehnert, R. Liebold, and R. Kühfusz. Compression molding. In D.B. Miracle and S.L. Donaldson, editors, *Composites*, volume 21 of *ASM Handbook*, pages 516–535. Springer Netherlands, 2001.

[17] M. Schneider, F. Ospald, and M. Kabel. Computational homogenization of elasticity on a staggered grid. *International Journal for Numerical Methods in Engineering*, 105(9):693–720, 2016.

[18] R.H. Todd, D.K. Allen, and L. Alting. *Manufacturing Processes Reference Guide*. Industrial Press, 1994.

[19] J. Zeman, J. Vondřejc, J. Novák, and I. Marek. Accelerating a FFT-based solver for numerical homogenization of periodic media by conjugate gradients. *Journal of Computational Physics*, 229(21):8065 – 8071, 2010.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

A THERMODYNAMICALLY CONSISTENT FINITE STRAIN MICRO-SPHERE FRAMEWORK FOR PHASE-TRANSFORMATIONS

R. Ostwald[1], **T. Bartel**[1], **and A. Menzel**[1,2]

[1]Institute of Mechanics, TU Dortmund University
Leonhard-Euler-Str. 5, D-44227 Dortmund, Germany
e-mail: richard.ostwald@udo.edu / thorsten.bartel@udo.edu / andreas.menzel@udo.edu

[2] Division of Solid Mechanics, Lund University
P.O. Box 118, SE-22100 Lund, Sweden
e-mail: andreas.menzel@solid.lth.se

Keywords: Finite strain, Micro-sphere, Phase-transformation

Abstract. *We extend a thermodynamically consistent finite strain micro-sphere framework elaborated by Carol et al. towards the modelling of phase-transformations to allow for the simulation of polycrystalline solids such as, e.g., shape memory alloys and shape memory polymers undergoing large deformations. The considered phase-transformation mechanism is based on statistical physics and allows the consideration of an arbitrary number of solid material phases. The specifically constructed, non-quadratic Helmholtz free energy functions considered in every micro-plane of the micro-sphere framework are extended to include individual Bain-type transformation strains for each of the phases. The total strains acting in each material phase are multiplicatively decomposed into elastic strains and transformation strains.*

1 INTRODUCTION

The physical mechanisms and effects accompanied by phase-transformations in solids present a large potential for industrial applications. For example, NiTi-based shape memory alloys (SMA) are used in medical applications [1], aerospace technology [2] and other industrial fields, cf., e.g., [3, 4].

During tensile deformation of such material—after exceeding the purely elastic regime of the, e.g. austenitic, parent phase—the onset of phase-transformations is accompanied by a plateau-type stress-strain relation, finally followed by a sharp rise in the specimen stress. The latter indicates the completion of the transformation process, resulting again in an elastic behaviour of the now completely transformed material. Considering a polycrystalline material at the meso-scale, however, it is important to note that an initially purely austenitic crystal will not transform to $100\,\%$ martensite. Unfavourably oriented grains within the polycrystalline arrangement might not transform at all—furthermore, experimental observations suggest that even favourably oriented grains usually only transform to a certain extent as elaborated in, e.g., [5].

This so-called grain locking effect within a crystal is related to the sequential transformation behaviour of individual grains. In other words, the initial onset of transformation in the most favourably oriented grain is accompanied by a change of the elastic stress state in its surrounding area, i.e. in its neighboured grains. This change of stress state can hinder transformation in a neighboured grain, especially if a formerly favourable stress state within a particular neighbouring grain turns into an unfavourable stress state.

Taking into account complex interactions at the micro-scale—such as specific microstructure arrangements and twin formations—leads to reliable micromechanical material models, such as the ones presented in, e.g., [6, 7, 8, 9, 10, 11, 12, 13, 14]. However, a possible disadvantage of such precise modelling approaches might be the high computational costs that usually accompany the detailed capturing of microstructural material effects. On the other hand, purely phenomenological models facilitate the solution of complex macroscopic initial boundary value problems (IBVPs) via finite element simulations. Typical phenomenological approaches such as the ones presented in [15, 16, 17], are usually derived within thermodynamical frameworks. Besides the application of the first and second law of thermodynamics, generalised irreversible forces and fluxes, cf. [18] amongst others, are considered in view of a consistent derivation of evolution equations for the inelastic constitutive variables. For the application of phenomenological models in view of efficient macro-scale simulations, the underlying specific modelling parameters need to be fitted to experimentally observed material behaviour for different, representative loading paths [19, 20, 21]. Such phase-transformation models are usually based on classic plasticity-type approaches regarding the onset of the transformation process. While early models were formulated in a one-dimensional setting with restriction to states of tensile stress [22], more sophisticated frameworks are established in, e.g., [16, 15, 23]. Constitutive frameworks that focus on the simulation of single crystals are provided in [24, 25, 26] amongst others.

Another class of thermodynamical models makes use of statistical considerations, capturing transformation probabilities that are derived based on statistical physics. In this context, energy barriers related to the Gibbs potentials of the individual phases need to be determined for the computation of the desired transformation probabilities [27, 28]. Such models have recently been implemented in small strain affine and non-affine micro-sphere frameworks, [29, 30] and [31, 32], respectively. The goal of this contribution is to provide an ansatz for a finite

strain micro-sphere phase-transformation model based on statistical physics, where we use a thermodynamically consistent microsphere framework established in [33] as a basis. For the sake of conceptual clarity, we restrict the subsequent formulation to two phases—one austenitic parent phase and a single martensite variant.

2 MICRO-PLANE CONSTITUTIVE FORMULATION

This section presents a finite strain generalisation of previous phase-transformation models by the authors, cf. [29, 30]. To this end, we consider a finite strain micro-plane constitutive formulation for Neo-Hookean-type elasticity introduced in [33] as a basis, which is extended towards phase-transformations. The micro-plane constitutive relation for phase-transformations is then embedded into a corresponding micro-sphere framework in view of the solution of three-dimensional initial boundary value problems.

2.1 DEFINITION OF VOLUME FRACTIONS

The volume fraction ξ_N^α associated to a particular phase $\alpha \in \{\mathrm{A}, \mathrm{M}\}$, where A represents austenite and M represents martensite, associated to the micro-plane with orientation N is defined as

$$\xi_N^\alpha := \lim_{v_N \to 0} \left(\frac{v_N^\alpha}{v_N} \right) \tag{1}$$

and must at any time obey the restrictions

$$\xi_N^\alpha \in \ [0,1] \subset \mathbb{R} \quad , \quad \sum_\alpha \xi_N^\alpha = 1 \quad , \quad \sum_\alpha \dot{\xi}_N^\alpha = 0 \quad , \tag{2}$$

as both the austenite and the martensite phase are assumed to possess identical time-independent referential mass densities $\rho^\alpha = \rho_0 = \mathrm{const}$.

2.2 MICRO-PLANE STRAIN MEASURE

Generally speaking, different micro-plane strain measures can be considered within a micro-sphere framework. However, in view of capturing a macroscopic Neo-Hookean type material response for the individual phases—and in line with [33]—we restrict the formulation to a certain normal strain measure. Specifically, we characterise the normal strain acting on the micro-plane as

$$\lambda_N = ||\boldsymbol{F} \cdot \boldsymbol{N}|| = \sqrt{\boldsymbol{N} \cdot \boldsymbol{C} \cdot \boldsymbol{N}} \tag{3}$$

with $\boldsymbol{F}$ representing the deformation gradient and $\boldsymbol{C} = \boldsymbol{F}^{\mathrm{t}} \cdot \boldsymbol{F}$ being the right Cauchy-Green deformation tensor. Note that the restriction to this strain measure induces a macroscopic Poisson's ratio of $\nu_{\mathrm{P}} = 0.25$.

2.3 MULTIPLICATIVE MICRO-STRAIN DECOMPOSITION

For the consideration of Bain-type transformation strains associated to the martensite phase, the total micro-plane strain measure (3) is multiplicatively decomposed into an elastic strain contribution λ_N^α and a transformation related contribution $\lambda_{\mathrm{tr}}^\alpha$ via

$$\lambda_N = \lambda_N^\alpha \, \lambda_{\mathrm{tr}}^\alpha \, , \tag{4}$$

so that the elastic strain contribution λ_N^α that enters the free energy potentials of the respective material phase α takes the form

$$\lambda_N^\alpha = \lambda_N \, [\lambda_{\mathrm{tr}}^\alpha]^{-1} \tag{5}$$

with the total strain measure λ_N defined in (3) and with an individual material constant $\lambda_{\rm tr}^{\alpha}$ for both austenite and martensite.

2.4 MICRO-PLANE HELMHOLTZ FREE ENERGY POTENTIALS

For each phase α situated within a micro-plane with orientation N we consider a micro-plane Helmholtz free energy $\psi_N^{\alpha} = \widehat{\psi}_N^{\alpha}(\lambda_N)$ of the form

$$\widehat{\psi}_N^{\alpha}(\lambda_N) = \mathsf{E}^{\alpha} \left[\frac{\left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right]^2}{2} + \frac{\left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right]^{-3}}{3} - \frac{5}{6} \right] , \tag{6}$$

with E^{α} denoting the scalar-valued micro-plane elasticity coefficent of phase α, $\lambda_{\rm tr}^{\alpha}$ being the corresponding transformation strain, and λ_N representing the micro-plane strain measure defined in (3). Note that this format of the Helmholtz free energy represents a natural extension of the energy term proposed in [33] towards the consideration of transformation strains $\lambda_{\rm tr}^{\alpha}$ within the individual phases α. For the transformation strain in the austenitic parent phase we have $\lambda_{\rm tr}^{\rm A} = [\lambda_{\rm tr}^{\rm A}]^{-1} = 1$, thus for austenite the Helmholtz free energy potential (6) simplifies to

$$\widehat{\psi}_N^{\rm A}(\lambda_N) = \mathsf{E}^{\rm A} \left[\frac{\lambda_N^2}{2} + \frac{\lambda_N^{-3}}{3} - \frac{5}{6} \right] , \tag{7}$$

which is directly related to the format proposed in [33] for the compressible extension of a micro-sphere model with vanishing initial microstresses.

Accordingly, the material micro-plane stress $S_N^{\alpha} = \widehat{S}^{\alpha}(\lambda_N)$ corresponding to each phase α takes the form

$$\begin{aligned} \widehat{S}_N^{\alpha}(\lambda_N) &= \frac{\partial \widehat{\psi}_N^{\alpha}(\lambda_N)}{\partial \lambda_N} && (8) \\ &= \mathsf{E}^{\alpha} \, \left[\lambda_{\rm tr}^{\alpha}\right]^{-1} \left[\left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right] - \left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right]^{-4} \right] . && (9) \end{aligned}$$

The overall Helmholtz free energy $\Psi_N = \widehat{\Psi}_N(\lambda_N)$ of the considered multi-phase mixture associated to the micro-plane with spatial orientation N is obtained from the contributions of the respective constituents, i.e.

$$\begin{aligned} \widehat{\Psi}_N(\lambda_N) &= \sum_{\alpha} \xi_N^{\alpha} \, \widehat{\psi}_N^{\alpha}(\lambda_N) && (10) \\ &= \sum_{\alpha} \xi_N^{\alpha} \, \mathsf{E}^{\alpha} \left[\frac{\left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right]^2}{2} + \frac{\left[\lambda_N \left[\lambda_{\rm tr}^{\alpha}\right]^{-1}\right]^{-3}}{3} - \frac{5}{6} \right] , && (11) \end{aligned}$$

where ξ_N^{α} is the volume fraction of phase α as defined in (1). The resulting overall material micro-plane stress measure $S_N = \widehat{S}_N(\lambda_N)$ of the mixture follows from the combination of (8)

and (10), namely

$$\widehat{S}_N(\lambda_N) = \frac{\partial \widehat{\Psi}_N(\lambda_N)}{\partial \lambda_N} \tag{12}$$

$$= \frac{\partial}{\partial \lambda_N}\left[\sum_\alpha \xi_N^\alpha\, \widehat{\psi}_N^\alpha(\lambda_N)\right] \tag{13}$$

$$= \sum_\alpha \frac{\partial}{\partial \lambda_N}\left[\xi_N^\alpha\, \widehat{\psi}_N^\alpha(\lambda_N)\right] \tag{14}$$

$$= \sum_\alpha \xi_N^\alpha\, \frac{\partial \widehat{\psi}_N^\alpha(\lambda_N)}{\partial \lambda_N} \tag{15}$$

$$= \sum_\alpha \xi_N^\alpha\, \widehat{S}_N^\alpha(\lambda_N) \tag{16}$$

$$= \sum_\alpha \xi_N^\alpha\, \mathsf{E}^\alpha\, [\lambda_{\mathrm{tr}}^\alpha]^{-1}\left[\left[\lambda_N\,[\lambda_{\mathrm{tr}}^\alpha]^{-1}\right] - \left[\lambda_N\,[\lambda_{\mathrm{tr}}^\alpha]^{-1}\right]^{-4}\right]\,. \tag{17}$$

2.5 MICRO-PLANE GIBBS FREE ENERGY POTENTIALS

The individual micro-plane Helmholtz free energy potentials ψ_N specified in Section 2.4, cf. (6), are transformed to a Gibbs potential $g_N^\alpha = \widehat{g}_N^\alpha(\lambda_N^*)$ using a Legendre transformation. Note that λ_N^* corresponds to a parametrisation of the micro-plane strain space, whereas λ_N reflects the actual, physical micro-plane strain that the considered material mixture is subjected to. With the considered micro-plane strain measure λ_N as the functional variable of the Helmholtz free energy of phase α, the Legendre transformation of the latter energy is accomplished by means of

$$\widehat{g}_N^\alpha(\lambda_N^*) = \widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^*\, \frac{\partial \widehat{\psi}^\alpha(\lambda_N)}{\partial \lambda_N}\bigg|_{\lambda_N} \tag{18}$$

$$= \widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^*\, \widehat{S}_N^\alpha(\lambda_N)\Big|_{\lambda_N} \tag{19}$$

$$= \widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^*\, S_N^\alpha\,, \tag{20}$$

where use has been made of (8). Note that the partial derivative in (18) reflects the material stress contribution corresponding to phase α of the mixture, the evaluation of which according to (9) for the current stretch state λ_N yields the specific stress contribution S_N^α of phase α present in the considered micro-plane.

For the evolution of volume fractions provided in Section 2.6, the Gibbs potential (25) that is parameterised in terms of a fictitious stretch λ_N^* takes the specific form

$$\widehat{g}_N^\alpha(\lambda_N^*) = \widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^*\, S_N^\alpha \tag{21}$$

$$= \mathsf{E}^\alpha\left[\frac{\left[\lambda_N^*\,[\lambda_{\mathrm{tr}}^\alpha]^{-1}\right]^2}{2} + \frac{\left[\lambda_N^*\,[\lambda_{\mathrm{tr}}^\alpha]^{-1}\right]^{-3}}{3} - \frac{5}{6}\right] - \lambda_N^*\, S_N^\alpha \tag{22}$$

where use has been made of (6).

In a similar fashion, the overall micro-plane Helmholtz free energy potential Ψ_N of the mixture specified in Section 2.4 is transformed to an overall Gibbs potential $G_N = \widehat{G}_N(\lambda_N^*)$ using

a Legendre transformation, specifically

$$\widehat{G}_N(\lambda_N^*) = \widehat{\Psi}_N(\lambda_N^*) - \lambda_N^* \left. \frac{\partial \widehat{\Psi}_N(\lambda_N)}{\partial \lambda_N} \right|_{\lambda_N} \tag{23}$$

$$= \widehat{\Psi}_N(\lambda_N^*) - \lambda_N^* \left. \widehat{S}_N(\lambda_N) \right|_{\lambda_N} \tag{24}$$

$$= \widehat{\Psi}_N(\lambda_N^*) - \lambda_N^* S_N \, . \tag{25}$$

Note that the partial derivative in (23) reflects the material stress currently acting within the mixture due to the externally applied material stretch λ_N, the evaluation of which according to (17) for the current stretch state λ_N yields the overall stress value S_N acting on the micro-plane under consideration.

The combination of (23) with (10) and (15) yields

$$\widehat{G}_N(\lambda_N^*) = \sum_\alpha \xi_N^\alpha \, \widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^* \left[\sum_\alpha \xi_N^\alpha \left. \frac{\partial \widehat{\psi}_N^\alpha(\lambda_N)}{\partial \lambda_N} \right|_{\lambda_N} \right] \tag{26}$$

$$= \sum_\alpha \xi_N^\alpha \left[\widehat{\psi}_N^\alpha(\lambda_N^*) - \lambda_N^* \left. \widehat{S}_N^\alpha(\lambda_N) \right|_{\lambda_N} \right] \tag{27}$$

$$= \sum_\alpha \xi_N^\alpha \, \widehat{g}_N^\alpha(\lambda_N^*) \, , \tag{28}$$

i.e. the overall Gibbs potential $\widehat{G}_N(\lambda_N^*)$ is obtained from the contributions of the individual constituents, which is consistent to the relation obtained for the overall Helmholtz free energy potential, cf. (10).

2.6 EVOLUTION OF VOLUME FRACTIONS

The approach used for the derivation of the differential equations governing the evolution of volume fractions is based on statistical physics, cf. [27]. In this context, a transformation probability matrix $\boldsymbol{Q} = \widehat{\boldsymbol{Q}}(\boldsymbol{\xi}) \in \mathbb{R}^{2\times 2}$ corresponding to an infinitesimal generator of a Markov process is introduced, see also [28]. The transformation probability matrix drives the evolution of volume fractions via

$$\dot{\boldsymbol{\xi}} = \widehat{\boldsymbol{Q}}(\boldsymbol{\xi}) \cdot \boldsymbol{\xi} \, , \tag{29}$$

wherein the notation $\dot{\bullet}$ denotes the material time derivative. For the two-phase transformation between austenite A and martensite M considered in the contribution at hand, the transformation probability matrix takes the form

$$\boldsymbol{Q} = \omega \begin{bmatrix} -P_{\mathrm{A}\to\mathrm{M}} & P_{\mathrm{M}\to\mathrm{A}} \\ P_{\mathrm{A}\to\mathrm{M}} & -P_{\mathrm{M}\to\mathrm{A}} \end{bmatrix} \neq \boldsymbol{Q}^{\mathrm{t}} \tag{30}$$

with ω being a material parameter denoted as transition attempt frequency and

$$P_{\mathrm{A}\to\mathrm{M}} = \widehat{P}_{\mathrm{A}\to\mathrm{M}}(b_{\mathrm{A}\to\mathrm{M}}) \in [0,1] \subset \mathbb{R} \tag{31}$$

as well as

$$P_{\mathrm{M}\to\mathrm{A}} = \widehat{P}_{\mathrm{M}\to\mathrm{A}}(b_{\mathrm{M}\to\mathrm{A}}) \in [0,1] \subset \mathbb{R} \tag{32}$$

representing the probability for the transformation of austenite A to martensite M and vice versa, respectively. Note that the sum of elements in each column of $\boldsymbol{Q}$ is zero, i.e. $\sum_i Q_{ij} = 0 \; \forall \; j$ holds, inducing a so-called conservative Markov process. It can be shown that this special structure of the system of evolution equations ensures that the physical restrictions (2) regarding the volume fractions ξ^{A} and ξ^{M} are fulfilled without the need to enforce any additional algebraic inequality constraints on the system of evolution equations, cf. also [27].

In particular, according to [27], the probability for the transformation from one phase α into another phase β follows from

$$P_{\alpha\to\beta} = \exp\left(\frac{-\Delta v \, b_{\alpha\to\beta}}{k\,\theta}\right) \quad , \tag{33}$$

with k being Boltzmann's constant, Δv denoting the (constant) transformation region's volume, θ representing the current temperature and $b_{\alpha\to\beta}$ being the energy barrier for the transformation under consideration, i.e. from phase α to phase β. Note that $b_{\alpha\to\beta} \neq b_{\beta\to\alpha}$ and thus $P_{\alpha\to\beta} \neq P_{\beta\to\alpha}$ holds in general. The energy barriers follow from the evaluation of certain Gibbs energy states as elaborated in detail in Section 2.7.

2.7 COMPUTATION OF GIBBS ENERGY BARRIERS

For the computation of Gibbs energy barriers that need to be overcome for the transformation from one phase into another, the minima of the individual Gibbs energy functions $\widehat{g}_N^{\alpha}(\lambda_N^*)$ specified in (22) as well as the Gibbs energy value at the intersection point of the two involved potentials need to be determined. To this end, in the context of the considered two-phase setup, we determine the minima of the austenite and martensite Gibbs free energies in the λ_N^* parameter space via

$$\lambda_{\min}^{\mathrm{A}} = \arg\min \widehat{g}_N^{\mathrm{A}}(\lambda_N^*) \tag{34}$$

and

$$\lambda_{\min}^{\mathrm{M}} = \arg\min \widehat{g}_N^{\mathrm{M}}(\lambda_N^*) \; , \tag{35}$$

respectively. The associated minimal Gibbs free energy values for both phases are then evaluated as

$$g_{\min}^{\mathrm{A}} = \widehat{g}_N^{\mathrm{A}}(\lambda_{\min}^{\mathrm{A}}) \tag{36}$$

and

$$g_{\min}^{\mathrm{M}} = \widehat{g}_N^{\mathrm{M}}(\lambda_{\min}^{\mathrm{M}}) \; . \tag{37}$$

Moreover, the intersection point of both potentials in the fictitious strain space parameterised by λ_N^* yields

$$\widehat{g}_N^{\mathrm{A}}(\lambda_N^*) = \widehat{g}_N^{\mathrm{M}}(\lambda_N^*) \tag{38}$$

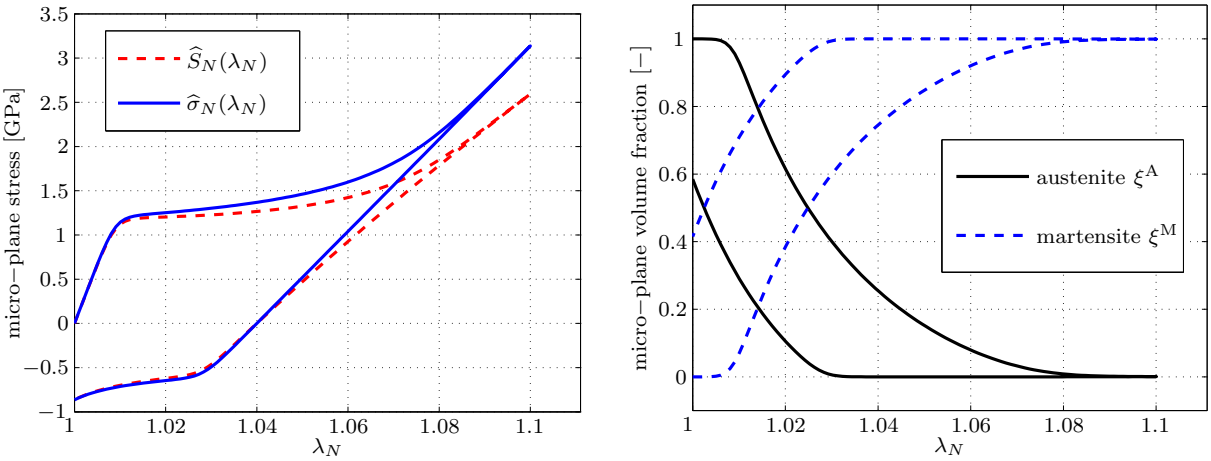
$$\Rightarrow \quad \widehat{g}_N^{\mathrm{A}}(\lambda_N^*) - \widehat{g}_N^{\mathrm{M}}(\lambda_N^*) = 0 \; , \tag{39}$$

which is solved using a minimisation procedure in terms of

$$\lambda_{\min}^{\mathrm{isc}} = \arg\min \left| \, \widehat{g}_N^{\mathrm{A}}(\lambda_N^*) - \widehat{g}_N^{\mathrm{M}}(\lambda_N^*) \, \right| \; , \tag{40}$$

where the superscript 'isc' stands for 'intersection'. The obtained solution $\lambda_{\min}^{\mathrm{isc}}$ therefore corresponds to the specific point in the fictitious strain space at which the Gibbs energy potentials of both phases take the same value, i.e.

$$\widehat{g}_N^{\mathrm{A}}(\lambda_{\min}^{\mathrm{isc}}) = \widehat{g}_N^{\mathrm{M}}(\lambda_{\min}^{\mathrm{isc}}) \; . \tag{41}$$



(a) Stress-stretch response.

(b) Micro-plane evolution of volume fractions.

Figure 1: Micro-plane results: evolution of stress and volume fractions due to an artificially prescribed time-proportional ramp-type micro-plane normal stretch $\lambda_N(t)$ up to a maximum stretch of $\lambda_N(t = 1\,\mathrm{s}) = 1.1$. The load is then reversed to a state of zero stretch, $\lambda_N(t = 2\,\mathrm{s}) = \lambda_N(t = 0) = 1$.

The Gibbs energy barrier $b_{\mathrm{A}\to\mathrm{M}}$ for the transformation from austenite A to martensite M then follows from

$$b_{\mathrm{A}\to\mathrm{M}} = \widehat{g}^{\mathrm{A}}_N(\lambda^{\mathrm{isc}}_{\mathrm{min}}) - g^{\mathrm{A}}_{\mathrm{min}} \,, \tag{42}$$

whereas the Gibbs energy barrier $b_{\mathrm{M}\to\mathrm{A}}$ for the reverse transformation, i.e. from martensite M to austenite A, follows in a similar way according to

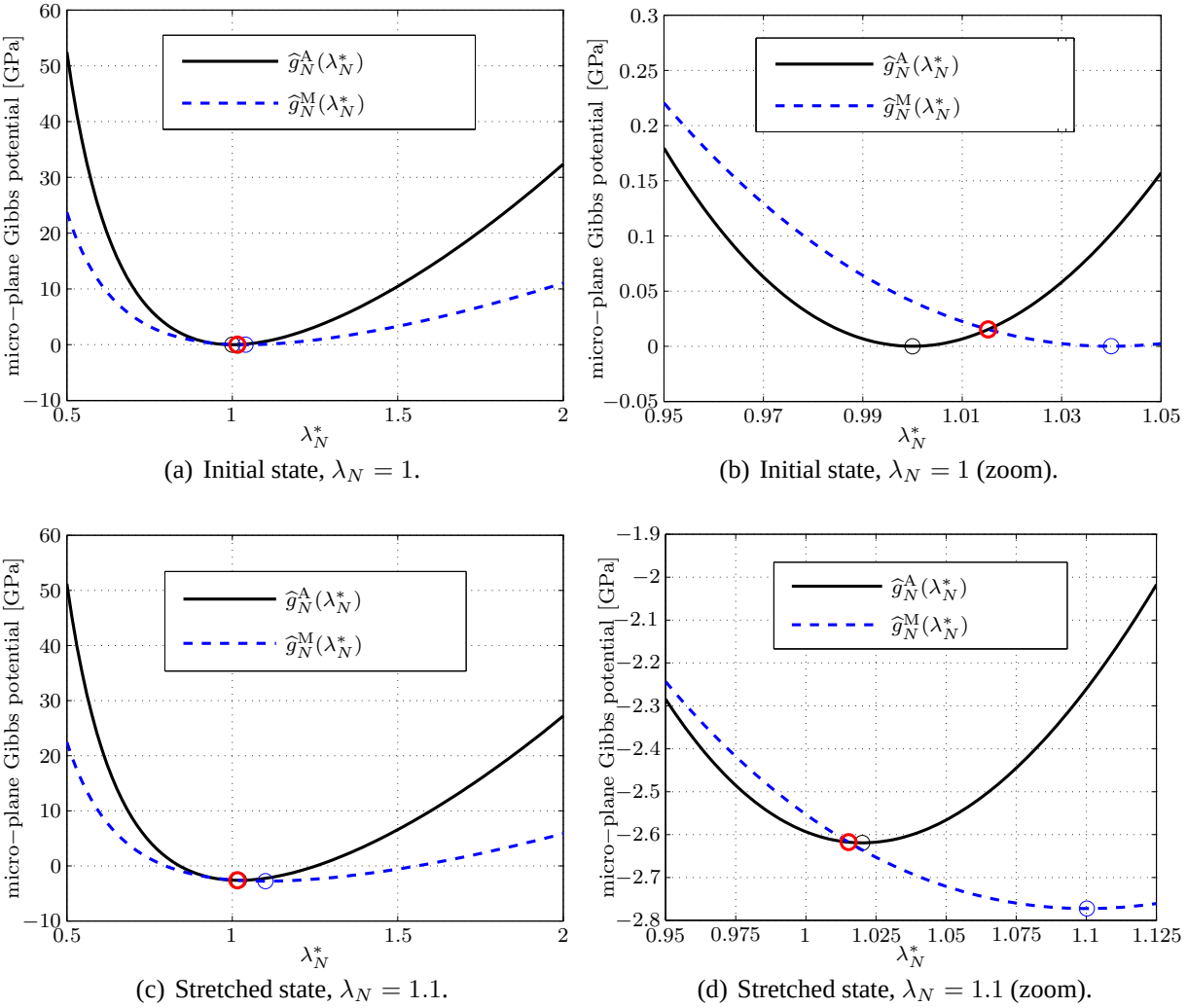
$$b_{\mathrm{M}\to\mathrm{A}} = \widehat{g}^{\mathrm{M}}_N(\lambda^{\mathrm{isc}}_{\mathrm{min}}) - g^{\mathrm{M}}_{\mathrm{min}} \,. \tag{43}$$

With the aforementioned quantities at hand, the system of evolution equations (29) is fully specified and can thus be solved with the help of standard numerical time-integration schemes.

2.8 NUMERICAL EXAMPLE ON THE MICRO-PLANE LEVEL

For the numerical example on the micro-plane level, we consider a stretch-driven deformation and evaluate the stress response (17) as well as the deformation-driven evolution of internal variables, i.e. volume fractions ξ^{A} and ξ^{M}, cf. (29). Given the two-phase setup considered in the contribution at hand, for the transformation strain of the austenitic parent phase A we set $\lambda^{\mathrm{A}}_{\mathrm{tr}} = 1$ and for the martensite phase M we choose $\lambda^{\mathrm{M}}_{\mathrm{tr}} = 1.04$. Further parameters are chosen as $\omega = 16\,\mathrm{s}^{-1}$, $\Delta v = 2.71 \times 10^{-8}\,\mathrm{mm}^3$, and $\theta = 273\,\mathrm{K}$, see Section 2.6.

The numerical results obtained for a prescribed micro-plane stretch λ_N with time-proportional quasi-static loading are depicted in Fig. 1. The stress-stretch relation shows a plateau-type stress evolution within the material as soon as the martensite transformation is activated at approximately $\lambda_N \approx 1.01$. At this deformation state, the austenitic parent phase starts to decrease, while at the same time the martensite volume fraction increases, cf. Fig. 1(b). As the martensitic transformation is completed, i.e. at $\xi^{\mathrm{M}} \to 1$ as $\lambda_N \to 1.08$, further stretch of the material induces a linear increase in elastic martensite stress, cf. Fig. 1(a). In addition to the micro-plane material stress measure $S_N = \widehat{S}_N(\lambda_N)$, cf. (17), we compute the micro-plane Cauchy stress $\sigma_N = \widehat{\sigma}_N(\lambda_N)$ via $\widehat{\sigma}_N(\lambda_N) = \lambda_N^4\, \widehat{S}_N(\lambda_N)/J^2$, see [33], where $J = \det(\boldsymbol{F})$ denotes the determinant of the macroscopic deformation gradient $\boldsymbol{F}$. In the context of the artificially applied micro-plane strain, however, we choose $J = \lambda_N$.



(a) Initial state, $\lambda_N = 1$.

(b) Initial state, $\lambda_N = 1$ (zoom).

(c) Stretched state, $\lambda_N = 1.1$.

(d) Stretched state, $\lambda_N = 1.1$ (zoom).

Figure 2: Micro-plane Gibbs potentials for austenite and martensite, g_N^{A} and g_N^{M}, respectively, at two selected material stretch states λ_N. Note that $\lambda_N = 1$ refers to the initial, undeformed configuration and that λ_N^* is a fictitious stretch used for parameterisation of g_N, cf. Section 2.5.

The Gibbs energy potentials of the individual phases as derived in (22) are provided in Fig. 2 for both the initial material state and the state of maximum stretch, i.e. for states of $\lambda_N(t = 0) = 0$ and $\lambda_N(t = 1\,\mathrm{s}) = 1.1$, respectively. The non-quadratic nature of both micro-plane Gibbs energy potentials—as resulting from the specific form of the underlying Helmholtz free energy definition (6)—is shown in Figs. 2(a) and 2(c).

A closer look at the intersection points of the individual phase potentials—cf. Figs. 2(b) and 2(d)—reveals the location of the Gibbs energy minima in relation to the Gibbs energy value at the intersection point. The minimum-related Gibbs energy values, (36) and (37), as well as the Gibbs energy value at the intersection point (41) need to be computed in order to obtain the Gibbs energy barriers, (42) and (43), which define the transformation probabilites (33) and thus drive the evolution of volume fractions (29).

3 MICRO-SPHERE APPLICATION

Following [33], the macroscopic Piola-Kirchhoff stress tensor $\boldsymbol{S}$ is computed based on the individual material micro-plane stress measures $\widehat{S}(\lambda_N)$ via integration over the unit hemisphere

in terms of

$$\boldsymbol{S} = \frac{3}{2\,\pi} \int_{\mathbb{H}^2} \widehat{S}(\lambda_N)\, \lambda_N^{-1}\, \boldsymbol{N} \otimes \boldsymbol{N}\, \mathrm{d}\omega \,. \tag{44}$$

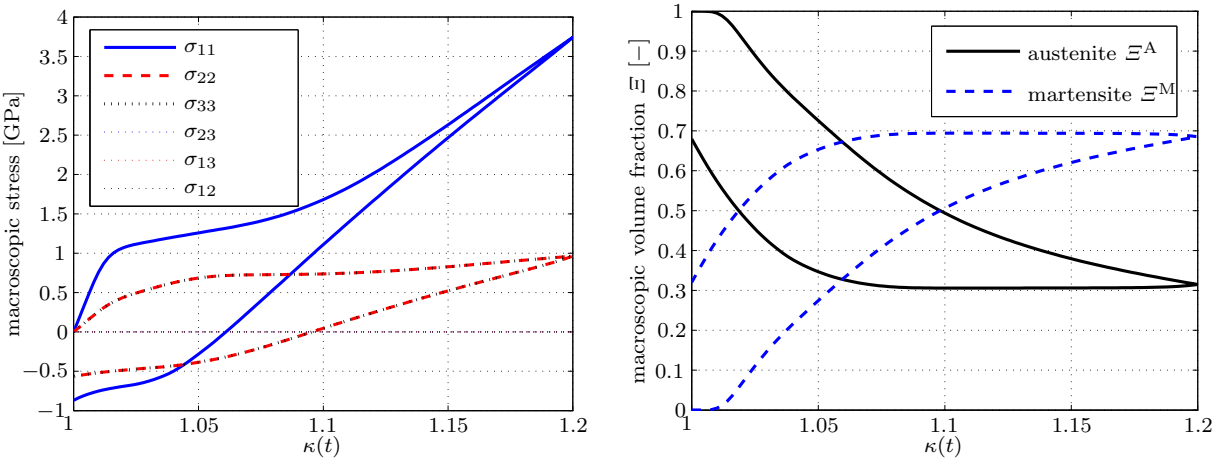
For the macroscopic Cauchy stress response $\boldsymbol{\sigma}$ depicted in Fig. 3, we apply the standard push-forward of $\boldsymbol{S}$ viz.

$$\boldsymbol{\sigma} = \frac{1}{J}\, \boldsymbol{F} \cdot \boldsymbol{S} \cdot \boldsymbol{F}^{\mathrm{t}} \,. \tag{45}$$

The quasi-static application of a macroscopic non-isochoric homogeneous state of tensile deformation

$$\boldsymbol{F}(t) = \boldsymbol{I} + \kappa(t)\, \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 \tag{46}$$

with $\kappa(t) \in [0, 0.2] \subset \mathbb{R}$ introduced as a time-proportional load scaling parameter results in the macroscopic material response provided in Fig. 3. The obtained Cauchy stress response, Fig. 3(a), shows a plateau-type stress characteristic that is related to the deformation-induced onset of martensitic transformation within the polycrystal. The macroscopic volume fractions $\Xi^{\bullet}$—homogenised in terms of a moment of zeroth order—show a saturation-type evolution, i.e. even for very large deformations the polycrystal will not transform to 100% martensite.



(a) Macroscopic Cauchy stress evolution.

(b) Macroscopic volume fractions.

Figure 3: Application of the micro-sphere framework: evolution of stress and volume fractions due to a quasi-statically applied macroscopic non-isochoric homogeneous state of tensile deformation in terms of $\boldsymbol{F}(t) = \boldsymbol{I} + \kappa(t)\, \boldsymbol{e}_1 \otimes \boldsymbol{e}_1$.

4 CONCLUSIONS

The contribution at hand introduces a constitutive formulation that extends an established finite strain micro-sphere framework for the simulation of Neo-Hookean-type materials towards phase-transformations between austenite and martensite. Individual micro-plane free energy terms are assigned to both phases, where the Helmholtz free energy of martensite is consistently extended to include a Bain-type transformation strain which is multiplicatively coupled to the overall micro-plane strain measure of the martensite fraction.

The numerical examples obtained at both the micro-plane level, see Section 2.8, and at the micro-sphere level, see Section 3, show that the proposed modelling framework—even though not yet fitted to experimental data—is able to capture the typical, experimentally observed stress-strain behaviour of, e.g., shape memory alloys. Moreover, the proposed micro-sphere

formulation naturally captures the polycrystalline transformation behaviour that is related to the so-called grain-locking effect as introduced in Section 1.

In conclusion, the proposed micro-sphere based finite strain modelling framework for phase-transformations can be considered as a promising foundation in view of later extensions towards coupled plasticity and temperature effects in polycrystals.

REFERENCES

[1] S. Miyazaki, "Medical and dental applications of shape memory alloys," *Shape Memory Materials, Cambridge University Press*, vol. 12, pp. 267–281, 1999.

[2] D. Hartl and D. Lagoudas, "Aerospace applications of shape memory alloys," *P. I. Mech. Eng. G: J. Aer.*, vol. 221, pp. 535–552, 2007.

[3] K. Otsuka and T. Kakeshita, "Science and technology of shape-memory alloys: new developments," *MRS Bull*, vol. 27, pp. 91–100, 2002.

[4] T. Merzouki, A. Duval, and T. B. Zineb, "Finite element analysis of a shape memory alloy actuator for a micropump," *Sim. Mod. Pract. Theory*, vol. 27, pp. 112–126, 2012.

[5] L. Brinson, I. Schmidt, and R. Lammering, "Stress-induced transformation behavior of a polycrystalline NiTi shape memory alloy: micro and macromechanical investigations via in situ optical microscopy," *J. Mech. Phys. Solids*, vol. 52, no. 7, pp. 1549–1571, 2004.

[6] S. Govindjee, A. Mielke, and G. Hall, "The free energy of mixing for n-variant martensitic phase transformations using quasi-convex analysis," *J. Mech. Phys. Solids*, vol. 50, pp. 1897–1922, 2002.

[7] T. Bartel and K. Hackl, "A micromechanical model for martensitic phase-transformations in shape-memory alloys based on energy-relaxation," *J. Appl. Math. Mech.*, vol. 89 (10), pp. 792–809, 2009.

[8] K. Hackl and R. Heinen, "An upper bound to the free energy of n-variant polycrystalline shape memory alloys," *J. Mech. Phys. Solids*, vol. 56, pp. 2832–2843, 2008.

[9] C. Collard and T. B. Zineb, "Simulation of the effect of elastic precipitates in SMA materials based on a micromechanical model," *Comp. Part B: Eng.*, vol. 43, no. 6, pp. 2560–2576, 2012.

[10] A. S. J. Suiker and S. Turteltaub, "Numerical modelling of transformation-induced damage and plasticity in metals," *Model. Simul. Mater. Sc.*, vol. 15, pp. 147–166, 2007.

[11] S. Turteltaub and A. S. J. Suiker, "Grain size effects in multiphase steels assisted by transformation-induced plasticity," *Int. J. Sol. Struct.*, vol. 43, no. 24, pp. 7322–7336, 2006.

[12] A. S. J. Suiker and B. Thijsse, "Nucleation, kinetics and morphology of displacive phase transformations in iron," *J. Mech. Phys. Solids*, vol. 61, no. 11, pp. 2273–2301, 2013.

[13] T. Bartel, B. Kiefer, K. Buckmann, and A. Menzel, "A kinematically-enhanced relaxation scheme for the modeling of displacive phase transformations," *J. Intel. Mat. Syst. Str.*, vol. 26, no. 6, pp. 701–717, 2015.

[14] T. Bartel and A. Menzel, "Partially relaxed energy potentials for the modelling of microstructures – application to shape memory alloys," *GAMM-Mitteilungen*, vol. 35, no. 1, pp. 59–74, 2012.

[15] D. Helm and P. Haupt, "Shape memory behaviour: modelling within continuum mechanics," *Int. J. Sol. Struct.*, vol. 40, pp. 827–849, 2003.

[16] F. Auricchio and R. Taylor, "Shape-memory alloys: modelling and numerical simulations of the finite-strain superelastic behavior," *Comput. Method. Appl. M.*, vol. 143, pp. 175–194, 1997.

[17] B. Raniecki, C. Lexcellent, and K. Tanaka, "Thermodynamic models of pseudoelastic behaviour of shape memory alloys," *Arch. Mech.*, vol. 44, pp. 261–284, 1992.

[18] C. Truesdell and R. Toupin, "The classical field theories," in *Handbuch der Physik* (S. Flügge, ed.), vol. 3, Berlin: Springer, 1960.

[19] D. Hartl and D. Lagoudas, "Constitutive modeling and structural analysis considering simultaneous phase transformations and plastic yield in shape memory alloys," *Smart Mater. Struct.*, vol. 18, no. 10, pp. 1–17, 2009.

[20] D. Helm, *Formgedächtnislegierungen – Experimentelle Untersuchung, phänomenologische Modellierung und numerische Simulation der thermomechanischen Materialeigenschaften*. Dissertation, Universität Gesamthochschule Kassel, 2001. Berichte des Instituts für Mechanik (Bericht 3/2001).

[21] D. Helm and P. Haupt, "Thermomechanical behaviour of shape memory alloys," *Proceedings of SPIEs Smart Structures and Materials*, vol. 4333, pp. 302–313, 2001.

[22] K. Tanaka, "A thermomechanical sketch of shape memory effect: one-dimensional tensile behavior," *Res. Mech*, vol. 18, pp. 251–263, 1986.

[23] S. Reese and D. Christ, "Finite deformation pseudo-elasticity of shape memory alloys – constitutive modelling and finite element implementation," *Int. J. Plasticity*, vol. 24, pp. 455–482, 2008.

[24] S. Govindjee and C. Miehe, "A multi-variant martensitic phase transformation model: formulation and numerical implementation," *Comput. Method. Appl. M.*, vol. 191, no. 3-5, pp. 215–238, 2001.

[25] E. Stein and O. Zwickert, "Theory and finite element computations of a unified cyclic phase transformation model for monocrystalline materials at small strains," *Comput. Mech.*, vol. 40, no. 3, pp. 429–445, 2007.

[26] E. Patoor, D. Lagoudas, P. Entchev, L. Brinson, and X. Gao, "Shape memory alloys, part I: general properties and modeling of singlecrystals," *Mech. Mater.*, vol. 38, pp. 391–429, 2006.

[27] M. Achenbach, "A model for an alloy with shape memory," *Int. J. Plasticity*, vol. 5, pp. 371–395, 1989.

[28] S. Govindjee and G. Hall, "A computational model for shape memory alloys," *Int. J. Sol. Struct.*, vol. 37, pp. 735–760, 2000.

[29] R. Ostwald, T. Bartel, and A. Menzel, "A computational micro-sphere model applied to the simulation of phase-transformations," *J. Appl. Math. Mech.*, vol. 90, no. 7-8, pp. 605–622, 2010.

[30] R. Ostwald, T. Bartel, and A. Menzel, "Phase-transformations interacting with plasticity – a micro-sphere model applied to TRIP steel," *Comp. Mater. Sci.*, vol. 64, pp. 12–16, 2012.

[31] R. Ostwald, T. Bartel, and A. Menzel, "A Gibbs-energy-barrier-based computational micro-sphere model for the simulation of martensitic phase-transformations," *Int. J. Numer. Meth. Eng.*, vol. 97, pp. 851–877, 2014.

[32] R. Ostwald, T. Bartel, and A. Menzel, "An energy-barrier-based computational micro-sphere model for phase-transformations interacting with plasticity," *Comput. Method. Appl. M.*, vol. 293, pp. 232–265, 2015.

[33] I. Carol, M. Jirásek, and Z. Bažant, "A framework for microplane models at large strain, with application to hyperelasticity," *Int. J. Sol. Struct.*, vol. 41, pp. 511–557, 2004.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

OPTIMAL CONTROL TUNING IN SMART STRUCTURES WITH DELAMINATIONS

Panagiotis Koutsianitis[1], Amalia Moutsopoulou[2], Georgios A. Drosopoulos[1], Georgios K. Tairidis[1], Georgia Foutsitzi[3] and Georgios E. Stavroulakis[1]

[1] Technical University of Crete
School of Production Engineering and Management, Chania, Greece
e-mail: {gestavroulakis, gdrosopoulos}@isc.tuc.gr , {panoskout, tairidis}@gmail.com

[2] Technological Educational Institute of Crete
Department of Mechanical Engineering, Heraklion, Greece
amalia@staff.teicrete.gr

[3] Technological Educational Institute of Epirus
Department of Accounting and Management, Arta, Greece
gfoutsi@gmail.com

Keywords: Instructions, ECCOMAS Congress, Computational Methods, Engineering Sciences, Proceedings.

Abstract. *An efficient strategy for calculation of delaminations in composite beams and intelligent structures is used in order to quantify structural uncertainties within a finite element model of a piezocomposite (multilayered plate theory). Furthermore the dynamical system is connected with robust and neurofuzzy control. The problem of positioning of actuators and sensors has been investigated. Model based simulations of increasing complexity illustrate some of the attractive features of the strategy in terms of accuracy as well as computational cost. This shows the possibility of using such strategies for the development of smart structural and systems.*

1 INTRODUCTION

Piezoelectric sensors and actuators are extensively employed in many practical applications such as intelligent structures due to their lightness and their capability of coupling strain and electric fields. In order to control structural vibrations, piezoelectric sensors and actuators can be easily bonded on the vibrating structure. In active vibration control, the patches are symmetrically bonded on top and bottom surfaces [1,2, 3]. An Hinfinity [4,5] controller, a fuzzy controller [20,21,26] or a hybrid neuro-fuzzy control [22,23,24,25] can be designed and optimally tuned [27, 28] to suppress the vibrations of the intelligent structure even when delamination phenomena occur [32, 33]. The position and the size of the actuators influence the response of the smart structure and determine the maximum admissible piezoelectric actuation voltage. The optimal sensor locations can be found, using genetic algorithms and multicriteria optimization [6,13]. In addition the numerical results can be compared with analytical ones obtained by the Laplace transform method [7,8]. Satisfactory vibration reduction has been observed for control cases under sinusoidal, random disturbances, confirming the reliability and validity of such algorithms. The results are very satisfactory since the vibration suppression is achieved.

2 FINITE ELEMENT FORMULATION

We base our model on the layerwise approach which takes into account the adhesive layer derived earlier in Refs [32, 33]. The structure consists of an elastic core bonded with piezoelectric sensor and actuator layers and adhesive layers between them. Delamination between the layers may appear and influences the effectiveness of active vibration control. Using a special adhesive finite element we are leading to the following equation of motion in terms of nodal variable *q*,

$$M\ddot{q}(t)+D\dot{q}(t)+Kq(t)=f_m(t)+f_e(t) \tag{1}$$

where *M* is the generalized mass matrix, *D* the viscous damping matrix, *K* the generalized stiffness matrix, f_m the external loading vector and f_e the generalized control force vector produced by electromechanical coupling effects. The exact form of these matrices are given in [32], [33].

To transform to state-space control representation, let (in the usual manner),

$$\dot{x}(t)=\begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} \tag{2}$$

Furthermore to express $f_e(t)$ as $Bu(t)$ we write it as $f_e^* u$ where f_e^* the piezoelectric force is for a unit applied on the corresponding actuator, and u represents the voltages on the actuators. Furthermore, $d(t)=f_m(t)$ is the disturbance vector [9, 10].

Then,

$$\dot{x}(t)=\begin{bmatrix} O_{2n\times 2n} & I_{2n\times 2n} \\ -M^{-1}K & -M^{-1}D \end{bmatrix} x(t)+\begin{bmatrix} O_{2n\times 2n} \\ M^{-1}f_e^* \end{bmatrix} u(t)+\begin{bmatrix} O_{2n\times 2n} \\ M^{-1} \end{bmatrix} \tag{3}$$

$$= Ax(t) + Bu(t) + Gd(t) = Ax(t) + \begin{bmatrix} B & D \end{bmatrix} \begin{bmatrix} u(t) \\ d(t) \end{bmatrix} = Ax(t) + \tilde{B}\tilde{u}(t) \tag{5}$$

The previous description of the dynamical system will be augmented with the output equation (some displacements or velocities are measured) [10, 11, 12],

$$y(t) = \begin{bmatrix} x_1(t) & x_3(t) & \cdots & x_{n-1}(t) \end{bmatrix}^T = Cx(t) \tag{6}$$

3 ROBUST CONTROL PROBLEM

Two control laws for the composite beam are designed in order to suppress the vibrations. Because of its linearity and easy implementation, the linear quadratic regulator (LQR)[12,13] is presented first. The response of the controlled nominal and damaged beams is investigated. In order to take into account the incompleteness of the information about the eventual damages and external additional influences a robust H_∞ controller is designed [14, 15]. A system analysis is made on condition that the system is not accurate but includes uncertainty that may be related to some kind of damage [9]

For practical applications both algorithms need several trial-and-error design iterations in order to provide appropriate control voltages, since the piezoelectric actuators can be depolled by high oscillating voltages. The effectiveness of the proposed control strategies is investigated with the help of numerical simulations [7, 9].

The following three steps are taken in the robustness analysis [16, 17]:

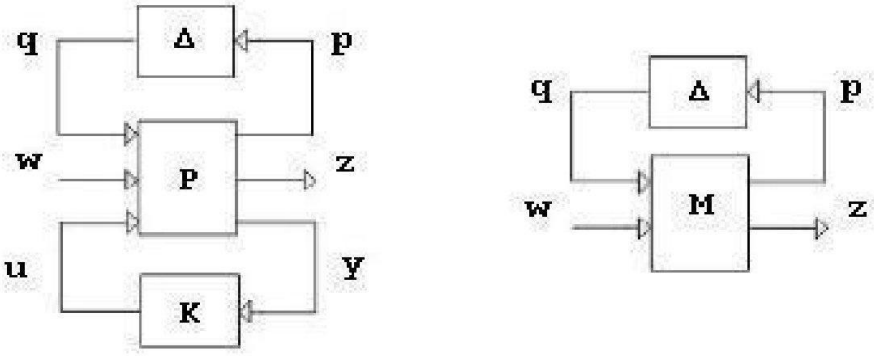


Figure 1 Uncertainty modelling

1. Expression of an *uncertainty* set by a mathematical model.
2. *Robust stability* (RS): check if the system remains stable for all plants within the uncertainty set.
3. *Robust performance* (RP): if the system is robustly stable, check whether performance specifications are met for all plants within the uncertainty set.

To perform the robustness analysis, the interconnection of Figure 1 will be used.
Δ define the uncertainty, M define the nominal system, w are the inputs (the mechanical force and the noise of the system), z are the outputs (the state vector and the control vector). The uncertainty included in Δ satisfies $\|\Delta\|_\infty \leq 1$.

The system (M, Δ) is *robustly* stable if,

$$\sup_{\omega \in \mathbb{R}_n} \mu_\Delta(\mathrm{M}_{11}(j\omega)) \prec 1 \tag{7}$$

The system (M, Δ) exhibits robust performance if,

$$\sup_{\omega \hat{I}_n} \mu_{\Delta_\alpha}(\mathrm{M}(j\omega)) \prec 1 \tag{8}$$

where,

$$\Delta_\alpha = \begin{bmatrix} \Delta_\rho & 0 \\ 0 & \Delta \end{bmatrix}$$

and Δ_ρ has the same structure as Δ but dimensions corresponding to (w,z). Unfortunately, only bounds on μ can be estimated.

To proceed let us assume uncertainty in the M, D and K matrices of the form,

$$\begin{aligned} &\mathrm{M} = \mathrm{M}_0(\mathrm{I} + m_p \delta_{M_u}) D = D_0(I + d_p \delta_D) \\ &K = K_0(I + k_p \delta_{\mathrm{K}}) \end{aligned} \tag{9}$$

with,

$$\|\Delta\|_\infty \overset{def}{=} \left\| \begin{bmatrix} \delta_{\mathrm{M}} & & \\ & \delta_D & \\ & & \delta_{\mathrm{K}} \end{bmatrix} \right\| \prec 1 \tag{10}$$

This means that we are allowing a percentage deviation from the nominal values [6,9].

With these definitions Eq. (1) becomes,

$$\Rightarrow \mathrm{M}_0 \ddot{q}(t) + D_0 \dot{q}(t) + K_0 q(t) = \tilde{D} q_u(t) + f_m(t) + f_e(t) \tag{11}$$

Where,

$$q_u(t) \overset{def}{=} \begin{bmatrix} \ddot{q}(t) \\ \dot{q}(t) \\ q(t) \end{bmatrix}$$

$$\tilde{D} = -\begin{bmatrix} M_0 m_p & D_0 d_p & K_0 k_p \end{bmatrix} \begin{bmatrix} I_{2n\times 2n}\delta_{\mathrm{M}} & & \\ & I_{2n\times 2n}\delta_{\Delta} & \\ & & I_{2n\times 2n}\delta_{\mathrm{K}} \end{bmatrix} \tag{12}$$

Writing [5]in state space form, gives,

$$\dot{x}(t) = \begin{bmatrix} O_{2n\times 2n} & I_{2n\times 2n} \\ -M^{-1}K & -M^{-1}D \end{bmatrix} x(t) + \begin{bmatrix} O_{2n\times 2n} \\ M^{-1} f_e^* \end{bmatrix} u(t) + \begin{bmatrix} O_{2n\times 2n} \\ M^{-1} \end{bmatrix} d(t) + \begin{bmatrix} O_{2n\times 6n} \\ M^{-1}D \end{bmatrix} q_u(t)$$

$$\dot{x}(t) = Ax(t) + Bu(t) + Gd(t) + G_u q_u(t) \tag{13}$$

In this way we treat uncertainty in the original matrices as an extra uncertainty term.

4 FUZZY AND NEURO-FUZZY CONTROL

A fuzzy inference system can be developed within MALAB using the fuzzy toolbox. The control scheme usually consists of a Mamdani [20,21] or a Sugeno-type controller [22,23,24]. This system can be used for the study of smart structures such as smart beams [21,24,25], plates [26,27], etc.The controllers usually consist of two inputs and one output. As for inputs, the controller can take displacement and velocity or the electric potential and current and returns the control force. The membership functions have triangular, trapezoidal, Gaussian or sigmoid both for inputs and output. An example of triangular and trapezoidal functions is shown in Fig. 2. The inference system involves membership functions combined with use of logical operations. Namely, the decision is based on a set of if-then rules, thus the recurring system is a rule-based system. A set of rules as this one shown in Table 1, can be considered. The graphic representation of rules is given by the fuzzy surface shown in Fig. 2.

1. Displacement / 2. Velocity	3. Far up	4. Close up	5. Equilibrium	6. Close down	7. Far down
8. Up	9. Max	10. Med+	11. Low+	12. Null	13. Low-
14. Null	15. Med+	16. Low+	17. Null	18. Low-	19. Med-
20. Down	21. High+	22. Null	23. Low+	24. Med-	25. Min

Table 1. Fuzzy inference rules

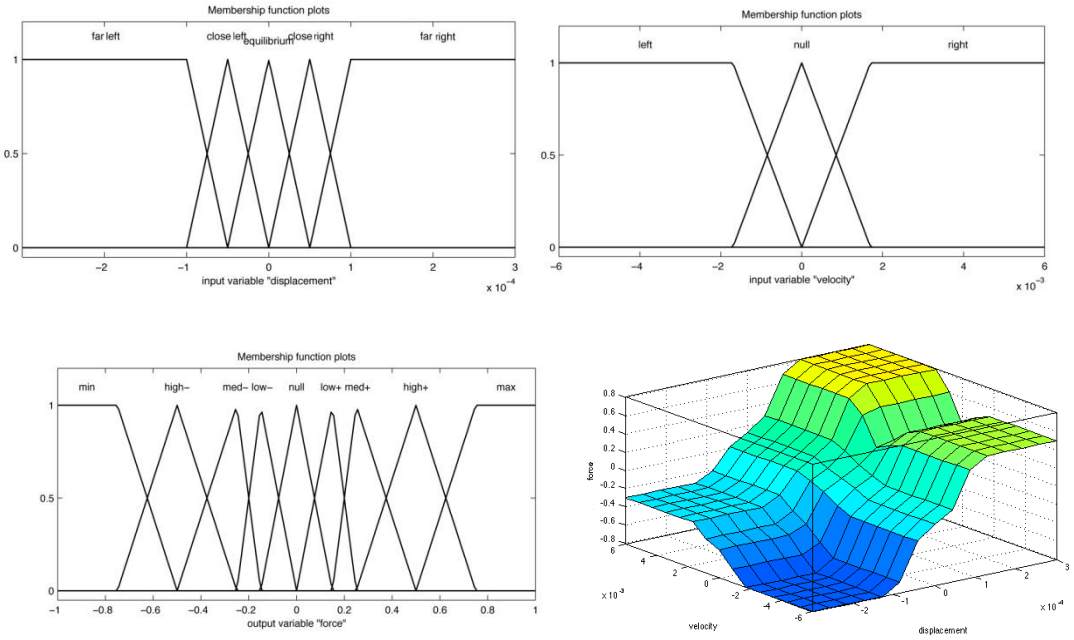


Figure 2: Membership functions prior optimization and fuzzy surface

5. OPTIMIZATION OF CONTROL

The results given solely by fuzzy control in previous investigations of our team [21, 24, 26], were very satisfactory in terms of displacement. However, the results regarding the velocity and the acceleration were not acceptable.

Thus, the need of optimization some of the parameters of the fuzzy controller arose. This fine tuning is a really exacting process, thus a simple, effective and well known procedure has to be chosen.

For example, a genetic algorithm can be used due to its simplicity and smooth behavior in such problems [27]. The algorithm was used in order to optimize the membership functions of the variables (inputs/outputs) or other characteristics of the fuzzy controller. In this case, a population of different solutions try to reach total optima via a stochastic iterative process.

Another method that may be useful for the fine tuning of the parameters of a controller, used for vibration suppression of smart structures, is the particle swarm optimization method [28, 29, 30] and the differential evolution [31]. These methods are similar to the genetic approach, however they can provide more accurate results.

6. DELAMINATION MODELING

For the simulation of delamination, a non-linear delamination law can be considered to the structural model investigated, as shown in Fig. 3:

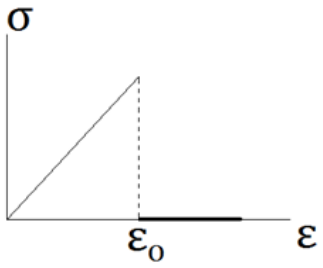


Figure 3: Stress-strain behavior in the adhesive

The above law leads to appropriate modification of the existing stiffness matrix of the structure in order to take into account the differences in the behavior before and after the delamination [32].

For each finite element, the element average strain ε_{zz}can be calculated as follows:

$$\varepsilon_{zz}^{(a_i)} = \frac{w^{(i+1)} - w^{(i)}}{h^{(a_i)}} \qquad (14)$$

where w is the vertical displacement of layers, whose behavior in delamination is investigated, and h(a) is the thickness of the adhesive layer. If the strain is less than ε_o, no delamination appears in the element. Otherwise, delamination takes place. In this case, the delamination can occur between the lower-middle and the middle-upper plate layer.

For the implementation of the non-linear delamination law, the Newton-Raphson incremental-iterative procedure has been used, as follows [33]:

Incremental step

Load enforcement

Start iterations

-Loop to the whole of elements

-Calculate the mean strain of each element

-Compare with ε_o

-Appropriate estimation of the tangential stiffness matrix

-Estimation of the internal force vector

-Solving equilibrium equs and finding incremental

and global displacements

-Error control and continuation iteration

orcontinue the incrementalstep

7. RESULTS AND DISCUSSION

For the numerical simulations a cantilever composite beam with piezoelectric layers bonded on its top and bottom and discretized with finite elements, is used. Periodic sinusoidal loading pressure acting on the side of the structure simulating a strong wind. A sinusoidal load with an amplitude of 15N and frequency of 6.5 *rad/sec*, has been considered. Fig 4 shows the dynamical response for the displacements of the uncontrolled and controlled beam with LQR control [12] and H_∞ control [18], for the four nodes of the beam. Fig 5shows the dynamical response for the rotations of the uncontrolled and controlled beam with LQR control and H_∞ control strategy, for the four nodes of the beam. The beam with H_∞ control keeps in equilibrium and we have zero ddisplacements, complete vibration reduction is achieved. The comparison of the open and closed loop frequency response of the system are shown in Figure 6, as shown in figure, there is a significant improvement in the effect of disturbance on error up to the frequency of 1000 Hz. Figure 7 shows the control voltages for the four nodes of the beam. The control voltages for the disturbance rejections of the beam are less than 500 volt.
Results are very good, and the beam remains in equilibrium. Reduction of vibrations is observed, while piezoelectric add-ons produce voltage within their tolerance limits (±500volt).

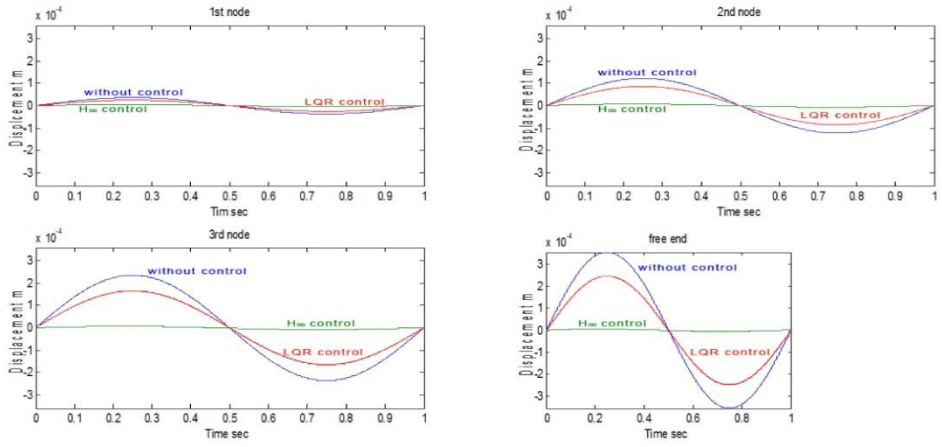


Figure 4: Response of the four nodes for the displacement with LQR and H_∞ control and response without control.

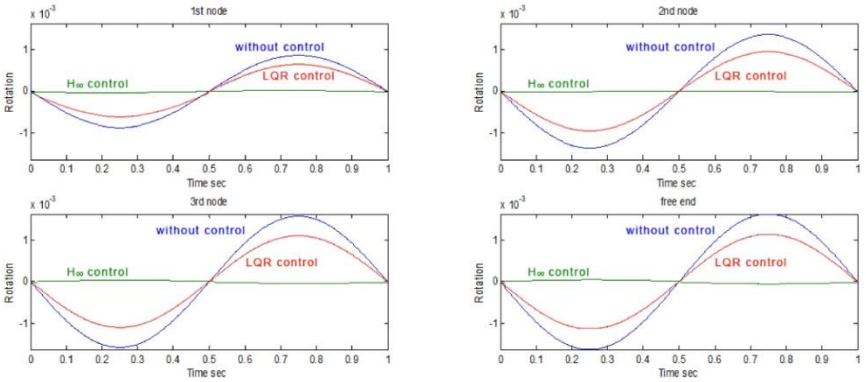


Figure 5: Response of the four nodes on the vibrating beam for the rotation with LQR and H_∞ control and without control.

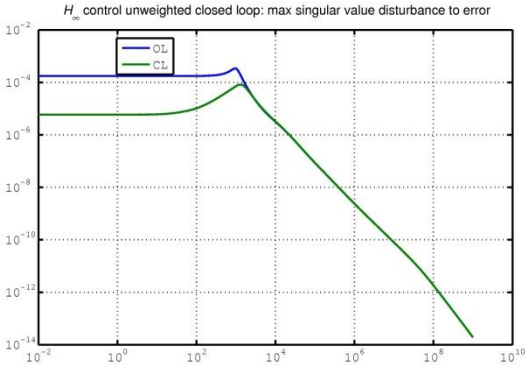


Figure 7: Singular Value for H_∞ control strategy.

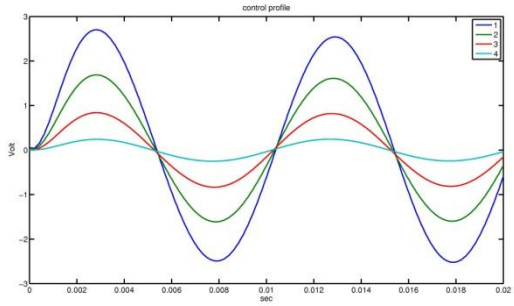


Figure 8: Response of the four nodes for the displacement with LQR and control and response without control.

As a conclusion one can say that structural uncertainty in smart piezocomposites can be quantified by using techniques of computational mechanics. This gives rise to more accurate design and optimal tuning of robust controllers, either following classical or soft computing techniques. Further investigation includes the optimal positioning of sensors and actuators and the incorporation of other uncertainties (crack initiation and propagation, fatigue).

.

REFERENCES

[1] B. Culshaw, *Smart structures: A concept or a reality*, Journal of Systems and Control Engg, Vol. 26, No. 206, 1992.

[2] S. Rao, M. Sunar, (994), *Piezoelectricity and its uses in disturbance sensing and control of flexible structures: A survey*, Applied mechanics Review, Vol. 17, No..2, pp. 113-119.

[3] T. Baily, J. E. Jr. Hubbard, *Distributed piezoelectric polymer active vibration control of a cantilever beam*, Journal of Guidance, Dynamics and Control, Vol. 8, No. 5, pp. 605-611, 1985.

[4] S. Hanagud, M. W. Obal, and A. J. Callise, *Optimal vibration control by the use of piezoelectric sensors and actuators*, J. Contr. Guidance, Vol. 15, No. 5, pp.1199-1206, 1992.

[5] W. S. Hwang and H. C.l Park, *Finite element modeling of piezoelectric sensors and actuators*, AIAA journal J, Vol. 31, No. 5, pp. 930-937, 1993.

[6] J. Moutsopoulou, A. Pouliezos, and G. E. Stavroulakis, *Modelling with Uncertainty and Robust Control of Smart Beams*, Paper 35, Proceedings of the Ninth International Conference on Computational Structures Technology. B.H.V. Topping and M. Papadrakakis, (Editors), Civil Comp Press, 2008.

[7] B. Miara, G. E. Stavroulakis, V. Valente, *Topics on mathematics for smartsystems*, Proceedings of the European Conference. Rome, Italy, 26-28 October 2006,World Scientific Publishers, Singapore, International, 2007.

[8] Bandyopadhyay, T.C. Manjunath, and M. Unapathy, *Modeling,Control, and Implementation of Smart Structures*, Springer ISBN-10 3-540-48393-4, 2007.

[9] A.J. Moutsopoulou, A. Pouliezos, G. E. Stavroulakis, *Nonsmooth and nonconvex optimization for the design and order reduction of robust controllers used in smart structures*, In: M. Papadrakakis et all., *Computational Methods in Erthquake Engineering*, Vol. 2, Springer,2012.

[10] Benjeddou, M.A. Trindade, and R. Ohayon, *New shear actuated smart structure beam finite element*, Journal of Guidance. Dynamics and Control.37, 378-383, 1999.

[11] B. Bona, M. Indri, and A. Tornamble, *Flexible piezoelectric structures approximate motion equations and control algorithms*, IEEE Auto.Control 42(1), 94-101, 1997.

[12] O. Bosgra, and H. Kwakernaak, *Design methods for control systems*, Course notes, Dutch Institute for Systems and Control 67, 2001.

[13] J. Burke, A. Lewis, and M.L. Overton, A *robust gradient sampling algorithm for nonsmooth nonconvex optimization*, Siam Journal on Optimization 15, 751-779, 2005.

[14] H. Kimura, *Robust stability for a class of transfer functions*, IEETransactions on Automatic Control 29, 788–793, 1984.

[15] H. Kwakernaak, *Robust control and Hinfinityoptimization*, Tutorial hper JFAC. Automatica, 29(2), 255-273, 1993.

[16] M. Millston, *HIFOO 1.5: Structured control of linear systems with a non trivial feedthrough*, M.S. thesis, Courant Institute of Mathematical Sciences, New York University, 2006.

[17] Packard, J. Doyle, and G. Balas, *Linear, multivariable robust control with a μ perspective*, ASME Journal of Dynamic Systems, Measurement and Control, 50th Anniversary Issue,115(2b), 310–319, 1993.

[18] G. Zames, *Feedback minimax sensitivity and optimal robustness*, IEE Trans. Autom. Control, 28, 585–601, 1983.

[19] G. Foutsitzi, G.E.Stavroulakis and E. Hadjigeorgiou, *Modelling and Simulation of Controlled Smart Structures*, Proceedings of the 9th HSTAM International Congress on Mechanics pp. 511-518, Limassol, Cyprus, 2010.

[20] G.K Tairidis, G.E. Stavroulakis, D.G. Marinova and E.C. Zacharenakis, *Classical and soft robust active control of smart beams*, ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Rethymno, Crete, Greece, 13-16 June, 2007.

[21] G.K. Tairidis, G.E. Stavroulakis, D.G. Marinova and E.C. Zacharenakis, *Classical and soft robust active control of smart beams*, in: M. Papadrakakis et al. eds, Computational Structural Dynamics and Earthquake Engineering, CRC Press, Balkema and Taylor & Francis, London, Ch. 11, 165-178, 2009.

[22] Papachristou, S. Salonikidis, G.K. Tairidis and G.E. Stavroulakis, *Adaptive fuzzy control of smart structures - ANFIS*, 7th International Congress on Computational Mechanics (GRACM), Athens, Greece, 30 June - 2 July, 2010

[23] G.E. Stavroulakis, I. Papachristou, S. Salonikidis, I. Papalaios and G. K. Tairidis: *Neurofuzzy control for smart structures*, invited lecture CSC2011, The Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, Chania, Greece, 6-9 September, 2011

[24] G.E. Stavroulakis, I. Papachristou, S. Salonikidis, I. Papalaios, G.K Tairidis, *Neurofuzzy Control for Smart Structures*, in Y. Tsompanakis and B.H.V. Topping eds, Soft Computing Methods for Civil and Structural Engineering, Saxe-Coburg Publications, Stirlingshire, UK, Ch. 7, 149-172, 2011

[25] G.K. Tairidis, I.Papachristou, M.Katagas and G.E. Stavroulakis, *Neuro - fuzzy control of smart structures*, 10th HSTAM International Congress on Mechanics, Chania, 25-27 May, 2013.

[26] G. Tairidis, G. Foutsitzi, G. Tsagkaris, G.E. Stavroulakis, *Fuzzy Control for Vibration Suppression of Smart Plates*, in B.H.V. Topping, P. Iványi, (Editors), "Proceedings of the Twelfth International Conference on Computational Structures Technology", Civil-Comp Press, Stirlingshire, UK, Paper 60, 2014

[27] G. Tairidis, G. Foutsitzi, P Koutsianitis and G.E. Stavroulakis, *Fine tuning of a fuzzy controller for vibration suppression of smart plates using genetic algorithms*, Advances in Engineering Software, 2016.doi: 10.1016/j.advengsoft.2016.01.019

[28] G.K. Tairidis, G.A. Foutsitzi, P.I. Koutsianitis and G.E. Stavroulakis, *Fine tuning of fuzzy controllers for vibration suppression of smart plates using particle swarm optimization*, 8th GRACM International Congress on Computational Mechanics, Volos, Greece 12 - 15 July, 2015.

[29] M. Marinaki, Y. Marinakis and G.E. Stavroulakis, *Fuzzy control optimized by PSO for vibration suppression of beams*, Control Engineering Practice, 18(6), 618-629, 2010.

[30] M. Marinaki, Y. Marinakis and G.E. Stavroulakis, *Fuzzy control optimized by a multi-objective particle swarm optimization algorithm for vibration suppression of smart structures*, Structural and Multidisciplinary Optimization, 43(1), 29-42, 2011

[31] M.Marinaki, Y. Marinakis, G. E. Stavroulakis, *Fuzzy control optimized by a Multi-Objective Differential Evolution algorithm for vibration suppression of smart structures*,Computers & Structures, Volume 147, 126-137, 2015

[32] G.K. Tairidis, G.A. Foutsitzi, P. Koutsianitis, G. Drosopoulos and G.E. Stavroulakis, *Design and testing of fuzzy controllers on smart structures in the pres-*

ence of delamination, Proceedings of the Fourth International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, Y. Tsompanakis, J. Kruis and B.H.V. Topping, (Editors), Civil-Comp Press, Prague, Czech Republic, 1-4 September, 2015.

[33] P. Koutsianitis, G.A. Drosopoulos, G.K. Tairidis and G.E. Stavroulakis, *Optimally tuned fuzzy control for smart, possibly damaged piezocomposites*, PCM-CMM-2015 3rd Polish Congress of Mechanics & 21st Computer Methods in Mechanics, 8-11 September 2015, Gdansk, Poland

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

DEPENDENCE OF THE MECHANICAL PROPERTIES OF PENTAMODE MATERIALS ON THE LATTICE MICROSTRUCTURE

Ada Amendola[1], Francesco Fabbrocino[2], Luciano Feo[1], Ferdinando Auricchio[3], Fernando Fraternali[1]

[1] Department of Civil Engineering, University of Salerno
Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy
e-mail: {adaamendola1,l.feo,f.fraternali}@unisa.it

[2] Department of Engineering, Pegaso University
Piazza Trieste e Trento 48, 80132 Naples, Italy
francesco.fabbrocino@unipegaso.it

[3]Department of Civil Engineering and Architecture, University of Pavia
via Ferrata 3, 27100 Pavia, Italy
auricchio@unipv.it

Keywords: Pentamode Lattices, Additive Manufacturing, Effective Elastic Moduli, Inelastic Response

Abstract *Pentamode materials consist of diamond-like lattices featuring five soft modes of deformation. Such materials have been proposed for transformation acoustics and elasto-mechanical cloak, but their potential in different engineering fields is still only partially explored. One characteristic in particular appears very appealing in this type of lattices: the dependency of the mechanical properties of the structure on the microstructure of the lattice rather than on the chemical nature of the employed materials. This characteristic implies the possibility to easily tune the effective mechanical properties of pentamode lattices to comply with the specific demands of the material at the meso- and macro-scales, with the aim of designing novel engineering devices, such as, e.g., innovative seismic isolators. The present work includes numerical and experimental studies investigating the dependence of the mechanical properties of pentamode materials on aspect parameters at the microscopic and macroscopic scales. It is shown that such materials exhibit elastic and post-elastic responses similar to those of elastomeric bearings alternating rigid layers and soft rubber layers.*

1 INTRODUCTION

Extremal materials that are receiving increasing interest are the so-called pentamode lattices, which consist of diamond-like lattices featuring five soft modes of deformation (unit cell with four rods meeting at a point) [1]. Such lattices exhibit very low shear moduli (theoretically equal to zero) [1], and may be able to stop or dramatically attenuate shear waves [2]. Physical models of pentamode materials have been fabricated through additive manufacturing (AM) techniques over the last few years, both at the macro- [3] and at the micro-scale [4]. Schittny et al. [3] have studied the experimental behavior of macroscopic, polymeric samples of pentamode lattices in the elastic regime: the results obtained by such authors prove that the elastic moduli of pentamode materials are strongly related to the geometry of the lattice micro-structure, being markedly affected by the dimensions of the rods forming the lattice, and particularly sensitive to the ratio between the diameter d of the connections between the rods and the lattice constant a. The experimental Young's modulus E of the lattice has been found approximately three times stiffer than the experimental shear modulus G. The results presented in [3] also show that the ratio between the bulk modulus B and the shear modulus G strongly increases by reducing the contact area between the rods. Similar results have been found at the microscale by Kadic et al. [4]. Polymeric samples of pentamode lattices have been fabricated by such authors using dip-in direct-laser-writing (DLW) optical lithography, a technique well suited to build three-dimensional structures across a range of scales from micro to small [5]. Experimental and numerical results given in [3] and [4] demonstrate that the mechanical behavior of pentamode lattices replicates that of fluids in the limit of $d \to 0$, when the shear modulus G tends to zero. For finite, nonzero values of d, it has been found that the B/G ratio is extremely high, even if G is positive and nonzero [3][4]. Because of their unusual mechanical features, pentamode materials have been proposed for transformation acoustics and elasto-mechanical cloak (refer, e.g., to the recent paper [6] and the references therein), but their potential in different engineering fields is still only partially explored.

The present study investigates the elastic and post-elastic responses of additively manufactured pentamode lattices confined between stiffening plates. We analyze metallic samples obtained through additive manufacturing by Electron Beam Melting (EBM) of a powder of the titanium alloy Ti-6Al-4V [7][8][10], for different aspect ratios of the unit-cell, and two macroscale aspect ratios (slender and thick samples). We find that the confinement of pentamode lattices between stiffening plates greatly affects the elastic response of the overall structure, compared with the case of unconfined, infinite pentamode lattices [11]. The obtained results highlight several similarities between the elastic response of confined pentamode materials and the analogous response of elastomeric bearings composed of rigid steel or fiber-reinforced composite layers and soft layers of natural or synthetic rubber [12]-[18]. Concerning the post-elastic response, we observe that the examined pentamode materials feature acceptable energy dissipation capacity, and ranges of supplemental damping in line with values of common isolation devices. The remainder of the paper is structured as follows.

We begin in Sect. 2.1 by illustrating the EBM-manufactured pentamode materials analyzed in the present study. Next, we provide a description of the in-house experimental setup employed to subject such materials to lateral and vertical force-displacement tests (Sect. 2.2). Section 3 presents the results of laboratory tests aimed at determining the elastic moduli (Sect. 3.1) and the post-elastic energy dissipation properties (Sect. 3.2) of the analyzed materials. Concluding remarks and directions for future research are presented in Sect. 4.

2 MATERIALS AND METHODS

2.1 Physical models of confined pentamode materials

We manufactured pentamode lattices confined by stiffening plates and made by the titanium alloy Ti-6Al-4V (hereafter simply denoted by Ti6Al4V) through the Arcam S12 EBM facility at the Department of Materials Science and Engineering, University of Sheffield. Such an additive manufacturing (AM) technology allows the manufacture of features with size down to 0.4 mm by progressively depositing, heating and melting layers of Ti6Al4V powder, with the melted regions in each layer defined according to a CAD model of the specimen to be manufactured [7][8]. It is worth noting that the size of the specimen designed by CAD does not correspond perfectly to the built object. The beam scan strategies and the surface roughness will result in larger (in diameter) printed members [10]. The main properties of the employed titanium alloy, when in the fully dense state, are given in Table 1 [19].

mass density [g/cm^3]	4.42
yield strength [MPa]	910.00
Young's modulus [GPa]	120.00
Poisson's ratio	0.342

Table 1. Main physical and mechanical properties of the fully dense isotropic polycrystalline Ti6Al4V titanium alloy [19].

Figure 1 shows the extended face-centered-cubic (fcc) unit cell of a pentamode lattice formed by sixteen rods that are composed of two truncated bi-cones featuring large diameter D at the mid-span and small diameter d at the extremities [2][3][4][5][6]. Upon selecting the lattice constant $a = 30$ mm and $D = 2.72$ mm ($D/a \approx 9\%$), we manufactured pentamode specimens with the unit cell shown in Figure 1, using three different values of d: $d_1 = 0.49$ mm ($d_1/a = 1.6\%$); $d_2 = 1.04$ mm ($d_2/a = 3.5\%$); and $d_3 = 1.43$ mm ($d_3/a = 4.8\%$, cf. Table 2). It is worth remarking that the limit $d/a \to 0$ corresponds to a perfectly pin-jointed lattice (stretching-dominated response), while the case with $d/a > 0$ corresponds to a lattice featuring nonzero bending rigidities of nodal junctions and rods, which deforms through both stretching of rods and bending of rods and nodes [11].

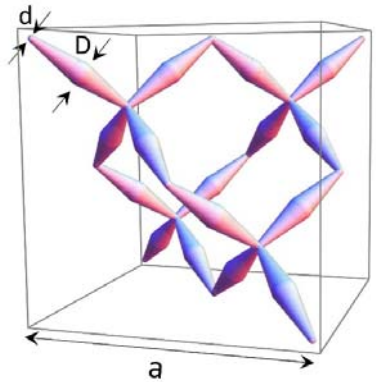


Figure 1: Extended fcc unit cell of the pentamode lattices analyzed in the present study.

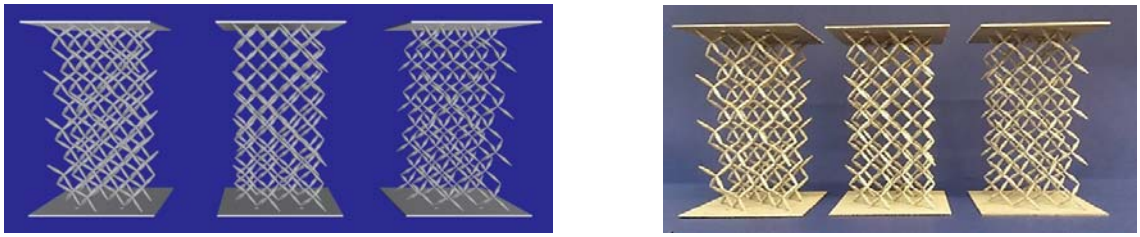
	a [mm]	D [mm]	d_1 [mm]	d_2 [mm]	d_3 [mm]
Built size	30	2.72	0.49	1.04	1.43
(CAD size)	(30)	(2.71)	(0.45)	(0.90)	(1.35)

Table 2: Geometrical parameters of the EBM built pentamode materials (CAD sizes in brackets).

We additively manufactured two different sets of pentamode specimens:

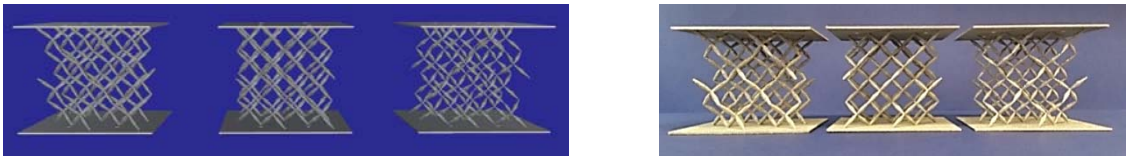
(a) *"Slender Pentamode Materials"* (SPM): obtained by replicating the extended fcc unit cell of Figure 1 $2x2$ times in the horizontal plane , 4 times along the vertical axis, and confining the $2x2x4$ lattice between Ti6Al4V plates with 80 mm edge and $t_p = 1$ mm thickness (see Figure 2);

(b) *"Thick Pentamode Materials"* (TPM): obtained by replicating the extended fcc unit cell of Figure 1 $2x2$ times in the horizontal plane, 2 times along the vertical axis, and confining the $2x2x2$ lattice between the above mentioned plates (see Figure 3).



(a) (b)

Figure 2: CAD (a) and EBM (b) models of slender (SPM) specimens ($2x2x4$ extended unit cells).



(a) (b)

Figure 3: CAD (a) and EBM (b) models of thick (TPM) specimens ($2x2x2$ extended unit cells).

Hereafter, we name SPM1, SPM2 and SPM3 the SPM-specimens with $d = d_1$, $d = d_2$, and $d = d_3$, respectively. Similarly, TPM1, TPM2 and TPM3 $d = d_1$, $d = d_2$, and $d = d_3$, respectively. We built three different SPM specimens and one TPM specimen for each analyzed d/a ratio.

It is well known that the EBM manufacturing process creates solid material with some level of porosity (typically below 0.2% if suitable parameters for the alloy are used in the machine) instead of fully dense materials for both solid structures [9] and those intended to contain free space [8]. Based on a recent, detailed study of porosity in EBM titanium made with the same conditions used for the current material [9], we hereafter assume the internal porosity of such a material is of the order of 0.2%, and its Young modulus E_m is therefore approximately equal to that of the fully dense Ti6Al4V alloy (Table 1).

2.2 Experimental setup

An in-house experimental setup was assembled to apply lateral force-displacement histories under constant vertical load on the pentamode specimens described in the previous section (Figure 4).

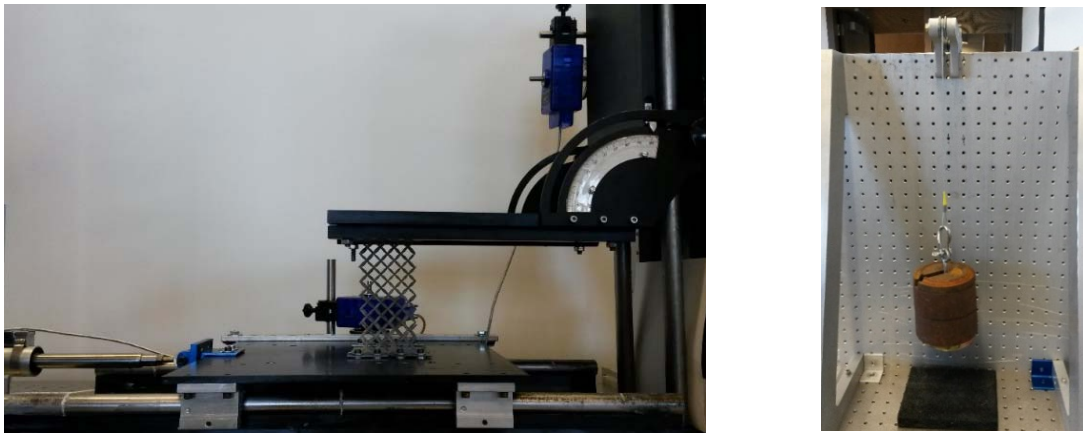


Figure 4: Left: overall view of the experimental setup. Right: counter balancing system introduced to remove platen self-weight.

An L-shaped plate, sliding vertically over linear bearings, allows the application of the vertical load through calibrated weights. The selfweight of the plate is removed with a counter balancing system of pulley and weights (Figure 4). The top base of the specimen is firmly connected to the plate by clamping strips located all around the base edges. The same clamping mechanism is applied to the bottom base of the specimen that is connected to the horizontally sliding table (Figure 5). Horizontal displacements are applied through an actuator connected to the sliding table and reacting against a solid metal block. The force rating of the actuator is 300 lb (1.3 kN), with a displacement range of +/-100 mm. Forces are acquired by a load cell installed between the back of the actuator and the reaction block. Vertical and lateral displacements are measured with rotary motion sensors.

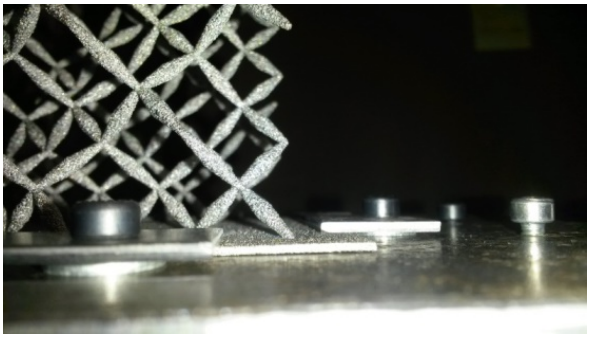


Figure 5: Connection of the specimen to the sliding table.

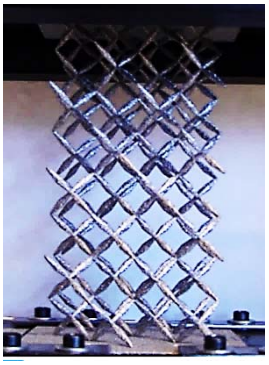
The same setup allows the application of the vertical load through calibrated weights while no horizontal motion is applied. With this procedure compression tests were completed by manual application of weights and automatic recording of vertical displacements of the plate. Preliminary to any actual test, the setup was subjected to a shakedown phase to assess the frictional force components and to verify alignments and parallelisms. The test data were consistently processed with the removal of friction from the testing equipment.

3 EXPERIMENTAL RESULTS

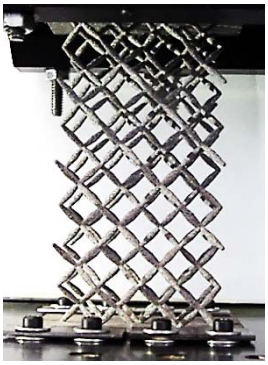
The present section illustrates the results of the laboratory tests that we conducted to characterize the mechanical response of the pentamode materials illustrated in Sect. 2.1. We begin by estimating the elastic moduli of slender and thick pentamode materials through cyclic lateral force-displacement tests at constant vertical loading, and monotonic vertical force-displacement tests under zero horizontal force (Sect.3.1). Next, we investigate the inelastic response of the analyzed materials via (cyclic and monotonic) lateral force-displacement tests conducted up to specimen's failure (Sect. 3.2). Figure 6 illustrates pictures of selected slender and thick specimens extracted from in-situ videos of lateral force-displacement tests.



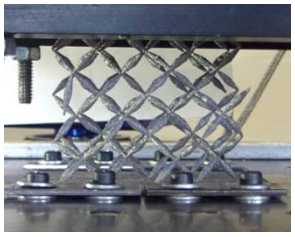
(a): SPM1



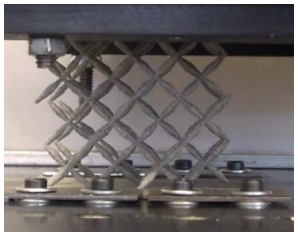
(b): SPM2



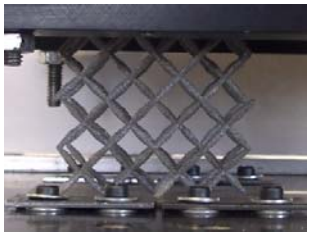
(c): SPM3



(d): TPM1



(e): TPM2



(f): TPM3

Figure 6: Frames from in-situ videos of lateral displacement tests on SPM and TPM specimens.

3.1 Estimation of the elastic moduli

The estimation of the elastic moduli of slender and thick specimens was conducted through cyclic lateral force (F_h) vs. lateral displacement (δ_h) tests in displacement control at the displacement rate $\dot{\delta}_h = 0.8$ mm/sec (hereafter also referred to as "*shear tests*"). Such a study considered three different (constant) values of the applied vertical load: F_v = 22.67 N, F_v = 45.34 N, and F_v = 68.00 N, and one specimen for each examined d/a ratio (cf. Figure 7-Figure 10; Table 3-Table 5). The analyzed values of the vertical load were accurately chosen so that to not cause permanent deformation of the specimens. We also conducted cyclic F_h-δ_h tests at the strain rates of 0.8 mm/sec and 2.4 mm/sec on the specimen SPM3 (Table 4), and vertical force (F_v) vs. vertical displacement (δ_v) compression tests on SPM and TPM specimens, on considering one specimen for each examined d/a ratio ("*compression tests*", cf. Table 3 and Table 5).

3.1.1 Slender specimens

Figure 7 illustrates the results of cyclic tests conducted on slender specimens for constant F_v = 45.34 N, and lateral displacements $\delta_h \approx \pm 3mm$ (three cycles for each specimen at $\dot{\delta}_h = 0.8$ mm/sec). The results in Figure 7 highlight an albeit limited hysteretic response of the SPM specimens in the examined rage of lateral displacements (especially in the case of the SPM3 specimen). Such a phenomenon is explained by internal stress intensification effects causing pseudo-elastic response, due to micro-plasticity events, before macro-yielding [23]. This effect would be increased in lattice samples as the structure results in areas of stress concentration being present in the sample, with the potential for the stress to reach high levels locally, while the specimen is macroscopically not highly loaded. Comparing the results in Figure 7 and the envelopes of cyclic and monotonic tests provided in Sect. 3.2 (cf. Figure 11 and Table 6), we note lateral displacements in the range $|\delta_h| \leq 3mm$ reached in Figure 7 are well below the deflection points of the F_h-δ_h curves in in Figure 11. Hereafter, we refer to such points where there is large scale departure from linearity in the force-displacement behavior as "macro-yield" points [23]. Very similar results were obtained for F_v = 22.67 N and F_v = 68.02 N. We refer to the line connecting the extreme points of the cyclic lateral force displacements curves as the *"effective"* secant response of the specimen [12]. Similarly, we let $K_{h,eff}$ denote the effective (secant) stiffness defined as follows

$$K_{h,eff} = \frac{\left|F_h^{+}\right| + \left|F_h^{-}\right|}{\left|\delta_l^{+}\right| + \left|\delta_l^{-}\right|} \tag{1}$$

where the symbols + and – denote the maximum and minimum value, respectively.
Table 3 shows the mean values and standard deviations of the experimental values of the secant stiffness $K_{h,eff}$, which were obtained for the SPM specimens under the analyzed values of the vertical load F_v, and the strain rate $\dot{\delta}_h = 0.8$ mm/sec. The same table also shows the values of the effective shear modulus defined as follows

$$G_{eff} = \overline{K}_{h,eff} \cdot \frac{h}{A} \tag{2}$$

where $\overline{K}_{h,eff}$ denotes the mean value of $K_{h,eff}$; $h = 4a + t_p$ denotes the total height of the specimen (including the heights of the pentamode lattice and the terminal bases); and $A = 2a \times 2a$ denotes the lattice covered area of the terminal plates. The results in Figure 7 and Table 3 highlight a marked increase of $K_{h,eff}$ with increasing values of the "microstructure" aspect ratio d/a. Such a phenomenon can be ascribed to the higher bending stiffness of the lattice with larger nodal junctions [11]. A small dispersion of $K_{h,eff}$ with the vertical load F_v is observed in the displacement regime $|\delta_h| \leq 3mm$.

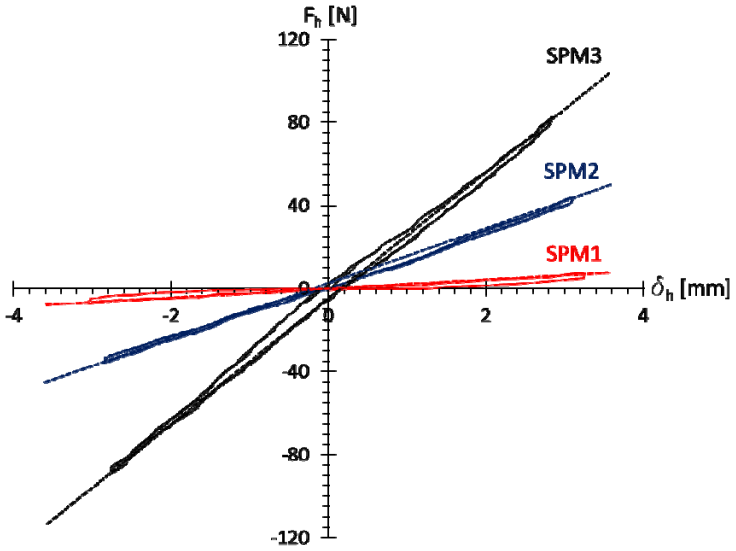


Figure 7: Lateral force-displacement curves under constant vertical load ($F_v = 45.34$ N) for slender specimens (dashed lines: effective secant responses).

Table 3 includes data extracted from compression tests, which refers to the effective vertical stiffness $K_{v,eff}$ defined as follows

$$K_{v,eff} = \frac{F_v^{\max}}{\delta_v^{\max}} \tag{4}$$

where $F_v^{\max}$ denotes the maximum applied vertical load, and $\delta_v^{\max}$ denotes the corresponding vertical displacement. Table 3 shows statistics of the experimental values of the vertical secant stiffness $K_{v,eff}$ obtained for each SPM specimen through three different measurements, and the values of the effective compression modulus defined as follows

$$E_{eff} = \overline{K}_{v,eff} \cdot \frac{h}{A} \tag{5}$$

where $\overline{K}_{v,eff}$ denotes the mean value of $K_{v,eff}$. By examining the results of shear and compression tests for current specimens, we find that the E_{eff}/G_{eff} ratio is equal to 7.12 in SPM1, 7.61 in SPM2 and 7.90 in SPM3. It is worth observing that such a ratio is approximately equal to one for (unconfined) many-cells pentamode lattices, according to the theory of elastic networks presented in Ref. [11]. It is also interesting comparing our current findings with the numerical and experimental results presented by Schittny et al. in [3] for 3D printed pentamode lattices in polymeric materials with the same macroscopic aspect ratio ($2x2x4$ extended unit cells). Such lattices feature $a = 15$ mm, and exhibit similar microscopic aspect ratios, as compared to the current specimen, i.e.: $D/a \approx 9\%$, and d/a varying between 1.5% and 9.0%. Schittny et al. experimentally observe the following ratios between the Young modulus E and the shear modulus G: $E/G \approx 3.67$ for lattices with $d/a = 1.6\%$ (lattices with microscopic aspect ratios similar to SPM1 specimens); $E/G \approx 5.52$ for $d/a = 3.5\%$ (similar to SPM2), and $E/G \approx 5.21$ for $d/a = 4.8\%$ (similar to SPM3). Numerical results obtained by such authors through 3D finite element simulations [3][4] instead predict E/G ratios ranging

between 4.29 and 4.97 for the same lattices. The significant increase of the E_{eff}/G_{eff} ratio of SPM specimens over the E/G ratio predicted by the theory presented in [11] ($E/G \approx 1.0$) is explained by the stiffening effect played by the terminal bases that confine such specimens. For what concerns the increase of E_{eff}/G_{eff} over the E/G ratios given in [3], it is worth noting that the current metallic specimens feature larger and stiffer bases, as compared to the polymeric specimens studied Schittny et al. in [3].

We now pass to study the dependence of the effective moduli G_{eff} and E_{eff} on the solid volume fraction ϕ of the lattice (defined as the volume of rods per unit volume of material). We construct linear and quadratic fitting models of the experimental data presented in Table 3, via the NonlinearModelFit function of *Mathematica*® 10 (models of the form $k_1\phi$ and $k_2\phi^2$, respectively). The results in Figure 8 show that quadratic models better fit the scaling laws of G_{eff}/E_m and E_{eff}/E_m with ϕ, featuring remarkably lower Root Mean Square Errors (RMSE) than linear models. Such an observation leads us to conclude that SPM specimens exhibit bending-dominated response under the examined tests [20][21][22].

		SPM1	SPM2	SPM3
$\phi \times 10^3$		8.14	11.75	15.04
$K_{h,eff}$ (N/mm)	mean (std Dev)	2.10 (0.159)	13.23 (0.237)	31.38 (0.598)
G_{eff} (kPa)		70.43	444.77	1054.69
$K_{v,eff}$ (N/mm)	mean (std Dev)	14.92 (0.278)	100.69 (≈ 0)	242.95 (≈ 0)
E_{eff} (kPa)		501.49	3384.47	8165.70

Table 3: Solid volume fraction ϕ and experimental values of the effective horizontal stiffness $K_{h,eff}$; effective shear modulus G_{eff}; effective vertical stiffness $K_{v,eff}$; and effective compression modulus E_{eff} of SPM specimens.

Let us now investigate the influence of strain rate effects on the shear response of a SPM3 specimen, by comparing experimental values of $K_{h,eff}$ under cyclic shear tests featuring $\dot{\delta}_h = 0.8$ mm/sec and $\dot{\delta}_h = 2.4$ mm/sec. The results in Table 4 show that such values of $K_{h,eff}$ differ each other only by 2.6%, which suggests that the examined strain rates play a limited role of on the elastic lateral response of the current specimen.

	$\dot{\delta}_h = 0.8$ mm/sec	$\dot{\delta}_h = 2.4$ mm/sec
$K_{h,eff}$ (N/mm)	30.73	29.95

Table 4: Experimental values of $K_{h,eff}$ for specimen SPM3 under different strain rates ($F_v = 45.34N$).

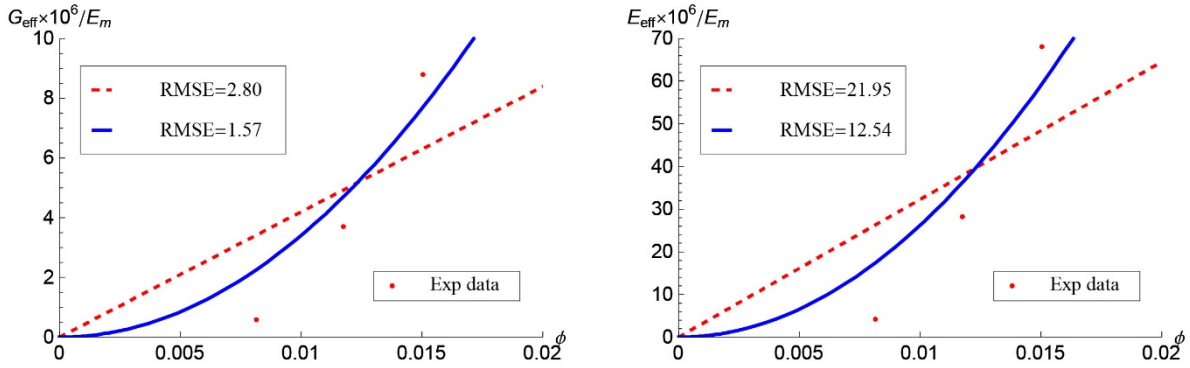


Figure 8: Linear and quadratic fitting models of the scaling laws of G_{eff}/E_m and E_{eff}/E_m with the solid volume fraction ϕ for SPM specimens.

3.1.2 Thick specimens

Figure 9 shows the results of cyclic lateral force-displacement tests performed on thick specimens at $\dot{\delta}_h = 0.8$ mm/sec, for $F_v = 45.34$ N (cf. also Table 5, which includes mean values and standard deviations of $K_{h,eff}$ for $F_v = 22.67$ N, $F_v = 45.34$ N, and $F_v = 68.02$ N).

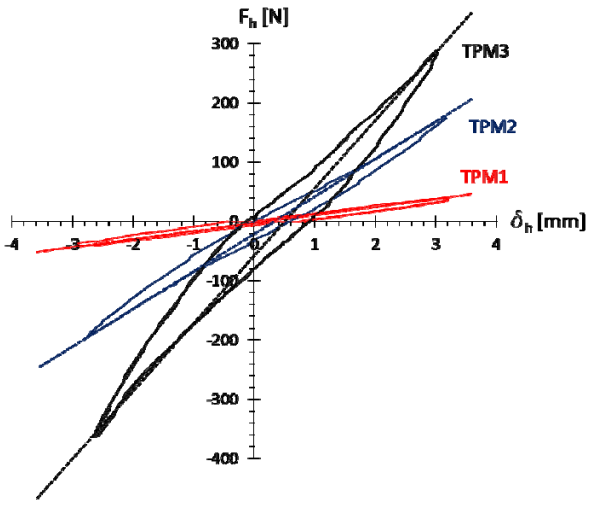


Figure 9: Lateral force-displacement curves under constant vertical load ($F_v = 45.34$ N) for thick specimens (dashed lines: effective secant responses).

The results in Figure 9 confirm the presence of the micro-scale hysteresis before macro-yielding already observed in slender specimens. Such a phenomenon is more pronounced in TPM specimens than slender specimens, especially in presence of large-size nodal junctions (TPM2 and TPM3). From the plots in Fig. 8 and the results in Table 5 we observe marked increases of $K_{h,eff}$ with increasing d/a ratios (i.e., when passing from TPM1 to TPM3), and small dispersions of such a quantity with the examined values of the vertical load F_v, as in the case of slender specimens. Estimates of the effective vertical stiffness $K_{v,eff}$ and the effective compression modulus E_{eff} for TPM specimens were obtained through vertical force-displacement tests performed as in the case of slender specimens (cf. Table 5).

		TPM1	TPM2	TPM3
$\phi \times 10^3$		8.14	11.75	15.04
$K_{h,eff}$ (N/mm)	mean (std Dev)	13.10 (0.117)	62.67 (0.270)	115.54 (0.662)
G_{eff} (kPa)		222.04	1061.90	1957.77
$K_{v,eff}$ (N/mm)	mean (std Dev)	38.83 (0.605)	175.37 (0.602)	503.89 (0.001)
E_{eff} (kPa)		657.98	2971.50	8538.12

Table 5: Solid volume fraction ϕ and experimental values of the effective horizontal stiffness $K_{h,eff}$; effective shear modulus G_{eff}; effective vertical stiffness $K_{v,eff}$; and effective compression modulus E_{eff} of TPM specimens.

The E_{eff}/G_{eff} ratio for thick specimens is equal to 2.96 in TPM1, 2.80 in TPM2 and 4.36 in TPM3. For what concerns the scaling laws of G_{eff} and E_{eff} with ϕ, the results in Figure 10 show that quadratic models better fit experimental data of TPM specimens than linear models, as we already observed in Sect. 3.1.1 (bending-dominated response [20][21][22]).

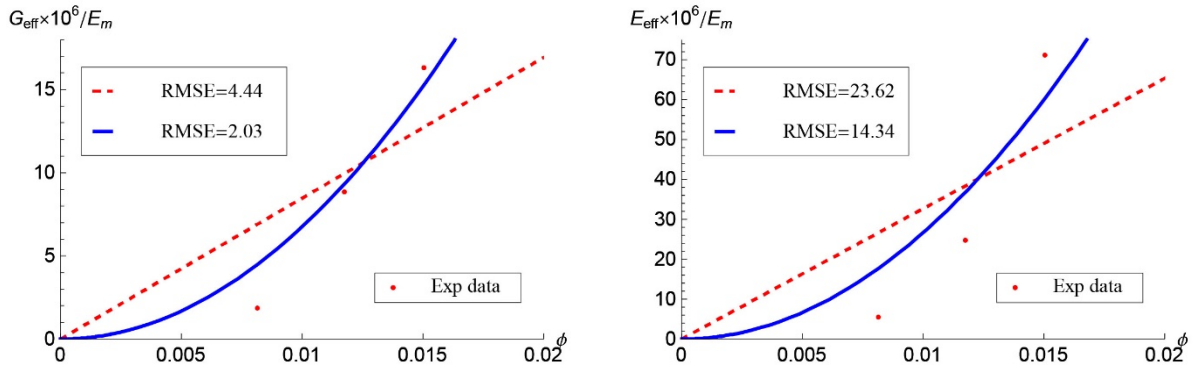


Figure 10: Linear and quadratic fitting models of the scaling laws of G_{eff}/E_m and E_{eff}/E_m with the solid volume fraction ϕ for TPM specimens.

3.2 Post-elastic response

We investigated the post-elastic response of SPM and TPM specimens by conducting cyclic lateral force - displacement tests in displacement control up to specimen's failure (under the vertical load $F_v = 44.45$N, and the strain rate $\dot{\delta}_h = 0.8$ mm/sec, cf. Figure 11-Figure 12). In the case of SPM specimens, we also performed monotonic lateral force-displacement tests up to specimen's failure (Figure 11). All such tests were carried out starting from the virgin state, for a single specimen of each microstructure aspect ratio d/a (cf. Sect. 2.1). Figure 11 shows the measured envelope force– displacement F_h-δ_h curves, where the labels a,b and c indicate the final point of the initial linear branch, the peak of horizontal force and the failure point, respectively. For these specimens, point c corresponds to the ultimate vertical bearing

capacity. The actual F_h-δ_h curves of the monotonic tests on SPM specimens showed marked oscillations in the unstable phase characterized by a decrease of F_h for increasing values of δ_h, due to progressive joint failure. The final collapse of the specimen occurred due to the rupture of several terminal nodes placed along the stretched diagonals (Figure 13). The results in Figure 11 highlight an initial hardening branch of the post-elastic response of each examined specimen. Such a branch is followed by a second post-elastic branch with negative tangent stiffness, which progressively leads the specimen to collapse. The coordinates of the above *a*-*b*-*c* points are provided in Table 6, together with the ductility values computed as follows

$$\mu = \frac{\delta_h - \delta_{h,el}}{\delta_{h,el}} \tag{6}$$

where δ_h is the current value of the lateral displacement, and $\delta_{h,el}$ is the lateral displacement in correspondence to point *a*.

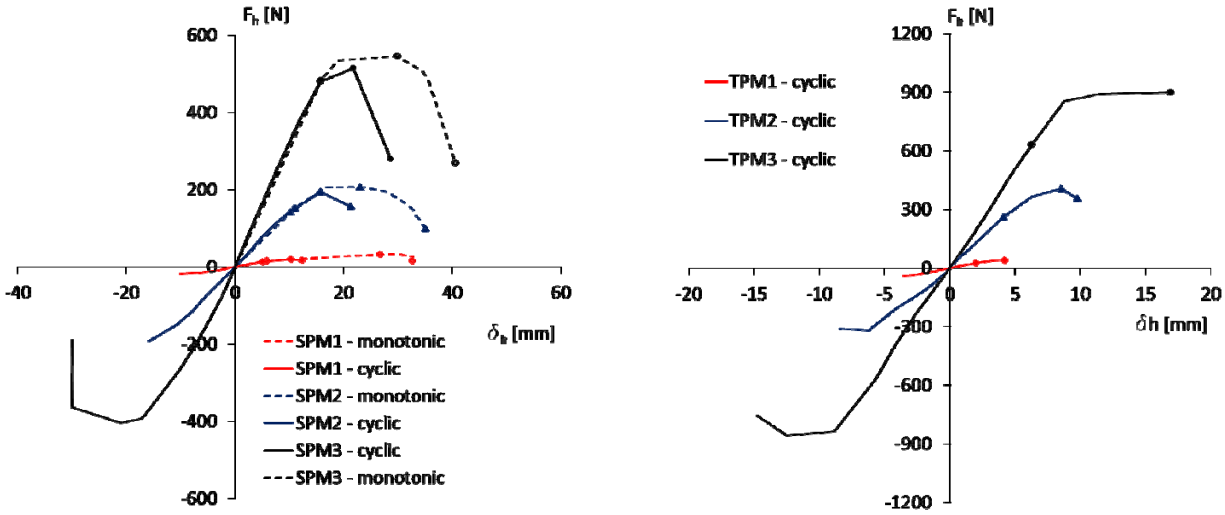


Figure 11: Measured envelops of cyclic (solid lines) and monotonic (dashed lines) lateral force-displacement tests.

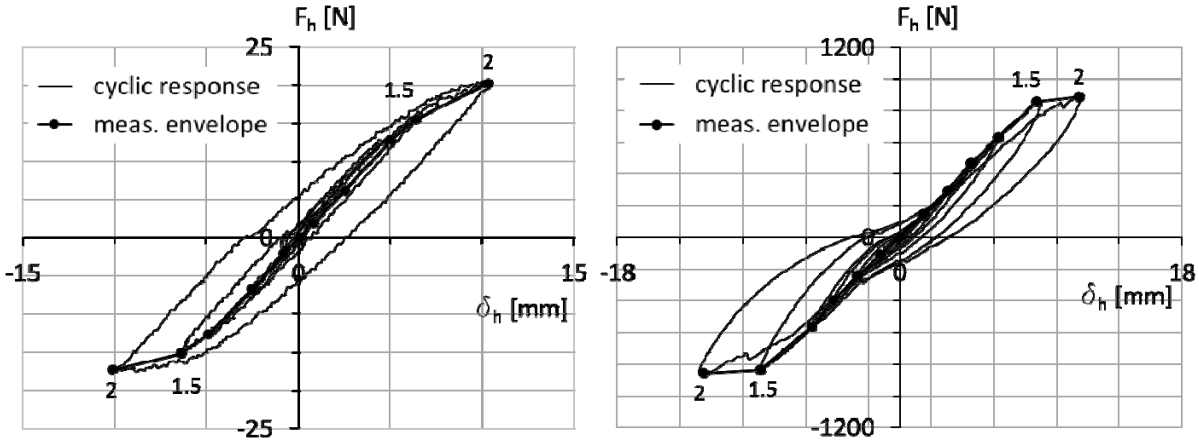
		F_a	δ_a	μ_a	F_b	δ_b	μ_b	F_c	δ_c	μ_c
		[N]	[mm]		[N]	[mm]		[N]	[mm]	
SPM1	monotonic	15.92	5.85	1.00	32.93	26.74	4.57	16.20	32.67	5.58
	cyclic	12.70	5.02	1.00	20.19	10.29	2.06	17.67	12.35	2.47
SPM2	monotonic	152.38	10.98	1.00	206.83	22.97	2.09	100.01	35.03	3.19
	cyclic	143.10	10.23	1.00	195.12	15.75	1.54	157.03	21.26	2.07
SPM3	monotonic	485.01	15.79	1.00	545.56	29.92	1.89	268.5	40.64	2.57
	cyclic	480.12	15.70	1.00	514.00	21.70	1.38	280.00	28.65	1.82
TPM1	cyclic	25.22	2.00	1.00	40.20	4.17	2.08	40.20	4.17	2.08
TPM2	cyclic	262.18	4.11	1.00	407.34	8.53	2.07	357.45	9.76	2.37
TPM3	cyclic	467.93	4.55	1.00	899.02	16.88	2.70	899.02	16.88	2.70

Table 6: Coordinates and ductility values of noticeable points of the curves in Figure 11.

The cyclic tests in Figure 12 where performed at controlled ductility on a SPM1 and a TPM3

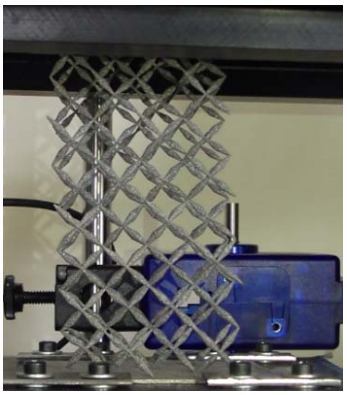
specimen, by letting the maximum lateral displacement to grow progressively in such a way that it results $\mu = 0.25$ in the first cycle; μ grows with step 0.25 in the next cycles up to $\mu = 1.00$; and, finally, μ grows with step 0.50 in the post-elastic cycles up to specimen's failure. Figure 11-left shows a comparison between the measured envelopes of cyclic tests and monotonic tests on SPM specimens, which highlights the fact that the envelopes of the monotonic tests always lie above those of the cyclic tests in the post-elastic range, due to (low-cycle) fatigue-damage [13].

Post-yield stiffness degradation is a common observation in a range of different tests and lattice designs of EBM manufactured Ti6Al4V porous structures (e.g [8][10]), due to progressive brittle fracture of individual struts of the lattice from the point of first macroscopic yield [7].

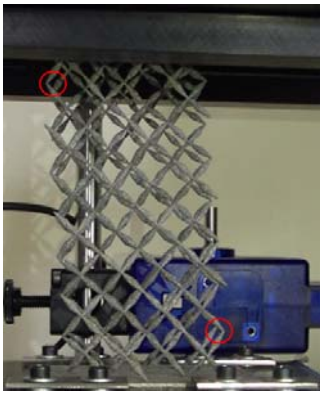


(a) (b)

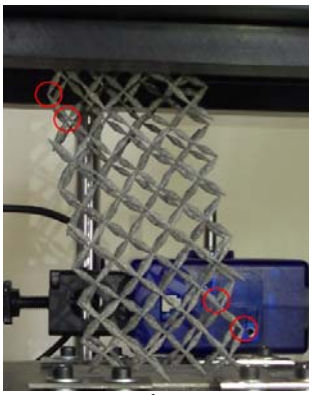
Figure 12: Lateral force - displacement responses of a SPM1 (a) and a TPM3 (b) specimen under cyclic tests at constant vertical load $F_v = 45.34$ N (solid curves: measured envelope curves; numbers indicate ductility).



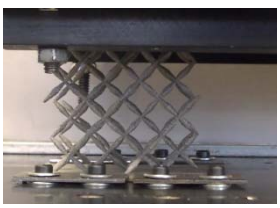
(a) SPM1 – *3rd elastic cycle* $\delta_h = 5.12$ mm, $\mu = 1.0$



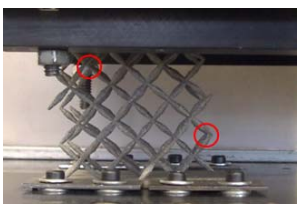
(b) SPM1 - *1st inelastic cycle* $\delta_h = 6.40$ mm, $\mu = 1.5$



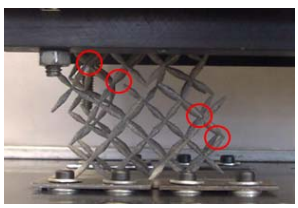
(c) SPM1 - *2nd inelastic cycle* $\delta_h = 10.29$, $\mu = 2.0$



(d) TPM3 - *4th elastic cycle* $\delta_h = 6.25$ mm, $\mu = 1.0$



(e) TPM3 - *1st inelastic cycle* $\delta_h = 8.74$ mm, $\mu = 1.5$



(f) TPM3 - *2nd inelastic cycle* $\delta_h = 11.45$ mm, $\mu = 2.0$

Figure 13. Frames from in-situ videos of lateral displacement tests on a SPM1 (a-b-c) and a TPM3 (d-e-f) specimen. The circled struts are affected by fracture damage.

Table 7 shows the Energy Dissipated per Cycle EDC the effective viscous damping ξ_{eff} ; the post-yield (K_d) vs. elastic (K_e) stiffness ratio $|K_d - K_e|/K_e$; and characteristic strength Q_d defined as force at zero horizontal displacement (cf. point 8.2.1.2.2 of [25]).

		EDC ($10^3\, kN \cdot mm$)	ξ_{eff} (%)	$\lvert K_d - K_e \rvert / K_e$ (%)	Q_d (N)
SPM1	$\mu_{\max} = 1.5$	191.44	25.11	25.73	1.52
	$\mu_{\max} = 2.0$	406.27	31.12	57.44	5.27
TPM3	$\mu_{\max} = 1.5$	$12.97 \cdot 10^3$	28.02	10.07	93.12
	$\mu_{\max} = 2.0$	$18.86 \cdot 10^3$	30.12	88.73	95.01

Table 7: Energy dissipation per cycle (E_{diss}); effective viscous damping (ξ_{eff}); post-yield vs. elastic stiffness ratio ($|K_d - K_e|/K_e$); and characteristic strength (Q_d) at different post-elastic cycles of lateral force – displacement test on SPM1 and TPM3 specimens.

The effective damping in Table 7 is defined in line with the European standard on anti-seismic devices EN 15129:2009 [25], and corresponds to the energy dissipated E_{diss} through the following formula

$$\xi_{eff} = \frac{E_{diss}}{2\pi K_{h,eff}\, \delta_{h,\max}^2} \tag{5}$$

where $K_{h,eff}$ is the effective secant stiffness, and $\delta_{h,\max}$ is the maximum lateral displacement of the current cycle (see point 3.1.10 of [25]). The results in Table 7 highlight progressive increases of ξ_{eff} , $|K_d - K_e|/K_e$, and Q_d with the amplitude of the maximum lateral displacement (i.e., with $\mu_{\max}$), due to progressive fracture of lattice struts [8] (cf. Figure 13). The results in terms of energy dissipated per cycle (and effective damping) while indicating a progressive increase with displacement show also values of supplemental damping (25% to 30%) that are suitable for seismic isolation applications. The failure mechanisms in Figure 13 indicate a quite symmetric pattern of broken elements between top and bottom of the specimen. It was also observed that during the tests the failure of struts happened in a rather smooth progression and in expected locations. This performance new configurations that can significantly increase the required structural ductility to be forseen.

4 CONCLUSIONS

We have experimentally investigated the response of EBM printed Ti6Al4V specimens of pentamode materials confined between stiffening plates, under lateral and vertical force-displacement tests. The examined tests were aimed at researching the effective shear and compression properties of such systems, and their inelastic response under cyclic and monotonic loading histories. We have examined a collection of specimens that differ in microstructural and macrostructural aspect ratios, which are related to the size of the nodal junctions

(microstructure aspect ratio) and the ratio between the number of unit cells placed in the vertical and horizontal directions (macrostructure aspect ratio).

Several results of the present study highlight analogies between the mechanical response of the analyzed systems and that of elastomeric bearings consisting of alternating layers of synthetic or natural rubber and stiffening layers made of steel plates or fiber-reinforced composites, which are commonly used as bridge support devices or seismic or vibration isolators for buildings [12]-[18]. We have observed that the ratio between the effective compression modulus and the effective shear modulus of the analyzed pentamode materials is markedly affected by the presence of the stiffening plates, as compared to available theoretical predictions of the ratio of unconfined pentamode lattices [11]. Such a stiffening effect is larger in the case of slender specimens and large-size nodal junctions, since the above ratio varies between 7.93, in case of the SPM3 specimen, and 2.96, in case of the TPM1 specimen (cf. Sect. 3.1). It replicates the stiffening role played by steel layers in elastomeric bearings. It is worth noting, however, that the absolute vertical and lateral stiffness coefficients of the examined pentamode materials increase when passing from slender to thick specimens (cf. Sect. 3.1).

Despite the very preliminary nature of this study on feasibility for base isolation applications, some observations support the potential of pentamode structures for use in anti-seismic system development and/or as shear-wave band gap systems [2]. The strict dependency of the response by the geometry of the lattice allows the design of structures of controlled performance. The ranges of supplemental damping observed for the specimens are well in line with values of common isolation devices. The levels of energy dissipated is acceptable and could be significantly improved by reducing material porosity, using different construction techniques, materials, and/or through the insertion of dissipative lead-cores within the pentamode isolator [12][15]. Additional horizontal ductility and vertical stiffness is expected to be developed also using different configurations that will provide more confinement to the single layers of the pentamode structure.

We address the design and modeling of highly dissipative pentamode bearings to future work, on employing either manual assembling procedures [26], layered structures showing different materials, prestressed members and/or lattice geometries in different layers [27][28], and additive manufacturing techniques in polymeric and metallic materials [3][4][7][29][30]. Additional directions of future work include the modeling of fracture damage of rods, via variational fracture [31], as well as numerical and experimental studies on the use of pentamode bearings for the seismic isolation of buildings, bridges, and historical constructions [32]-[36].

REFERENCES

[1] G.W. Milton, A.V. Cherkaev, Which Elasticity Tensors are Realizable? *Journal of Engineering Materials and Technology*, **117**(4), 483-493, 1995.

[2] A. Martin, M. Kadic, R. Schittny, T. Bückmann, M. Wegener, Phonon band structures of three-dimensional pentamode metamaterials, *Physical Review Letters B*, **86**, 155116, 2010.

[3] M. Schittny, T. Bückmann, M. Kadic, M. Wegener, Elastic measurements on macroscopic three-dimensional pentamode metamaterials. *Applied Physics Letters*, **103**, 231905, 2013.

[4] M. Kadic, T. Bückmann, N. Stenger, M. Thiel, M. Wegener, On the practicability of pentamode mechanical metamaterials. *Applied Physics Letters*, **100**, 191901, 2012.

[5] T. Buckmann, N. Stenger, M. Kadic, J. Kaschke, A. Frölich, T. Kennerknecht, C. Eberl, M. Thiel, M. Wegener, Tailored 3D Mechanical Metamaterials Made by Dip-in Direct-Laser-Writing Optical Lithography. *Advanced Materials,* **24**(20), 2710-4, 2012.

[6] T. Bückmann, M. Thiel, M. Kadic, R. Schittny, M. Wegener, An elastomechanical unfeelability cloak made of pentamode metamaterials. *Nature Communications*, **5**, 4130, 2014.

[7] A. Amendola, E. Hernandez-Nava, R. Goodall, I. Todd, R.E. Skelton, F. Fraternali, On the additive manufacturing and testing of tensegrity structures. *Composite Structures*, **131**, 66-71, 2015.

[8] E. Hernandez-Nava, C.J. Smith, F. Derguti, S. Tammas-Williams, F. Leonard, P.J. Withers, I. Todd, R. Goodall, The effect of density and feature size on mechanical properties of isostructural metallic foams produced by additive manufacturing. *Acta Materialia*, **85**, 387-95, 2015.

[9] S. Tammas-Williams, H. Zhao, F. Léonard, F. Derguti, I. Todd, P.B. Prangnell, XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by Selective Electron Beam Melting. *Materials Characterization*, **102**, 47-61, 2015.

[10] W. van Grunsven, E. Hernandez-Nava, G.C. Reilly, R. Goodall, Fabrication and mechanical characterisation of titanium lattices with graded porosity. *Metals*, **4**(3), 401-9, 2014.

[11] A.N. Norris, Mechanics of elastic networks. *Proceedings of the Royal Society of London A*, **470**(522):1-18, 2014.

[12] M.C. Constantinou, A.S. Whittaker, Y. Kalpakidis, D.M. Fenz, G.P. Warn, Performance of seismic isolation hardware under service and seismic loading, Technical Report MCEER-07-0012, 2007.

[13] R.I. Skinner, W.H. Robinson, G.H. McVerry, *An introduction to seismic isolation*. Wiley, 1993.

[14] J.M. Kelly, *Earthquake-resistant design with rubber*. London: Springer-Verlag, 1993.

[15] G. Benzoni , C. Casarotti, Effects of vertical load, strain rate and cycling on the response of lead-rubber seismic isolators. *Journal of Earthquake Engineering*, **13**(3):293-312, 2009.

[16] H. Toopchi-Nezhad, M.J. Tait, R.G. Drysdale, Bonded versus unbonded strip fiber reinforced elastomeric isolators: Finite element analysis. *Composite Structures*, **93**(2), 850-859, 2014.

[17] P.M. Osgooei, M.J. Tait, D. Konstantinidis, Three-dimensional finite element analysis of circular fiber- reinforced elastomeric bearings under compression. *Composite Structures,* **108**(1), 191-204, 2014.

[18] Osgooei PM, Tait MJ, Konstantinidis D. Finite element analysis of unbonded square fiber-reinforced elastomeric isolators (FREIs) under lateral loading in different directions. *Composite Structures*, **113**(1), 164-173, 2014.

[19] Metals Handbook, (2) Properties and selection: nonferrous alloys and special-purpose materials. ASM International, Metals Park, OH, 1990.

[20] L.R. Meza, S. Das, J.R. Greer, Strong, light weight, and recoverable three-dimensional ceramic nanolattices. *Science*, **6202**(345), 1322-1326, 2014.

[21] VS Deshpande, NA Fleck, MF Ashby. Effective properties of the octet-truss lattice material. *Journal of the Mechanics of Physics of Solids*, **22**, 409-428, 2001.

[22] T.A. Schaedler, A.J. Jacobsen, A. Torrents, A.E. Sorensen, J. Lian, J.R. Greer, L. Valdevit, W.B. Carter, Ultralight Metallic Microlattices. *Science*, **6058**(334), 962-965, 2011.

[23] R.K. Steele, A.J. McEvily, The high-cycle fatigue behavior of Ti-6Al-4V alloy. *Engineering Fracture Mechanics*, **8**, 31-37, 1976.

[24] I. Szachogluchowicz, L. Sniezek, V. Hutsaylyuk, Low cycle fatigue properties laminate AA2519-TI6AL4V. *Procedia Engineering*, **114**, 26-33, 2015.

[25] European Committee for Standardization. Anti-seismic devices, EN 15129. Brussels, Belgium, 2009.

[26] A. Amendola, G. Carpentieri, M. de Oliveira, R.E. Skelton, F. Fraternali, Experimental investigation of the softening stiffening response of tensegrity prisms under compressive loading. *Composite Structures*, **117**, 234-243, 2014.

[27] F. Fraternali, G. Carpentieri, A. Amendola, R.E. Skelton, V.F. Nesterenko, Multiscale tunability of solitary wave dynamics in tensegrity metamaterials, *Applied Physics Letters*, 105 (20), art. no. 201903, 2014.

[28] A. Amendola, C.J. Smith, R. Goodall, F. Auricchio, L. Feo, G. Benzoni, F. Fraternali, Experimental response of additively manufactured metallic pentamode materials confined between stiffening plates. *Composite Structures*, **142**, 254–262, 2016.

[29] R.E. Skelton, R. Montuori, V. Pecoraro, Globally stable minimal mass compressive tensegrity structures. *Composite Structures*, **141**, 346–354, 2016

[30] I. Farina, F. Fabbrocino, G. Carpentieri, M. Modano, A. Amendola, R. Goodall, L. Feo, F. Fraternali, On the reinforcement of cement mortars through 3D printed polymeric and metallic fibers. *Composites. Part B, Engineering*, 90, 76-85, 2016.

[31] F. Fraternali, Free discontinuity finite element models in two-dimensions for in-plane crack problems, *Theoretical and Applied Fracture Mechanics*, **47** (3), 274-282, 2007.

[32] A. Bossio, F. Fabbrocino, G.P. Lignola, A. Prota, G. Manfredi, Simplified model for strengthening design of beam-column internal joints in Reinforced Concrete Frames, *Polymers*, **7** (9), 1732-1754, 2015.

[33] G. Carpentieri, M. Modano, F. Fabbrocino, F. Fraternali, Optimal design and dynamics of truss bridges, *COMPDYN 2015 - 5th ECCOMAS* Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, 1731-1740, 2015

[34] F. Fraternali, G. Carpentieri, M. Modano, F. Fabbrocino, R.E. Skelton, A tensegrity approach to the optimal reinforcement of masonry domes and vaults through fiber-reinforced composite material. *Composite Structures*, **134**, 247-54, 2015.

[35] F. Fabbrocino, I. Farina, V.P. Berardi, A.J.M. Ferreira, F. Fraternali, On the thrust surface of unreinforced and FRP-/FRCM-reinforced masonry domes. *Composites. Part B, Engineering*, **83**, 297-305, 2015.

[36] M. Modano, F. Fabbrocino, A. Gesualdo, G. Matrone, I. Farina, F. Fraternali, On the forced vibration test by vibrodyne, COMPDYN 2015 - 5th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, 209-217, 2015.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

HOMOGENIZATION TECHNIQUES FOR ACCURATE AND APPROXIMATE ESTIMATES FOR OVERALL PROPERTIES OF MICROCRACKED VISCOELASTIC MASONRIES

A. Rekik[1], **A. Gasser**[1]

[1]Univ. Orléans, INSA-CVL, PRISME, EA 4229, 45072 Orléans, France
amna.rekik@ univ-orleans.fr, alain.gasser@ univ-orleans.fr

Keywords: Masonry, Creep, Microcraks, Homogenization techniques, Finite elements method.

Abstract. *Based on the coupling between brittle fracture mechanics and homogenization techniques, either analytical or finite elements homogenization, this work provides accurate (up to numerical errors) and approximate overall estimates for masonries accounting for their creep behaviour and a certain level of damage as it is the case for instance for refractory linings serving at high temperatures or middle-ages masonry building.*

1 INTRODUCTION

It is well known that creep phenomenon has a noticeable effects not only on new structures but also on historical monuments [20, 3, 17]. To model the creep behaviour of traditional mortar, various rheological models namely the USBR, Feng, Ross, typical and modified versions of the Burgers and Modified-Maxwell models may be investigated [5, 11]. On the other hand, there exist several approaches accounting for damage in viscoelastic materials [11, 12]. Indeed, the approach presented in [12] is based on a coupling between continuum damage mechanics and viscoelasticity through the generalized Kelvin-Voigt model. Accordingly, a three-dimensional phenomenological model was developed to describe the long-term creep of gypsum rock. The main disadvantage of this model is that it requires experimental investigation [11] or computational efforts to resolve nonlinear equation [12] function of internal damage variables.

In the literature, little attention is devoted to the prediction of the macroscopic creep behaviour of masonry. In this context, Brooks, Cecchi & Tralli, Cecchi & Taliercio and Taliercio [4, 7, 8, 22] proposed models to predict creep coefficients according to the properties of each masonry constituent. These models are based on analytical or numerical homogenization using the finite elements method (FEM) in order to deduce the macroscopic creep of undamaged (without cracks) masonry.

The objective of this study is to evaluate the effective and local behaviour of masonries exhibiting nonlinear behaviours mainly viscoelastic at short and/or long times especially when they are subjected to severe or long term loading such as historical monuments or refractory masonry linings working under high temperatures. More details about this analysis are provided in the references [13, 18].

2 METHODOLOGY AND HYPOTHESIS

In the present study, the coupling between the Griffith's theory and the dilute scheme [2] will be applied to provide the effective behaviour of a micro-cracked mortar.

As a first approach and for the sake of simplicity, only the mortar is assumed to be viscoelastic and microcracked. The distribution of microcracks is assumed to be isotropic. The overall creep behaviour of the mortar is obtained due to the coupling between the Griffiths theory [1] and mean-field homogenization [2]. At short and long terms, this procedure allows fast and easy approximation of elastic and viscous properties of the mortar assumed to follow the Generalized Maxwell rheological model. These properties which are explicit function of time and crack density parameter are determined without recourse to a heavy numerical inversion of the Laplace-Carson transform.

The second step of this work consists to carry out either a 'direct' (analytical or numerical) homogenization modeling of the periodic masonry or 'indirect' method based on two steps: (i) the differentiating of the mortar's constitutive law defining thus a thermo-elastic incremental behaviour and (ii) finite elements homogenization of the masonry periodic cell. The last step, as the 'direct' homogenization method, relies on the computation of the localization strain tensors in each brick and mortar in order to define overall stiffness and prestress in the periodic cell. More details about principles and results of these models are provided in [13, 18, 19].

For concision reason, this paper deals only with direct homogenization model coupling between brittle fracture theory in order to assess the effective microcracked mortar's behaviour and periodic homogenization. Basic steps followed by this model are summarized in Figure (2).

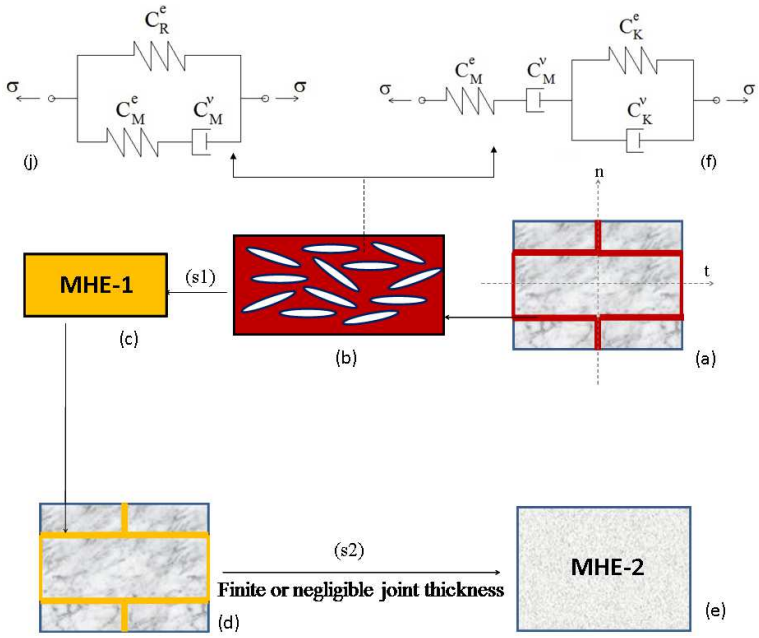


Figure 1: Main steps of the proposed model.

2.1 Microcracked mortar's effective behaviour

Let us firstly recall that the effective stiffness of an elastic porous medium with a homogeneous solid phase tensor C^s is:

$$\tilde{C} = C^s : (I - \phi A^p) \tag{1}$$

where A^p is the average of the strain localisation tensor $A(z)$ over the pore space Ω^p and ϕ is the porosity volume fraction. For a flat spheroid - usual $3D$ crack model - of aspect ratio $w << 1$, the Eshelby tensor S^E is a function of w. Accordingly the components A^p_{iikl} and A^p_{ikik} of the dilute localization tensor $A^p = (I - S^E)^{-1}$ in the pore are of the order of $1/w$ and therefore is the ratio of the normal strain ε_{nn} to the macroscopic strain E. Possible none negligible variations of $1/w$ is in contradiction with the assumption of linearity of the localization relationship. To overcome this difficulty, it was proposed [9] to consider the rate-type formulation of the problem i.e. the strain localization tensor should be replaced by a strain rate localization as in the following $\dot{\varepsilon}(z) = A(z) : \dot{E}$. Similarly, the rate-type formulation of the Eshelby problem for a spheroidal cavity reads:

$$\dot{\varepsilon}^p = (I - S^E(w))^{-1} : \dot{E} \tag{2}$$

Such hypothesis implies that the use of A^p in the homogenized stiffness (1) leads to an estimate of the tangent effective stiffness. Moreover, since the crack porosity ϕ is proportional to w [10], the tangent effective stiffness is mathematically independent of w. This renders the effective behaviour linear elastic. Note that the rate-type hypothesis is indispensable [10] to also avoid troubles related to possible large strain in the direction normal to the crack.
On the other hand, the extension of the linear homogenization schemes to non-aging viscoelasticity is based on the Laplace-Carson (LC) transform [21]. The effective stiffness $\tilde{C} =< C : A >$ becomes

$$\tilde{C}^* =< C^* : A > \tag{3}$$

in the LC space where C^* is the apparent elastic stiffness. The presence of microcracks implies the existence of nonlinearity at the local scale in the relationship between the crack strain and the

macroscopic strain. Accordingly the homogenization of a viscoelastic cracked medium is not as straightforward as (3). The basic idea consists in anticipating that both the microscopic strain field and the displacement discontinuity vectors $[u]_i$ linearly depend on the macroscopic stress. Such hypothesis, confirmed by [15], justifies the use of the LC transform which can be applied to the macroscopic strain related to the microscopic strain and the displacement discontinuity vectors $[u]_i$ between two lips of crack. In the framework of the stress-based dilute scheme [10], the displacement jump $[u]_i$ is linearly related to Σ. Indeed, the normal displacement jump (mode I) at the lips of a crack in an infinite matrix submitted to an isotropic asymptotic macroscopic stress $\Sigma^* = \Sigma^*\, i$ can be written:

$$[u_n]^* = \frac{4(1-\nu_s^*)}{\pi}\frac{\Sigma^*}{\mu_s^*}\sqrt{l^2-\rho^2}. \tag{4}$$

where l is the crack's radius and ρ is the position of a point M pertaining to the crack's plane. The tangential displacement jump (mode II) under an asymptotic shear stress $\Sigma^* = \Sigma^*\, n \otimes^s t$ where t is parallel to the crack's plane reads:

$$[u_t]^* = \frac{4(1-\nu_s^*)}{\pi(2-\nu_s^*)}\frac{\Sigma^*}{\mu_s^*}\sqrt{l^2-\rho^2}. \tag{5}$$

Under respectively an isotropic and deviatoric loading and assuming that all cracks have the same radius l, an integration over all orientations on the unit sphere yields the total crack contribution and allows the determination of the apparent effective bulk's and shear moduli as follows

$$\begin{array}{l} \dfrac{1}{\tilde{k}_{DL}^*} = \dfrac{1+d_cQ^*}{k_s^*} \quad \text{where} \quad Q^* = \dfrac{16}{9}\dfrac{(1-\nu_s^{*2})}{(1-2\nu_s^*)} \\ \dfrac{1}{\tilde{\mu}_{DL}^*} = \dfrac{1+d_cM^*}{2\mu_s^*} \quad \text{where} \quad M^* = \dfrac{32}{45}\dfrac{(1-\nu_s^*)(5-\nu_s^*)}{2-\nu_s^*} \end{array} \tag{6}$$

in which the crack density parameter and the symbolic Poisson's ratio of the safe matrix reads respectively: $d_c = Nl^3$ and $\nu_s^* = \frac{3k_s^*-2\mu_s^*}{6k_s^*+2\mu_s^*}$. Since the expressions (6) can not be satisfied rigorously, it was proposed in [15] to identify the best approximation of the effective behaviour in the class of Burgers' (Modified Maxwell's) model if the mortar in its safe state follows the Burgers' (Modified Maxwell's) model. The idea is to satisfy the series expansion of the dilute estimates of the bulk's (6)-a and shear (6)-b moduli to the first order at $p=0$ and $p\to\infty$ such that

$$\lim_{t\to\infty} f(t) = \lim_{p\to 0} f^*(p) \quad \text{et} \quad \lim_{t\to 0} f(t) = \lim_{p\to\infty} f^*(p) \tag{7}$$

The dilute symbolic moduli of a non-aging linear viscoelastic (n.a.l.v.) microcracked mortar following the MM's model can be approached by expressions similar respectively to those available for a safe mortar following the MM's rheological model

$$k_{MM}^* = k_R(d_c) + \frac{pk_M(d_c)\eta_M^s(d_c)/3}{k_M(d_c)+p\eta_M^s(d_c)/3} \quad \mu_{MM}^* = \mu_R(d_c) + \frac{p\mu_M(d_c)\eta_M^d(d_c)/2}{\mu_M(d_c)+p\eta_M^d(d_c)/2} \tag{8}$$

Following idea (7), it is possible to identify the MM's six parameters [14] and accordingly to determine an approximate creep function of a microcracked mortar which matrix follows the MM's model reads

$$\begin{array}{l} J_{MM}^{app}(t,dc) = \frac{1}{9k_R(d_c)}\left(1-\frac{k_M(d_c)}{(k_R(d_c)+k_M(d_c))}e^{-t/\tau_{MM}^s(d_c)}\right) \\ +\frac{1}{3\mu_R(d_c)}\left(1-\frac{\mu_M(d_c)}{(\mu_R(d_c)+\mu_M(d_c))}e^{-t/\tau_{MM}^d(d_c)}\right) \end{array} \tag{9}$$

in which the characteristic times of the spherical and deviatoric parts are respectively

$$\tau^s_{MM}(d_c) = \frac{\eta^s_M(d_c)(k_R(d_c)+k_M(d_c))}{3k_R(d_c)k_M(d_c)} \quad \tau^d_{MM} = \frac{\eta^d_M(d_c)(\mu_R(d_c)+\mu_M(d_c))}{2\mu_R(d_c)\mu_M(d_c)}. \tag{10}$$

2.2 Masonry's periodic cell global behaviour

For the viscoelastic periodic problem, the auxiliary problem reads

$$\begin{cases} \dot{E} + \varepsilon(\dot{u}^{per}) = \varepsilon(\dot{u}) \\ div(\sigma(t)) = 0 \\ \sigma(t).n \quad \text{anti-periodic on} \quad \partial Y \\ u^{per} \quad \text{periodic on} \quad \partial Y \\ \langle \sigma(t) \rangle = \Sigma(t) \end{cases} \tag{11}$$

where $\sigma(t)$ is the microscopic stress tensor state, $\varepsilon(\dot{u}(y))$ is the microscopic strain tensor state, u^{per} is the periodic displacement field and $\dot{E}$ is the macroscopic in-plane strain tensor rate. Following the procedure of Cecchi and Tralli [7] accounting for the joint thickness, the expression of the homogenized tensor $\tilde{A}^R$ of the periodic cell is provided in the same quoted reference. If only the mortar is assumed to be viscoelastic such that its stiffness tensor reads $A^m(t) = A^m(1+\phi_m(t))$, the main idea consists to substitute the Young's modulus E^m by $J^{-1}_m(t) = E^m(1+\phi_m(t))$.
Recently, based on the works of Cecchi & Sab, Cecchi & Barbieri [6] and Cecchi & Tralli, Cecchi & Taliercio [8] have proposed a homogenized compliance for viscoelastic undamaged masonry with mortar joints assimilated to interfaces. In this paper we propose to extend also this model to microcracked masonry with finite dimensions of microcracked mortar joints. The effective compliance of the damaged viscoelastic masonry reads then:

$$\begin{aligned} \tilde{S}^R_{1111}(t,dc) &= \frac{S^b_{1111}\frac{e_v}{a+e_h}J'v(t,dc)+4S^b_{1111}\frac{e_h}{b+e_v}J'_h(t,dc)+4\frac{e_v e_h}{(b+e_v)^2}J'_v(t,dc)J''_h(t,dc)}{4\frac{e_h}{b+e_v}J''_h(t,dc)+\frac{e_h}{a+e_h}J'_v(t,dc)} \\ \tilde{S}^R_{1212}(t,dc) &= S^b_{1212} + \frac{e_h}{a+e_h}J''_h(t,d_c) + \frac{e_v}{b+e_v}J''_v(t,dc) - \frac{e_v^2(b+e_v)(a+e_h)J''^2_v(t,dc)}{4\frac{e_h}{b+e_v}J'_h(t,dc)+\frac{e_v}{a+e_h}J''_v(t,dc)} \\ \tilde{S}^R_{2222}(t,dc) &= S^b_{2222} + e_h a + e_h J'_h(t,dc), \\ \tilde{S}^R_{1122}(t,dc) &= S^b_{1122} \end{aligned} \tag{12}$$

The damaged masonry's orthotropic effective properties are then the following: $\tilde{E}_{11}(t,d_c) = \frac{1}{\tilde{S}^R_{1111}(t,d_c)}$, $\tilde{E}_{22}(t,d_c) = \frac{1}{\tilde{S}^R_{2222}(t,d_c)}$, $\tilde{\mu}_{12}(t,d_c) = \frac{1}{\tilde{S}^R_{1212}(t,d_c)}$, $\tilde{\nu}_{ij}(t,d_c) = -\tilde{E}_{ii}(t,d_c)\tilde{S}^R_{iijj}(t,d_c)$ where $(i,j) \in \{1,2\}$. These moduli are explicit function of the crack density - damage parameter -, time and ratios $\frac{e_h}{a+e_h}$, $\frac{e_v}{b+e_v}$ instead of respectively $\frac{e_h}{a}$ and $\frac{e_v}{b}$. It is worth noting, that for the case of undamaged mortar interfaces, Cecchi & Taliercio have proven that this solution is more consistent with a numerically homogenized solution based on the finite elements method for ratios $E_b/E_m \geq 20$ when the mortar follows a Generalized Maxwell model. In the following, we assume that this condition ensuring the accuracy of the Cecchi & Taliercio's analytical model with the additional assumption of finite joints dimensions is also available for damaged mortars.

3 Results

The proposed model is applied to the case of a masonry with hybrid mortar's joints which material's properties are given in table 1. The properties of the safe bricks are the following:

E_M (MPa)	τ_M (s)	E_R (MPa)	τ_R (s)	ν_m
4038	46490	2112	90866	0.22

Table 1: Elastic and viscous moduli of hybrid mortar [11]

$\nu_b = \nu_m$ and $E_b = 100E_m$ with dimensions 55x250 mm^2. Mortar's joints thicknesses are $e_h = e_v = 10\ mm$. Note that the characteristic times for the spherical and deviatoric hybrid mortar's behaviour are assumed to be equal $\tau^s = \tau^d = \tau$ for the MM's and Burgers' models. The Young's modulus E_R and Poisson's ratio ν_R coincide respectively with E_K and ν_K, properties of the Kelvin-Voigt's spring.

3.1 Mortar's effective creep function

Figure (2) illustrates the evolutions of normalized inverse dilute creep functions of respectively the Burgers' and MM's models with respect to the damage parameter d_c and the time t.

For a given crack density parameter $d_c = 0,\ 0.1$ or 0.2, Figure (2)-b demonstrates that the MM's model yields to a constant function $J^{-1}(t, d_c)/J^{-1}(0, 0)$ with variation of the time beyond $t = 10^6$(s) (i.e. almost 11 days). As expected, the increase of d_c decreases the level of the asymptotic limits reached by this function. The difference between the MM's curves for different d_c is not negligible (around 15%) unlike that observed for the Burgers curves which are very close especially at short and long terms. Figure (2)-a shows that the Burgers model leads to vanishing inverse creep functions for $t \geq 310^8$ (s) i.e. 1157 days for every crack density value $d_c \geq 0$.

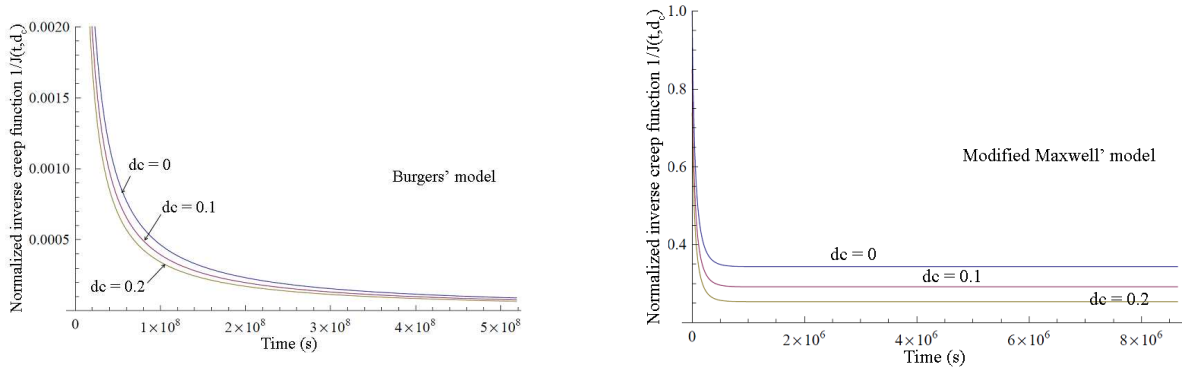


Figure 2: Variation of $J^{-1}_{MM}(t, d_c)/J^{-1}_{MM}(0, 0)$ versus time for safe mortar ($d_c = 0$) and microcracked mortars ($d_c = 0.1,\ d_c = 0.2$). Mortar following (a) the Burgers's and (b) MM's model.

3.2 Case of a compressed masonry panel

We study the case of a masonry panel of dimensions $L = 1560\ mm$ (length) and $H = 1040\ mm$ (height) studied in [7] subjected to boundary conditions BC-1 with three distributed loads at the top and two lateral edges (see Figure (3)). Bricks are assumed to be elastic or rigid. The mortar inside the joints is assumed to be microcracked with a matrix which obeys to linear viscoelastic behaviour following the Burgers or Modified Maxwell's model. As the arrangement of the bricks is regular, the effective behaviour of the panel is assumed to be well estimated by that of a periodic cell (see Figure 2-a). The panel can then be modeled as a homogeneous

material which properties coincide with those of the equivalent material MHE-2 (see Figure (1)-e). For the sake of simplicity, it is assumed that the crack density d_c is set equal to zero at the initial time $t_0 = 0$ and evolves linearly with time t as follows

$$d_c = \dot{d}_c \, t \tag{13}$$

where, here, the rate of the crack density $\dot{d}_c$ is assumed to be a positive constant lower than $0.002/\text{day}$ (i.e. if $t = 100$ days, then d_c reaches a maximum of 0.2). Indeed, beyond this limit, the dilute estimate will not be appropriate. Of course, as well known, the increase of the damage rate reduces the stiffness of the masonry cell. Under boundary conditions BC-1, the stress

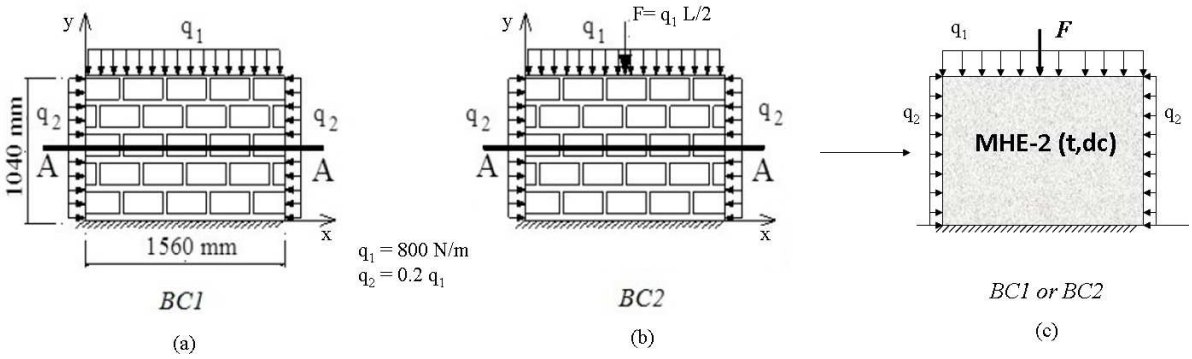


Figure 3: Boundary conditions BC-1 applied to the masonry panel [7]

(strain) fields σ_{yy} and σ_{xy} (ε_{yy} and ε_{xy}) are concentrated at both left and right superior corners of the compressed wall. The magnitude of the stress field σ_{yy} is similar for both the MM's and Burgers' models. However, the Burgers estimates of the stress fields σ_{xx} and σ_{xy} are superior to those predicted by the MM's model. These trends are confirmed by the evolutions of the stress

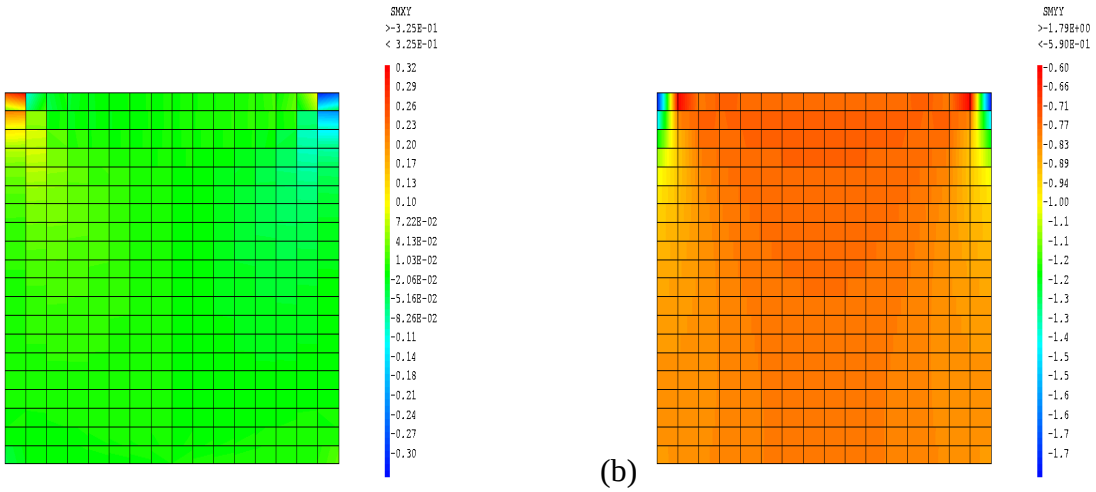


Figure 4: *BC-1 boundary conditions: MM's model predictions of σ_{xy} (a) and σ_{yy} (b) maps for $d_c = 0.15$ at $t = 1000$ days. Here crack density $d_c = \dot{d}_c t$ where $\dot{d}_c = 1.5\ 10^{-4}/\text{day}$.*

components with respect to the axis x as shown on figure (6). Note that unlike the stress (strain) fields σ_{yy} and σ_{xx} (ε_{yy} and ε_{xx}) which are symmetric by reference to the axis of symmetry of the panel ($x = L/2$), the shear stress σ_{xy} (shear strain ε_{xy}) is anti-symmetric. For the MM's model,

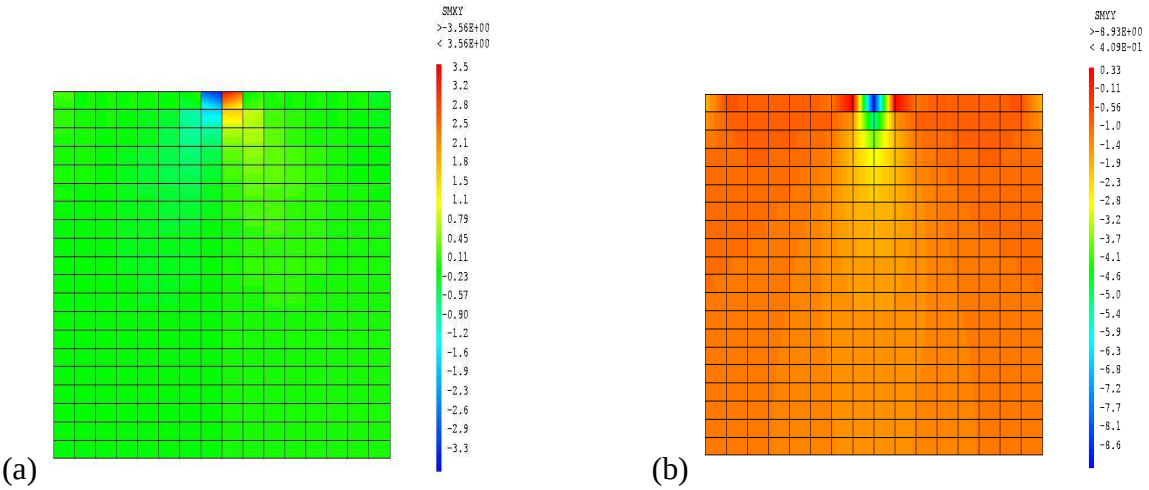


Figure 5: *BC−2 boundary conditions: MM's model predictions for the maps of stresses σ_{xy} (a) and σ_{yy} (b) for(a) $d_c = 0$ at $t = 0$ and (b) $d_c = 0.15$ at $t = 1000$ days. Here crack density $d_c = \dot{d}_c t$ where $\dot{d}_c = 1.5\ 10^{-4}$/day.*

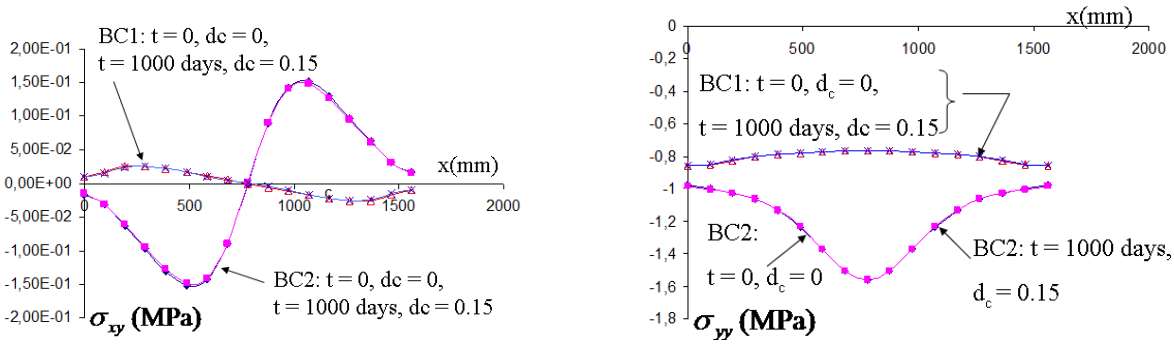


Figure 6: Boundary condition $BC - 1$ or BC−2: trends in σ_{xy} (a) and σ_{yy} (b) at the panel section A-A for the Burgers' (a) and Modified Maxwell's model at time $t = 0$ with $d_c = 0$ and $t = 1000$ days with $d_c = 0.15$. Here crack density $d_c = \dot{d}_c t$ where $\dot{d}_c = 1.5\ 10^{-4}$/day.

it is observed [18] that the strain fields increase slightly with time and damage. Moreover, this model predicts small strains unlike the Burgers model. This rheological model predicts for the masonry large strains similarly to the case of polymer materials. This quantitative aspect for the Burgers predictions is available at $t \geq 25$ days in a safe or damaged state. This result violates the hypothesis of small strains adopted in this study. These results motivate to avoid the use of a mortar following the Burgers model in masonries since such a model leads to vanishing masonry's stiffnesses and large strains increasing thus the risk of failure. Such results motivate to avoid modeling traditional mortars with short-term gathered properties using the Burgers model since it leads to vanishing masonry's stiffnesses and large strains increasing thus the risk of failure. These trends for the Burgers model are not valid if the Maxwell's relaxation time of the mortar is too high ($\tau_M \geq 10^7(s)$) as it is the case f hybrid mortar with long-term characteristic times (Table-12). In the later case, the Burgers model is expected to provide local fields predictions similar to the MM's estimates. The boundary conditions BC-1 are preferable to BC-2 since the later increases the stress and strain levels throughout the wall. Owing to Figures (4) and (5), at $t = 1000$ days, while failure occurs in the wall's area located around the application's point of the concentrated load F under BC-2, there is no failure in the wall submitted to conditions BC-1.

4 CONCLUSION

This paper proposes 'direct' (analytical and numerical) homogenization models for non-aging viscoelastic microcracked masonries based on the extensions of the Cecchi $\&$ Barbieri [6], Cecchi $\&$ Tralli [7] and Cecchi $\&$ Taliercio [8] models which are available for uncracked masonries. The herein proposed models rely on the coupling between the Griffith's theory and the periodic homogenization (analytical and numerical) schemes in order to derive easily and with low computational effort - without recourse to numerical inversion of the Laplace transform - the effective creep function of a microcracked non-aging viscoelastic mortar. Notice that this 'direct' homogenization model is preffered to the 'indirect' incremental homogenization which requires the additional computation of prestress in the masonry and is dependent of the time increment. The main result of this study dealing with microcracked mortars following the Modified Maxwell and Burgers models, is that unlike the MM's model which estimates well the creep of masonry at short to long terms, the relevance of this version of the Burgers model is limited to mortars with too high Maxwell's relaxation times ($\tau_M \geq 10^7(s)$) otherwise it predicts vanishing stiffness at short and medium terms and exaggerated local strains. In addition, the effects of the damage parameter - crack density d_c - and the ratio between the mortar's and bricks stiffness at short and long terms have been assessed [18].

REFERENCES

[1] H.D. Bui, *Mécanique de la rupture fragile*, Edition Masson, 1978.

[2] M, Bornert, T., Bretheau, P., Gilormini, *Homogénéisation en mécanique des matériaux, Tome-1 : Matériaux aléatoires élastiques et milieux périodiques*. Hermes Science, 2001.

[3] L. Binda, G. Gatti, G. Mangano, C. Poggi, G. Sacchi-Landriani, The collapse of civic tower of Pavia, *Masonry International*, 6(1), 11-20, 1992.

[4] J.J. Brooks, Composite modeling of masonry deformation, *Materials $\&$ Structures*, 23, 241-251, 1990.

[5] K. K. Choi, S.L. Lissel, M.M.R. Taha, Rheological modelling of masonry creep. *Canadian Journal of Civil Engineering*, 34, 1506-1517, 2007.

[6] A. Cecchi, A. Barbieri, Homogenisation procedure to evaluate the effectiveness of masonry strengthening by CFRP repointing technique, *Applied and Theoretical Mechanics*, 1 (3), (2008)

[7] A. Cecchi, A. Tralli, A homogenized viscoelastic model for masonry structures. *International Journal of Solids and Structures*, 49, 1485-1496, 2012.

[8] A. Cecchi, A. Taliercio, A comparison between numerical and analytical homogenized models for visco-elastic brickwork. *XXI Congresso AIMETA Associazione Italiana di Meccanica Teoricae Applicata*, Torino, Italy, 1-10, 2013.

[9] V. Deudé, L. Dormieux, D. Kondo, V. Pensée, Propriétés élastiques non linéaires d'un milieu mésofissuré. *Comptes Rendus Mécanique* (2002) 330 (8), 587–592.

[10] L. Dormieux, D. Kondo, Stress-based estimates and bounds of effective elastic properties: the case of cracked media with unilateral effects. *Computational Materials Science* (2009) 46 (1), 173–179.

[11] S. Ignoul, L. Schueremans, J. Tack, L. Swinnen, S. Feytons, L. Binda, D.V. Gemert, K.V. Balen, Creep behaviour of masonry structures - failure prediction based on a rheological model and laboratory tests. Structural Analysis of Historical Constructions. *Conference, New Delhi*, 2007.

[12] B. Nedjar, R. Le Roy, An approach to the modeling of viscoelastic damage: application to the long-term creep of gypsum rock materials. *International Journal for Numerical and Analytical Methods in Geomechanics* 37, 1066-1078, 213.

[13] T. T. N. Nguyen, A. Rekik, A. Gasser, A multi-level approach for micro-cracked masonry. *11th World Congress on Computational Mechanics (WCCM XI)*, Barcelona, July (2014).

[14] T. T. N. Nguyen, Approche multi-échelles pour des maçonneries viscoélastiques. *PhD University of Orléans*, France (2015).

[15] S. T. Nguyen, Propagation de fissures et endommagement par microfissures dans un milieu viscoélastique linéaire non veillissant. PhD, *University Paris-Est*, France (2010).

[16] E. Papa, A. Taliercio, A visco-damage model for brittle materials under monotonic and sustained stresses, *International journal for numerical and analytical methods in geomechanics*, 29 (3), 287-310, 2005.

[17] E. Papa, A. Taliercio, Modeling of the long-term behaviour of historical masonry towers, *in Book: Learning from Failure: Long-term Behaviour of Heavy Masonry Structures*, 153-173, 2008.

[18] A., Rekik, N.T.T., Nguyen, A. Gasser, Multi-level modeling of viscoelastic microcracked masonry *International Journal of Solids and Structures*, 2015, 63-83, http://dx.doi.org/10.1016/j.ijsolstr.2015.11.002.

[19] A., Rekik, A. Gasser, Numerical homogenization model for effective creep coefficients of microcracked masonry, submitted to *International Journal of Solids and Structures*, 2016.

[20] N.G. Shrive, A. Huizier, Letter to the Editor, *The masonry Society Journal*, 9(2),1991

[21] R. A. Schapery, Approximate methods of transform inversion for viscoelastic stress analysis. *Proc. U.S. Nat. Congr. Appl. Mech. ASME 4th*. Vol. 2. (1962) 1075–1085.

[22] A. Taliercio, Closed-form expressions for the macroscopic in-plane elastic and creep coefficients of brick masonry. *International Journal of Solids and Structures*, 51, 2949-2963, 2014.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

APPLICATION OF POLAR ELASTICITY TO THE PROBLEM OF PURE BENDING OF A THICK PLATE

Svitlana Fedorova[1] **and Jiří Burša** [2]

[1] Institute of Solid Mechanics, Mechatronics and Biomechanics
Faculty of Mechanical Engineering
Brno University of Technology
Technická 2896/2
616 69 Brno
Czech Republic

e-mail: svtlnfdrv2@gmail.com

[2] dtto

e-mail: bursa@fme.vutbr.cz

Keywords: Polar Theory, Fibres, Couple Stress, Anisotropic.

Abstract. *Polar elasticity theory is employed to derive exact solution for the pure bending problem of the transversely isotropic plate. It is then compared to the conventional elasticity solution and the corresponding composite plate solution. The material parameters of the polar model are chosen so that it represents the effective behaviour of the composite plate. The influence of the additional material constant related to the fibre size effect and affecting bending stiffness of the plate is illustrated.*

1 INTRODUCTION

The recently developed polar elasticity theory [1, 2] deals with fibre-reinforced materials containing fibres that possess bending stiffness. The contribution of the fibre bending stiffness is accounted for by employing couple stresses and introducing additional elastic parameters to the constitutive equations. To gain a better understanding of the additional constant and the role of couple stress in polar theory [2] we focus on exact analytical (polar elasticity) solutions for the pure bending of thick infinite plates.

The paper is organized in the following manner: in Section 2 the displacement solution is derived for the transversely isotropic thick plate under pure bending (and the conventional – Cauchy theory of elasticity); in Section 3 the solution is derived for the case of longitudinal modulus being a periodic function of position; in Section 4 the first problem is analysed from the standpoint of couple stress theory. In Section 5 solutions of all the three models, taken with the matching effective properties and the same effective loading, are compared.

2 TRANSVERSELY ISOTROPIC MATERIAL (CONVENTIONAL THEORY OF ELASTICITY)

Let us consider a rectangular plate subjected to a known bending moment M_3 [Fig. 1]. The material is transversely isotropic and characterized by the following constants: E_1, E_2, ν_{12}, $\nu_{23}= \nu_{32}$, G_{21}.

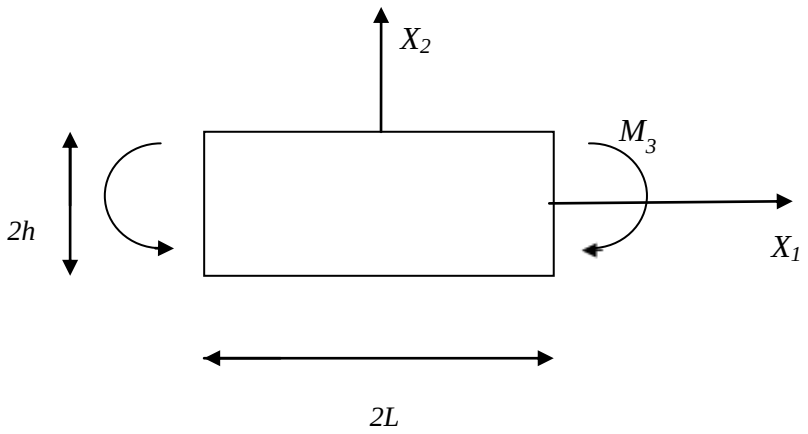


Fig.1: A thick plate with infinite length in X_3 direction.

The following boundary conditions have to hold for the tractions t_1 and t_2 on the specified faces:

$$X_2 = \pm h \Rightarrow t_2 = 0 \tag{1a}$$

$$X_1 = \pm L \Rightarrow \begin{array}{l} \int_{-h}^{h} t_1 dx_2 = 0, \\ \int_{-h}^{h} t_1 x_2 dx_2 = M_3. \end{array} \tag{1b}$$

Equilibrium equations (with negligible volume loads) must be satisfied as well:

$$\frac{\partial \sigma_{ji}}{\partial x_j} = 0 \quad , \tag{2}$$

The following stress distribution satisfies both (1) and (2):

$$\sigma_{22} = 0, \quad \sigma_{11} = \alpha x_2, \quad \sigma_{12} = 0. \tag{3}$$

The stress distribution throughout the plate is the same at any cross-section (Fig.2).

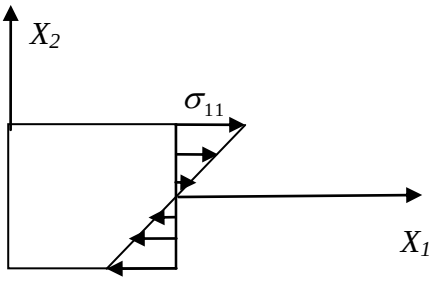


Fig.2: The stress distribution at any cross-section.

Under plane strain conditions (ε_3=0) and for stresses in the form (3) the constitutive equations yield:

$$\begin{aligned} \varepsilon_1 &= \frac{1-\nu_{21}\nu_{12}}{E_1}\sigma_1 \\ \varepsilon_2 &= \frac{-\nu_{12}(1+\nu_{23})}{E_1}\sigma_1 \end{aligned} \tag{4}$$

By integrating (4) we obtain

$$\begin{aligned} u_1 &= \int \varepsilon_1 dx_1 + g_1(x_2) \\ u_2 &= \int \varepsilon_2 dx_2 + g_2(x_1) \end{aligned} \tag{5}$$

$$\begin{cases} u_1 = \dfrac{1-\nu_{21}\nu_{12}}{E_1}\alpha x_1 x_2 + g_1(x_2) \\ u_2 = \dfrac{-\nu_{12}(1+\nu_{23})}{2E_1}\alpha {x_2}^2 + g_2(x_1) \end{cases} \tag{6}$$

With the additional condition of zero shear strain

$$\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} = 0 \tag{7}$$

we find that

$$\frac{1-\nu_{21}\nu_{12}}{E_1}\alpha x_1 + g_1'(x_2) + g_2'(x_1) = 0 \tag{8}$$

and

$$\begin{cases} g_2'(x_1) = -\dfrac{1-\nu_{21}\nu_{12}}{E_1}\alpha x_1 + c, \\ g_1'(x_2) = -c, \end{cases} \tag{9}$$

$$\begin{cases} g_2(x_1) = -\dfrac{1-\nu_{21}\nu_{12}}{2E_1}\alpha x_1^{\,2} + cx_1 + c_1, \\ g_1(x_2) = -cx_2 + c_2, \end{cases} \tag{10}$$

where the constants c_1, c_2, c are set to zero (c_1, c_2 refers to rigid body translation, c to the rigid body rotation). The final solution in displacements then acquires the form

$$\begin{cases} u_1 = \dfrac{1-\nu_{21}\nu_{12}}{E_1}\alpha x_1 x_2, \\ u_2 = \dfrac{-\nu_{12}(1+\nu_{23})}{2E_1}\alpha x_2^{\,2} - \dfrac{1-\nu_{21}\nu_{12}}{2E_1}\alpha x_1^{\,2}. \end{cases} \tag{11}$$

3 HETEROGENEOUS MATERIAL WITH PERIODIC PROPERTIES

We still consider a plate subjected to a known bending moment M_3 [Fig. 1]. The material is now characterized by the following parameters: $E_1 = E_0 + \tilde{E}\cos(\frac{\pi N}{h}x_2)$, $E_2 = \left(\frac{N}{2h}\int_0^{2N/h}\frac{dx_2}{E_1(x_2)}\right)^{-1}$, ν_{12}, ν_{23}, G_{21}.

The same boundary conditions (1) and equilibrium equations (2) must be satisfied. In order to ensure pure bending, the applied normal stress at the ends of the plate must accommodate to the strain distribution $\varepsilon_1 = kx_2$, where the curvature is introduced as $k = -u_{2,11}$. The curvature can be defined in relation to the load as follows:

$$M_3 = \frac{k}{1-\nu_{1T}\nu_{T1}}\int_{-h}^{h} E_1 x_2^2 dx_2 \tag{12}$$

Thus the stresses are

$$\sigma_{22} = 0, \quad \sigma_{11} = \frac{E_0 + \tilde{E}\cos(\frac{\pi N}{h}x_2)}{1-\nu_{1T}\nu_{T1}}kx_2, \quad \sigma_{12} = 0. \tag{13}$$

Following the logics of the previous section and the Hooke's law relations for plane strain (4) we arrive to the displacement solution

$$\begin{cases} u_1 = kx_1x_2, \\ u_2 = \int_0^x \varepsilon_2 dx_2 - \frac{1}{2}kx_1^2. \end{cases} \tag{14}$$

4 TRANSVERSELY ISOTROPIC MATERIAL (POLAR THEORY OF ELASTICITY)

Now let us consider the pure bending of the plate with the additional elastic constant $\boldsymbol{d_{31}}$ indicative of the bending stiffness of the substructure. The loading conditions stay the same as on the Fig. (1).
The boundary conditions on the faces in this case are as follows:

$$x_2 = \pm h: \quad t_2 = 0 \tag{15a}$$

$$x_1 = \pm L: \begin{array}{l} \int_{-h}^{h} t_1 dx_2 = 0, \\ \int_{-h}^{h} t_1 x_2 dx_2 + \int_{-h}^{h} m_{13} dx_2 = M_3 \end{array} \tag{15b}$$

where the couple stress [1] $\boldsymbol{m_{13}}$ is introduced.
The stress components are divided into the symmetric and antisymmetric parts

$$\sigma_{ij} = \sigma_{(ij)} + \sigma_{[ij]}. \tag{16}$$

Equilibrium equations in the polar theory for the present case are augmented by the equation connecting the antisymmetric shear stress and the couple stress:

$$\begin{array}{l} \frac{\partial \sigma_{ji}}{\partial x_j} = 0 \ , \\ 2\sigma_{[21]} = \frac{\partial m_{13}}{\partial x_1}. \end{array} \tag{17}$$

The following stress and couple stress distribution satisfies both (15) and (17) (also see Fig.3):

$$\begin{array}{l} \sigma_{22} = 0\,, \ \ \sigma_{21} = 0\,, \\ \sigma_1 = \alpha x_2\,, \ \ \sigma_{12} = 0\,, \\ m_{13} = const = \beta\,. \end{array} \tag{18}$$

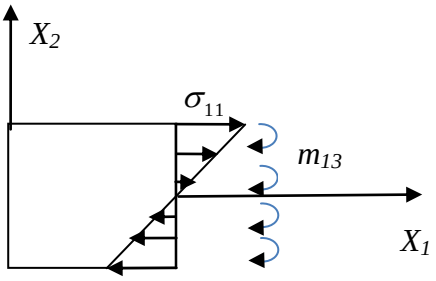


Fig.3: Distribution of the non-zero stress and couple stress components in a section of the plate.

The strain-stress relations hold in the form of (4) with the additional equation [1]:

$$m_{13} = d_{31} u_{2,11} \tag{19}$$

which can be rewritten here as

$$\frac{\partial^2 u_2}{\partial x_1^{\,2}} = \frac{\beta}{d_{31}}. \tag{20}$$

From the integration of (4) and zero shear condition (7) we obtain the displacement solution identical to (11):

$$\begin{cases} u_1 = \dfrac{1-\nu_{21}\nu_{12}}{E_1}\,\alpha x_1 x_2, \\ u_2 = \dfrac{-\nu_{12}\left(1+\nu_{23}\right)}{2E_1}\,\alpha x_2^{\,2} - \dfrac{1-\nu_{21}\nu_{12}}{2E_1}\,\alpha x_1^{\,2}. \end{cases} \tag{21}$$

If we now insert the second equation of (21) into (20) we see that this solution is valid only if

$$-\frac{1-\nu_{21}\nu_{12}}{E_1}\,\alpha = \frac{\beta}{d_{31}}\ . \tag{22}$$

So (22) is the result of requirement of zero shear strain (20). It means that for the given material constant d_{31} the applied stress parameter α and the applied couple stress value $\boldsymbol{\beta}$ must relate via (22) in order to ensure pure bending.

As this section shows, it can be concluded that the displacement solution in the form of (21) for the pure bending problem is valid for

a) transversely isotropic plate (d_{31}=0) under end loading $\boldsymbol{\sigma_{11}=\alpha x_2}$ or
b) transversely isotropic polar material plate ($d_{31}\neq 0$) under the end loading $\boldsymbol{\sigma_{11}=\alpha x_2}$ and $\boldsymbol{m_{13}=const=\beta}$ under the condition that (22) holds.

It can be verified that the units of the elastic constant d_{31} correspond to Newtons [N].

5 COMPARISON AND DISCUSSION

On the basis of the previous developments we compare the solutions of the three corresponding models under the given value of the applied bending moment. The model from the Section 2 is taken as a reference model. Its periodically changing stiffness can be regarded as an approximation of the fibre reinforced composite. This model we refer to as the PS (periodical stiffness) model. If we aim to replace the PS model with a homogeneous one, we can do it in two ways. Firstly, the model from the Section 1, taken with the corresponding properties, can serve this goal. This model we refer to as the EC (effective classical) model. Secondly, the model from the Section 3, further called the EP (effective polar) model, can simulate the given problem. The goal is to compare results from two of the alternative homogeneous models (ES and EP) with the PS model and to evaluate their accuracy under varying N which can be regarded as analogous to the nominal number of fibres per height in a unidirectional fibre composite. The properties of both homogeneous models are set as effective (averaged) properties of the PS model. In addition, the bending stiffness parameter in the EP model is set as $d_{31} = \dfrac{2h^2 \widetilde{E}}{\pi^2 N^2 \left(1 - \nu_{21}\nu_{12}\right)}$ and serves as a correction parameter which ensures that the overall bending stiffness of the plate in the EP model equals that one of the PS model. All models are loaded at both ends of the plate by two equilibrated couples with magnitude (per unit width) of $M=-500\ MPa{\cdot}mm^2$ inducing the same bending moment in all sections along the plate length. The dimensions are: $2L=200\ mm$ (length of the plate), $2h=50\ mm$ (thickness).

The following properties are set in the computations:

	PS model	**EC model**	**EP model**
E_1, *MPa*	$1000 + 999\cos(\frac{\pi N}{h} x_2)$	1000	1000
E_2, *MPa*	44,71	44,71	44,71
ν_{21}	0,3	0,3	0,3
ν_{23}	0,3	0,3	0,3
d_{31}, $MPa{\cdot}mm^2$			8690 for N=4 1390 for N=10

Table 1: Material constants set for different models.

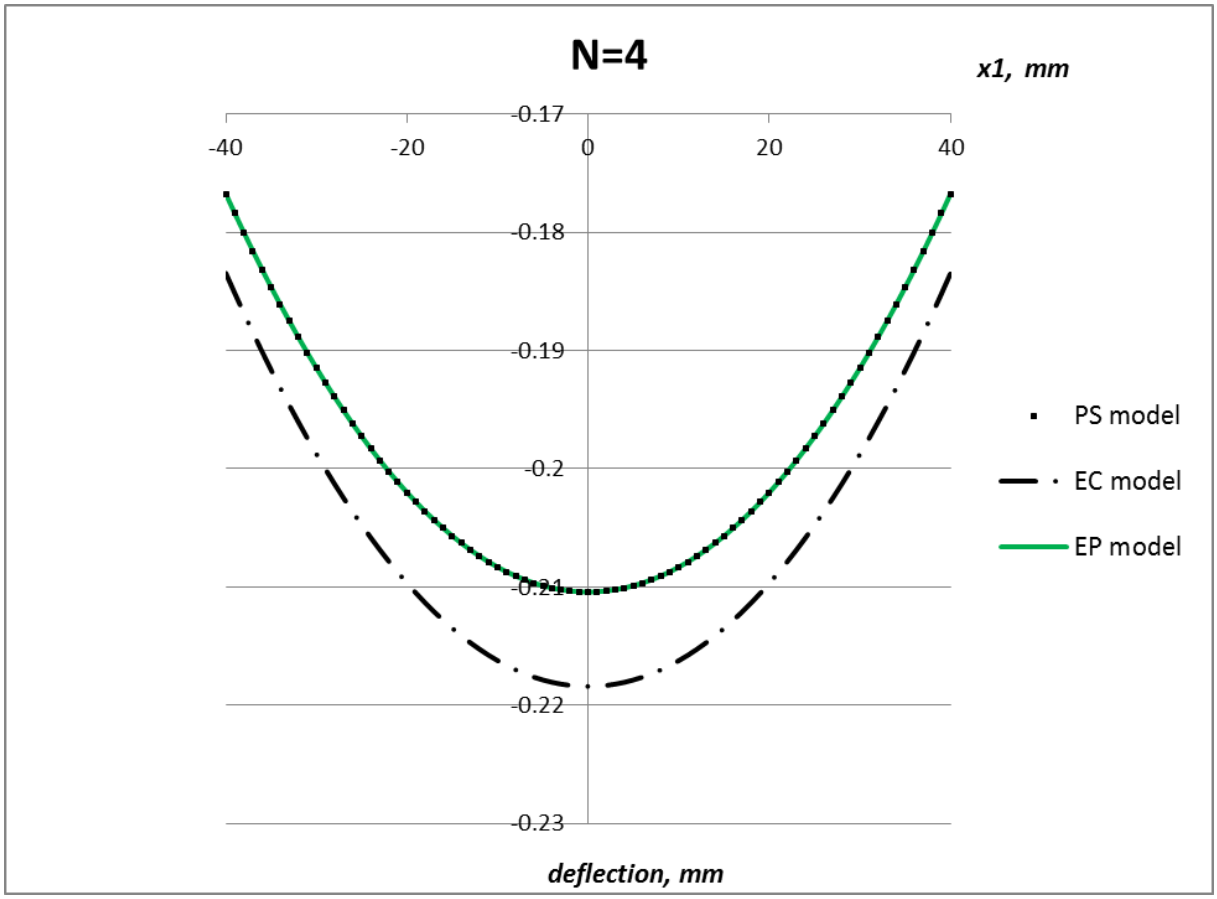


Fig.4: Plate deflection for different models, N=4.

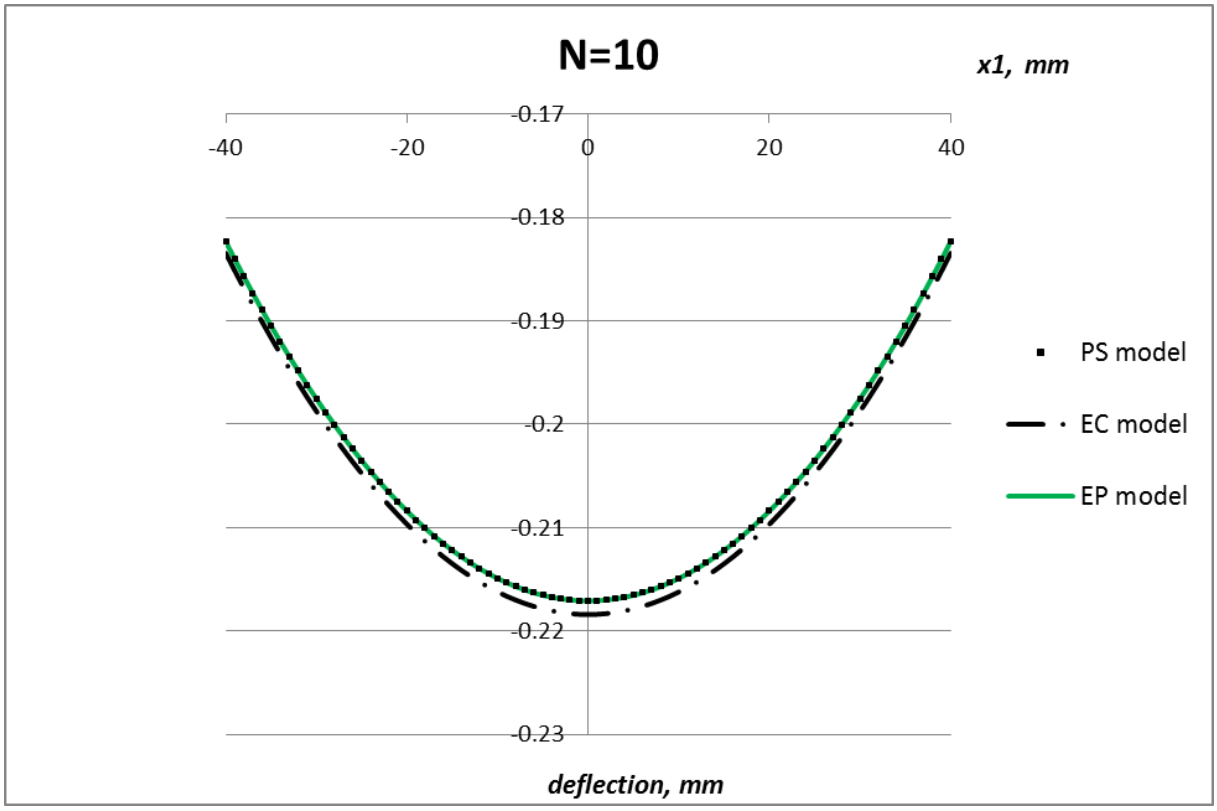


Fig.5: Plate deflection for different models, N=10.

The motivation for such comparison sprung from chapter 4 in [1] where the authors consider Euler-Bernoulli beam with the cosinusoidal Young's modulus distribution to point out the inaccuracy of the conventional theory of fibre-reinforced materials [3] in cases where the fibres are not infinitesimally thin.

In polar theory, the length scale can be introduced via the additional elastic constant d_{31}. In the present computations, the formula for d_{31} contains $(h/N)^2$ which is the (squared) dimension of a representative volume unit related to the nominal fibre thickness. (In a similar vein this constant is chosen in [4, 5].) Therefore the EP model gives correct results (equal to the PS model) for N=4 while the conventional theory of elasticity (EC model) neglects the bending stiffness of fibres and consequently underestimates the stiffness because the fibre thickness is not negligible. With increasing N the bending stiffness of fibres decreases and the correct solution (obtained by both PS and EP models) becomes closer to the EC model.

6 CONCLUSIONS

The new pure bending elasticity solution is derived for the transversely isotropic polar material. It is compared with the conventional theory solutions in order to demonstrate how the size effect can be taken into account in the homogeneous polar model. With the nominal number of fibres increasing their bending stiffness decreases and the polar elasticity model converges to the conventional elasticity model.

In the given example, the additional elastic constant has a role of a correcting parameter. It corrects the error in bending stiffness resulting from the averaging of the Young's modulus.

Acknowledgments. This work was supported by Czech Science Foundation project number 13-16304S.

REFERENCES

[1] A.J.M. Spencer, K.P. Soldatos, Finite deformations of fibre-reinforced elastic solid with fibre bending stiffness. *Int J Nonlinear Mech*, **42**, 355-368, 2007.

[2] K.P. Soldatos, Foundation of polar linear elasticity for fibre-reinforced materials. *J Elasticity*, **114**(2), 155-178, 2013.

[3] A.J.M. Spencer, *Deformations of Fibre-reinforced Materials*. Oxford University Press, London, 1972.

[4] A.F. Farhat, K.P. Soldatos, Cylindrical bending and vibration of polar material laminates. *Mech Adv Mater Struct*, **22**, 885-896, 2015.

[5] M.A. Dagher, K.P. Soldatos, On small azimuthal shear deformation of fibre-reinforced cylindrical tubes. *Journal of Mechanics of Materials and Structures*, **6**(1-4), 141-168, 2011.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ARTIFICIAL MICROSTRUCTURE GENERATOR FOR DUAL-PHASE STEELS

Y. Hou[1], A. Dumon[2], P. Culière[2] and M. Rachik[1]

[1]Laboratoire Roberval, CNRS UMR7337, Sorbonne universités, Université de technologie de Compiègne
Centre de recherche Royallieu, CS60319, 60203, Compiègne Cedex, France
e-mail: yuliang.hou@utc.fr, mohamed.rachik@utc.fr

[2] ESI Group
99 Rue des Solets, BP 80112, Rungis Cedex, France
e-mail: alexandre.dumon@esi-group.com, pierre.culiere@esi-group.com

Keywords: Dual-phase Steels, Microstructure Generator, Representative Volume Element, Homogenization, Proper Orthogonal Decomposition.

Abstract. *The microstructure of low alloyed ferritic-martensitic Dual-Phase (DP) steels consists of hard coarse grained martensite islands embedded in a soft ferrite matrix. Therefore, the macroscopic mechanical properties of DP steels mostly derive from their microstructures, such as volume fractions, morphology of martensite, phase distributions and ferrite grain size. Recently, micromechanical approaches are used to predict ductility and failure mode of DP steels under varying mechanical loading scenarios. In this work, an artificial microstructure generator inspired by topology optimization was developed to construct representative volume element (RVE) with predefined design parameters within a mofidied Voronoï tessellation. Micromechanical modeling of DP steel was performed on the generated artificial RVE. The plastic flow behavior of each single phase in DP steel were calculated by using a dislocation based theory. After numerical simulation, the flow curve on macro scale can be obtained from an asymptotic expansion homogenization (AEH) scheme. This approach allows studying the influence of individual microstructure features on local and global stress-strain response. To improve the robustness of this artificial microstructure generator, a proper orthogonal decomposition (POD) reduction was introduced to identify the optimal design parameters. This approach used a collection of artificial microstructures (snapshots) to find the most representative one of the real microstructure.*

1 INTRODUCTION

Advanced high strength steels (AHSSs), such as Dual-Phase (DP), Transformation Induced Plasticity (TRIP), Complex Phase (CP) and martensitic steels, have been developed and used in the automotive industry for the sake of reducing the weight of car body, improving passive safety performance, and fuel efficiency. DP steel usually consists of hard martensite islands embedded in a soft ferrite matrix, which makes it exhibit several mechanical characteristics, such as relatively high ultimate tensile strength (UTS), low yield to tensile strength ratio, absence of yield point elongation and a good balance of strength and formability [1, 2].

Recently, real microstructure based models of DP steel are created using experimental techniques, e.g., scanning electron microscopy (SEM), X-ray and neutron diffraction, and electron backscatter diffraction (EBSD), or statistical descriptions to study the influence of microstrutural featuers on the effective mechanical properties [3, 4, 5, 6]. However, highly inhomogeneous materials are formed during industrial welding, forging and heat treatment processes. Local microstructure varies significantly within the material, and is difficult to capture experimentally. It is comparatively easier to use a computed artificial microstructure based on local phase proportions and chemical compositions, which can be obtained from metal forming and heat-treatment simulations, e.g., SysWeld. This method particularly requires the development of an artificial microstructure with similar statistical properties to replace the real one.

Based on statistical descriptions of DP steel, artificial microstructures are commonly generated using geometry primitives (e.g., spheres, polygons or polyhedra). Al-Abbasi and Nemes [7] developed a micromechanical model for DP steel, which is dispersed of spherical martensitic particles with two different sizes in a ferritic matrix. However, this model has other disadvantages such as, inexact geometric representation and no reliable data close to the interface between different phases. Alternatively, Voronoï tessellation [8] is considered as an efficient tool for approximating the microstructure in DP steel.

In this study: firstly, a mofidied Voronoï tessellaion is periodically generated from Halton (quasi-random) sequence [9], which statistically exhibits low discrepancy, to provide adequate grain morphology. Secondly, two design parameters are proposed to control the phase assignment process in the DP steel. An algorithm related to density-based topology optimization method is introduced to solve the phase distribution problem. Thirdly, a proper orthogonal decomposition (POD) approach [10, 12] is applied to identify the optimal design parameters for a DP590 steel. Finally, micromechanial modeling are performed on the generated RVE using these optimal design parameters and the results corroborate experimental materials behavior.

2 MODIFIED VORONOI TESSELLATION

The concept of Voronoï tessellation allows to generate artificial polycrystalline aggregates with randomly distributed and oriented grains or cells for metallic or ceramic materials. This kind of aggregate is a nearest neighbor diagram determined from a set of generating points. Since the resulted diagram is mainly affected by the distribution of the Voronoï generating points, a modified point set has been applied to overcome the drawbacks (e.g., the inexact estimation of grain size and nearest neighboring grain number) found in the standard tessellation generated from a pseudo-random sequence [2]. Alternatively, the modified point set is generated using Halton (quasi-random) sequence, which statistically exhibits low-discrepancy [9].

Since periodic microstructures have favorable numerical properties in the context of computational homogenization, these seeds are repeated three times in each direction to ensure the periodicity of the modified Voronoï tessellation. In order to present the advantages of the mod-

ified Voronoï tessellation, an example has been compared with the standard one, as shown in Fig. 1 and 2. Two sets of seeds are generated using Halton (quasi-random) and pseudo-random sequences, as shown in Fig. 1a and b, respectively. The pairwise distance of each seed follows the same Gaussian distribution, as shown in Fig. 1c and d, in which the average value is 0.52 µm and the standard deviation is 0.25. These two parameters can be applied to control the average size of the generated Voronoï cells.

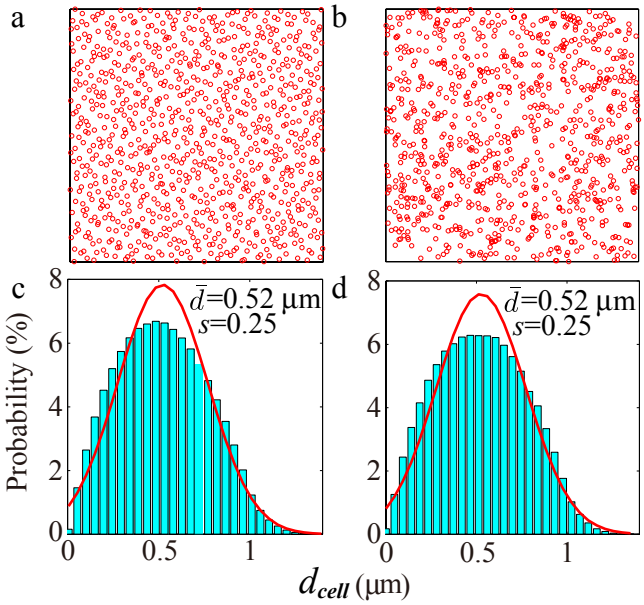


Figure 1: Scatter of periodic (a) modified and (b) standard Voronoï generating seeds including 900 points (30 points in each direction). Distribution of pairwise distance in each case: (c) modified and (d) standard generating seeds.

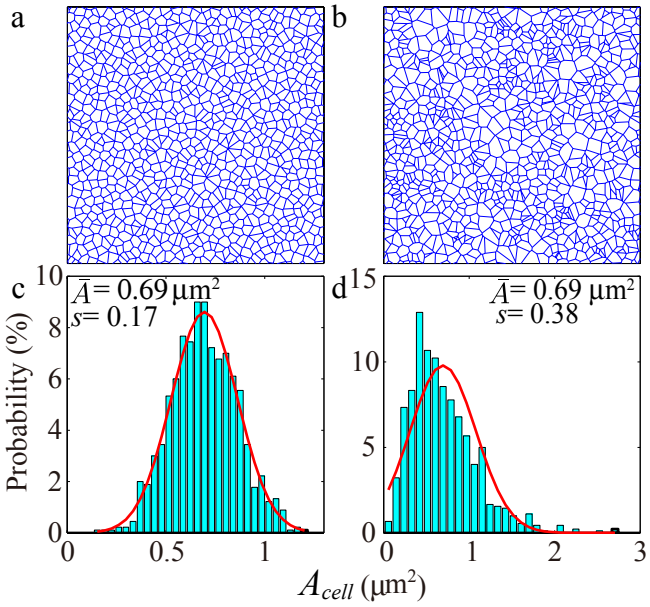


Figure 2: Square periodic Voronoï tessellations including 900 cells generated for: (a) modified one using Halton (quasi-random) sequence and (b) standard one using pseudo-random sequence. Distribution of cell area in each case: (c) modified and (d) standard tessellations.

Fig. 2a and b show the modified and standard Voronoï tessellations including 900 cells generated from the two seed sets. Fig. 2c and d also illustrate the area distribution of Voronoï cells for the modified and standard tessellations, respectively. From the comparison, both of the two generated Voronoï tessellations have the average cell area of $\bar{A}_{cell} = 0.69$ µm^2. However, the modified tessellation from Halton sequence shows a standard deviation of 0.17 which is less than half of the standard one (0.38). This clearly indicates that the modified Voronoï cells are more regular than the standard one. It can effectively avoid the appearance of bad aspect ratio and extremely small or large cells in the artificial polycrystalline aggregate, as shown in Fig. 3. This might cause that the variability of cell size is underestimated, while the number of nearest neighboring cells is overestimated.

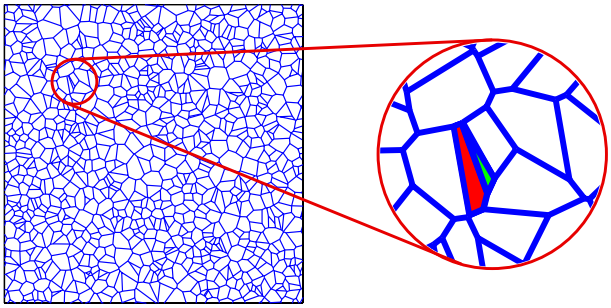


Figure 3: Voronoï cells with bad aspect ratio (red) and extremely small cell (green) in the standard Voronoï tessellation.

Usually, big martensite clusters are formed in the standard tessellation using the proposed phase assignment algorithm, which affect the formation of shear bands along the phase interface. It definitely leads to an inexact flow behavior prediction of DP steel. Moreover, the modified Voronoï tessellation is generated directly from Halton sequence, no additional computational cost on iteration to control the grain size distribution and grain morphology.

3 PHASE ASSIGNMENT ALGORITHM

With the modified Voronoï tessellation, an automate process need to be developed to assign the selected cells to represent different phases. Within this automate process, the modified Voronoï tessellation can be considered as a fixed grid, which is similar in material topology optimization design. In each modified Voronoï cell, a material density function is proposed to determine its phase property: martensite cells with density $\rho = 1$, the red layer in Fig. 4, while ferrite ones with density $\rho = 0$, the blue layer in Fig. 4. By defining constraints and an objective function, this process is deduced to a 0-1 decrete value optimization problem, also known as "black-and-white" design. Therefore, the material interpolation algorithms in topology optimization can be referred ti achieve artificial microstructures with proper phase distribution, as shown in Fig. 4.

In order to control the phase assignment, two design parameters proposed by Fillafer et al. [2], are utilized in this study to confine the solution space. These two parameters, the martensite phase fraction P_M and the neighboring coefficient of martensite grains C_M, which consider not only the martensite phase fraction but also the correlation of different martensite islands. The expressions of these two parameters are given as:

$$P_M = \frac{A_M}{A_T}, \quad C_M = \frac{2L_{MM}}{2L_{MM} + L_{FM}} \tag{1}$$

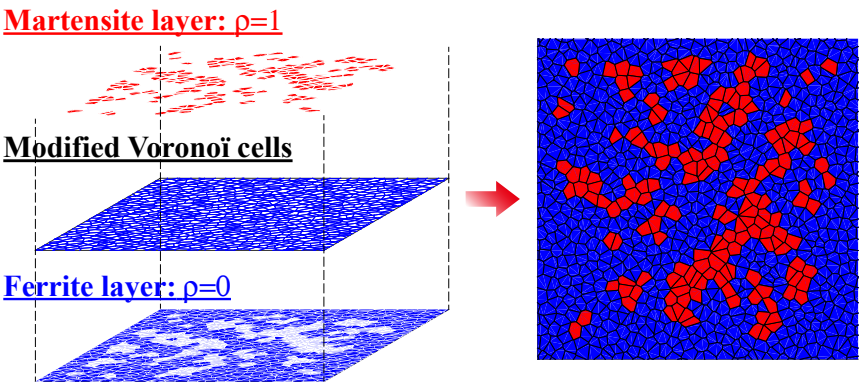


Figure 4: Automate phase assignment process and final DP microstructure.

where A_M and A_T denote martensite phase and total area, L_{MM} and L_{FM} are length of specific martensite-martensite and ferrite-martensite grain boundaries, respectively. Albeit it is identified that the value of P_M can vary between 0 and 1, for a given martensite phase fraction, C_M cannot satisfy the same range [2]. In other words, these two design parameters are not mutually independent.

According to the algorithm related to density-based topology optimization method, the design parameters can be rewritten in matrix form:

$$P_M = \frac{\mathbf{a}^T \boldsymbol{\rho}}{\mathbf{a}^T \mathbf{I}}, \quad C_M = \frac{\boldsymbol{\rho}^T \mathbf{E} \boldsymbol{\rho}}{\mathbf{I}^T \mathbf{E} \boldsymbol{\rho}} \tag{2}$$

where $\mathbf{a}$ is the grain area vocter, $\mathbf{E}$ is a correlation matrix of martensite islands' neighboring coefficient, $\mathbf{I}$ is uniform vector, and $\boldsymbol{\rho}$ is material density vector which only consists of 0 or 1:

$$\rho_e(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad \mathbf{x} \in \Omega \backslash \Omega_M \\ 1 & \text{if} \quad \mathbf{x} \in \Omega_M \end{cases} \tag{3}$$

Noting from the definition of $\boldsymbol{\rho}$, a distributed and discrete value problem is formulated in the matrix form. To solve this problem, the most commonly used approach in topology optimization is to replace the integer by continuous variables. And then, a penality factor p is introduced to derive the martensite density distribution same as the so-called "black-and-white" solution [11, 13]. Therefore, a power form of the material density in each cell, also named "pseudo-density", can be rewritten as:

$$Q_e = Q_F + \rho_e^p (Q_M - Q_F) \tag{4}$$

where Q_F and Q_M are the material density of ferrite and martensite, respectively. Therefore, a structural optimization problem can be given as:

$$\begin{aligned} \textbf{Argmin} \quad & J = \| C_M - C_M^{target} \| \\ s.t.: \quad & \textstyle\int_\Omega Q_e(\mathbf{x}) \cdot \mathrm{a} \, d\Omega \leq P_M^{target} \cdot A_T \\ & Q_e \in [Q_F, Q_M] \end{aligned} \tag{5}$$

where the target design parameters P_M^{target} and C_M^{target} are predefined based on statistical descriptions of DP steel. Here, a heuristic updating scheme [11, 13] is introduced to solve this

optimization problem:

$$Q_e^{new}(\mathbf{x}) = \begin{cases} \max(0, Q_e - \delta) & \text{if} \quad Q_e B^\eta \leq \max(0, Q_e - \delta) \\ \min(0, Q_e + \delta) & \text{if} \quad Q_e B^\eta \geq \min(0, Q_e + \delta) \\ Q_e B & \text{otherwise} \end{cases} \tag{6}$$

where δ and η donate a positive move limit and a numerical damping coefficient, respectively. And, the term B is calculated using the optimality condition:

$$B = \frac{\dfrac{\partial J}{\partial Q_e}}{\lambda \dfrac{\partial P_M}{\partial Q_e}} \tag{7}$$

In Eq. (7), the Lagrangian multiplier λ is calculated using bisection algorithm to ensure the satisfaction of phase fraction constraint. Other two terms, $\partial J/\partial Q_e$ and $\partial P_M/\partial Q_e$ can be obtained in matrix form.

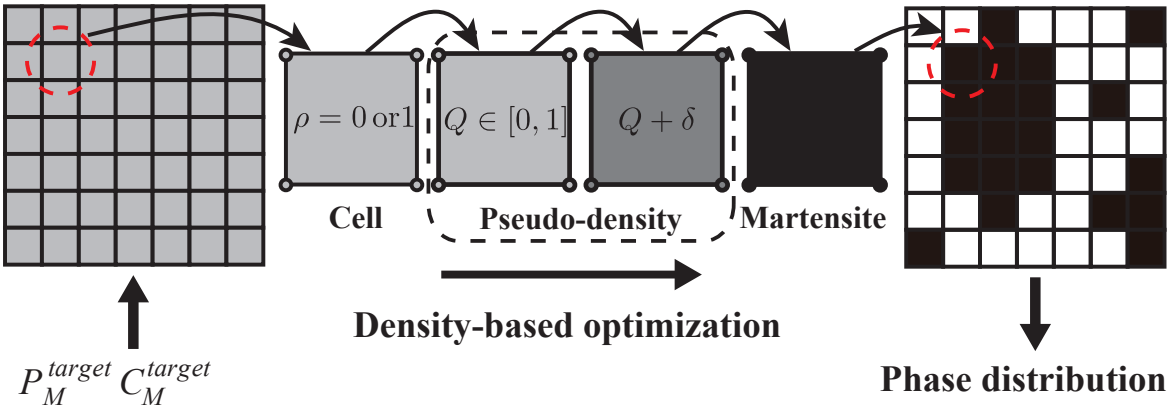


Figure 5: Illustration of the phase assignment algorithm inspired by topology optimization.

A schematic illustration of this phase assignment algorithm is shown in Fig. 5. In summary, the algorithm consists of the following step:

1. the initial material density of each Voronoï cell is given with the value of the target martensite phase fraction P_m^{target}, the initial grayscale grids in Fig. 5 (left);

2. the discrete material density is deduced to the continuous "pseudo-density", after the intergration of the penalization factor in Eq. (4). Meanwhile, this discrete value problem also becomes a continuous one;

3. iterations of the updating scheme in Eq. (6) is performed to achieve a convergent material density in each cell. The intermediate grayscale color in Fig. 5 (middle) shows the simultaneous material density after the first iteration;

4. a convergent solution is reached to present the optimal phase distribution. It is shown as the final grids in Fig. 5 (right), white and black fill colors indicate ferrite and martensite phases, respectively.

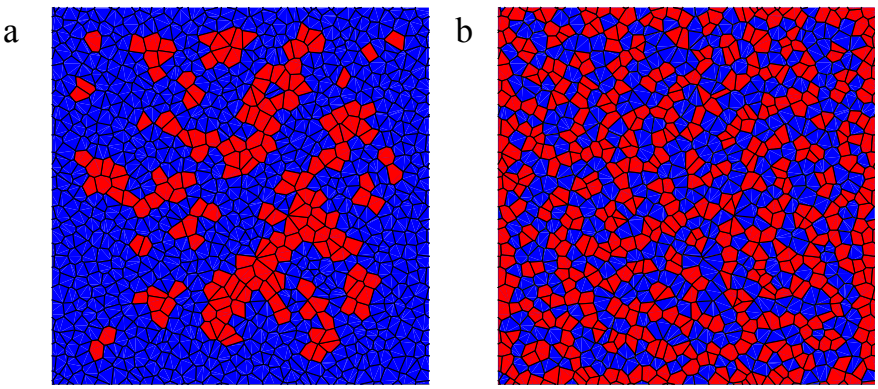


Figure 6: Example artificial microstructures with 900 grains: (a) $P_M = 0.22$, $C_M = 0.29$ ($P_M^{target} = 0.22$, $C_M^{target} = 0.3$) and (b) $P_M = 0.5$, $C_M = 0.5$ ($P_M^{target} = 0.5$, $C_M^{target} = 0.5$).

Following the aforementioned steps, artificial DP microstructures are constructed with a good convergence. That is, if the design parameters and the Voronoï tessellation are fixed, our generator will find only one unique optimal microstructure. Using the modified Voronoï tessellation described in Section 2 and predefined sets of design parameters, two example artificial microstructures are constructed as shown in Fig. 6.

In Fig. 6a, the target design paratmeters are set as $P_M^{target} = 0.22$ and $C_M^{target} = 0.3$. A fast and stable convergence is achieved in 9 iterations, the resulted microstructure is with $P_M = 0.22$ and $C_M = 0.29$. As discussed previously, due to the mutual independence of the two parameters, there exists a slight dissimilarity between the target and resulted parameters. While, an exact microstructure is constructed with target design parameters $P_M^{target} = 0.5$ and $C_M^{target} = 0.5$, as shown in Fig. 6b. It only takes 158 iterations to reach the convergence. This illustrates the flexibility of the proposed phase assignment algorithm in handling additional phase distribution parametric variables.

4 RESULTS AND DISCUSSION

In this study, the flow behavior of ferrite and martensite is obtained from a dislocation based theory [5], in which material parameters are calculated based on the local chemical composition. Numerical tensile test is performed on artificial RVEs generated with various design parameters. With the asymptotic expansion homogenization (AEH) method [14], each one can provide a prediction of flow behavior. In order to validate the proposed artificial microstructure generator, the generated RVE model is verified with a DP590 steel. The optimal design parameters $P_M = 0.22$ and $C_M = 0.3$ for the investigated DP590 steel are identified using a POD reduction approach, which considers the experimental measurement, as *a priori*. Therefore, the RVE generated by predefining the optimal design parameters within the modified Voronoï tessellation (Fig. 6a), is called as the optimal artificial RVE in the following content. Linear elements with 0.25 µm and plane strain assumption are applied in this study. Meshed FE models of the optimal artificial RVE of DP590 is shown in Fig. 7.

Tensile test simulation is performed on the optimal artificial RVE model by prescribing a periodic boundary condition. In the periodic boundary condition, a global plastic strain of 17% is imposed to deform the RVE models, which guarantees the material below the UTS. The resulted flow curves is compared with the experimental measurement, as shown in Fig. 8. A small discrepancy is observed between the simulated results and the measured curve. This can be attributed to the utilization of plane strain assumption, while the real specimen is subjected to 3D stress states. The similar difference has also been reported by Ramazani [6],

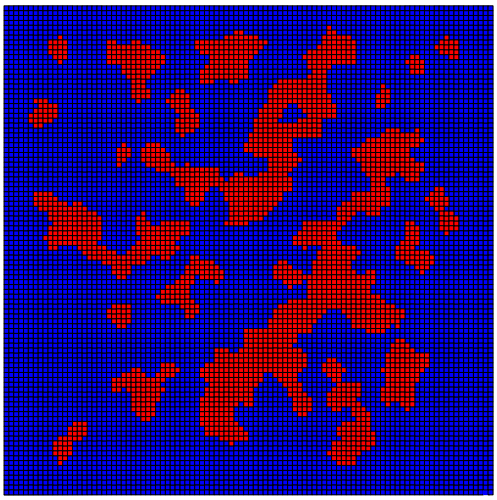


Figure 7: FE model with 0.25 µm element size and 25 µm edge length of the optimal artificial RVE for DP590 steel.

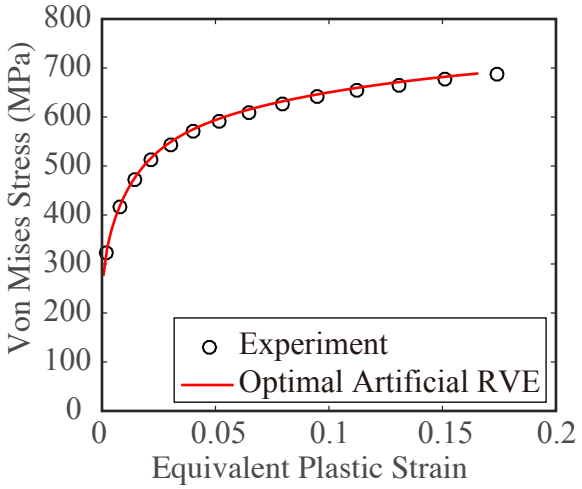


Figure 8: Comparison of flow curves between the experimental measurement and the predicted result of the optimal artificial RVE.

and it increases in the volume fraction of martensite. Nevertheless, the optimal artificial RVE provides an accurate prediction of flow behavior for the DP590 steel.

Fig. 9 show the equivalent plastic strain distribution in the optimal artificial RVEs at various global plastic strains. Although no material fracture or damage model is introduced, it is clearly found that shear bands are formed in the ferrite matrix near martensite grains (Fig. 9c). The shear bands and localized plastic strain are caused by the heterogeneous microstructure of DP590 steel. Further details are shown in Fig. 10, which indicates the equivalent plastic strain distribution at global strain of 15.8% in the optimal artificial RVE, and shear bands are located in the red ellipses. The direction of these shear bands is around 45° to the tensile loading. Integration of a ductile fracture model (e.g., GTN model) in the RVE model can be reasonably assumed to show nucleation of voids and microcracks in the strain localization zone, with the accumulation of plastic flow. Moreover, it is notable that, the validation of the optimal artificial microstructure is performed with the uniaxial tensile test. Due to the heterogeneity of the DP microstructure, additional investigations can be implemented in the case of biaxial or shear tests. This work will be performed in future studies.

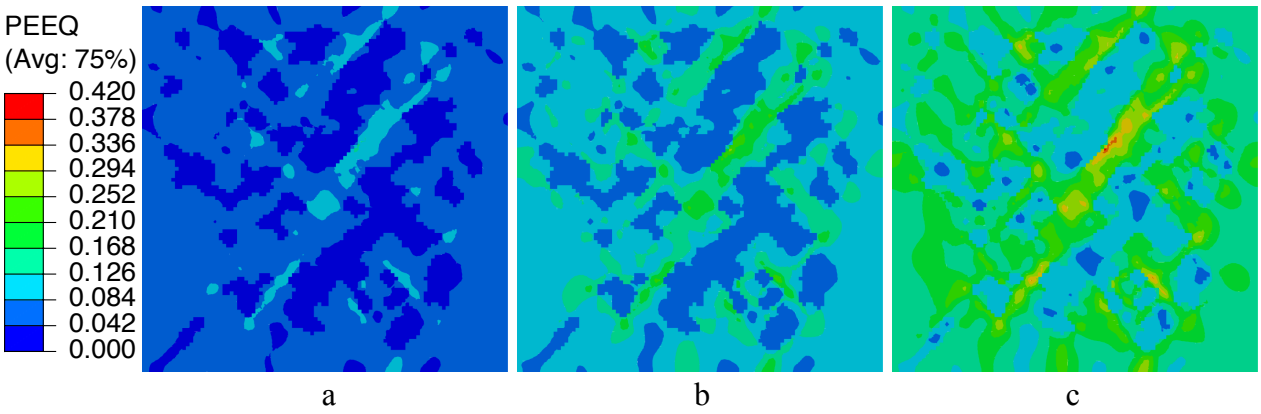


Figure 9: Equivalent plastic strain distribution in the optimal artificial RVE at global plastic strain levels of: (a) 5.09%, (b) 10.01% and (c) 15.03%

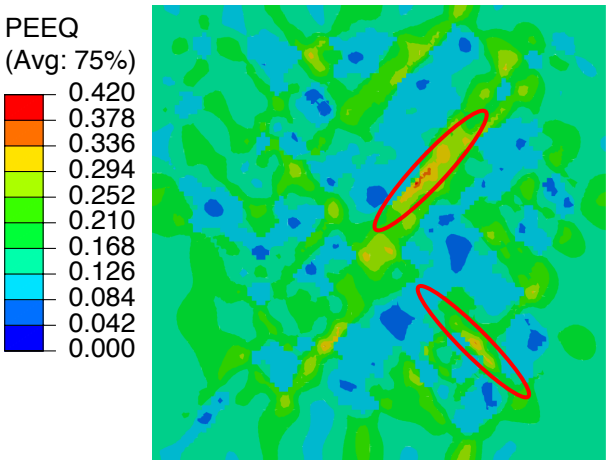


Figure 10: Equivalent plastic strain distribution at global plastic strain of 15.8% the optimal artificial RVE, shear bands in red ellipses.

5 CONCLUSION

In this study, we proposed a novel artificial generator inspired by topology optimization to reconstruct the microstructure of a DP steel. A modified Voronoï tessellation generated from Halton sequence, was utilized as the polycrystalline aggregate in the generator. Two microstructure parameters were introduced to control the phase assignment process in the DP steel. The novel phase assignment algorithm deduced the discrete value problem to a continuous one by defining a so-called "pseudo-density" within the modified Voronoï tessellation. Artificial DP microstructures were convergently generated with the optimal design parameters.

In order to validate the novel artificial generator, comparison of overall flow behavior was preformed between the experimental measurement and numerical simulation for a DP590 steel. It was shown that the optimal artificial RVE simulation can provide an accurate prediction. Moreover, shear bands were observed along the interface of different phase in each case. The path of shear bands is around 45° to the loading direction. It also concurs that, with the introduction of ductile fracture model, voids and microcracks can be formed in the strain localization zones. The promising results confirm the robustness of the proposed generator.

REFERENCES

[1] U. Liedl, S. Traint, E.A. Werner, An unexpected feature of the stress–strain diagram of dual-phase steel. *Computational Materials Science*. **25**, 180-192, 2002.

[2] A. Fillafer, C. Krempaszky, E.A. Werner, On strain partitioning and micro-damage behavior of dual-phase steels. *Materials Science and Engineering: A*. **614**, 122-128, 2014.

[3] X. Sun, K.S. Choi, A. Soulami, W. Liu, M.A. Khaleel, On key factors influencing ductile fractures of dual phase (DP) steels. *Materials Science and Engineering: A*. **526**, 140-149, 2012.

[4] H.T. Hossein, B. Anbarlooie, Javad. Kadkhodapour, G. Shadalooyi, Microstructural deformation pattern and mechanical behavior analyses of DP600 dual phase steel. *Materials Science and Engineering: A*. **600**, 108-121, 2014.

[5] A. Ramazani, K. Mukherjee, U. Prahl, W. Bleck, Modelling the effect of microstructural banding on the flow curve behaviour of dual-phase (DP) steels. *Computational Materials Science*. **52**, 46-54, 2012.

[6] A. Ramazani, K. Mukherjee, H. Quade, U. Prahl, W. Bleck, Correlation between 2D and 3D flow curve modelling of DP steels using a microstructure-based RVE approach. *Materials Science and Engineering: A*. **560**, 129-139, 2013.

[7] F.M. Al-Abbasi, J.A. Nemes, Micromechanical modeling of dual phase steels. *International Journal of Mechanical Sciences*. **45**, 1449-1465, 2003.

[8] G. Voronoï, Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Recherches sur les parallélloèdres primitifs, Deuxième mémoire*. **134**, 198-287, 1908.

[9] J.H. Halton, G.B. Smith, Radical inverse quasi-random point sequence, Algorithm 247. *Commun. ACM*. **7**, 701, 1964.

[10] L. Xia, B. Raghavan, P. Breitkopf, W. Zhang, Numerical material representation using proper orthogonal decomposition and diffuse approximation. *Applied Mathematics and Computation*. **224**, 450–462, 2013.

[11] L. Xia, P. Breitkopf, Design of materials using topology optimization and energy-based homogenization approach in Matlab. *Structural and multidisciplinary optimization*. **52**, 1229-1241, 2015.

[12] L. Meng, P. Breitkopf, B. Raghavan, G. Mauvoisin, O. Bartier, X. Hernot, Identification of material properties using indentation test and shape manifold learning approach. *Structural and multidisciplinary optimization*. **21**, 120-127, 2001.

[13] O. Sigmund, A 99 line topology optimization code written in Matlab. *Structural and multidisciplinary optimization*. **21**, 120-127, 2001.

[14] P.W. Chung, K.K. Tamma, R.R. Namburu, Asymptotic expansion homogenization for heterogeneous media: computational issues and applications. *Composites Part A: Applied Science and Manufacturing*. **32**, 1291-1301, 2001.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ON THE STRUCTURAL DESIGN OF IMPERFECTION SENSITIVE LAMINATED COMPOSITE SHELL STRUCTURES SUBJECTED TO AXIAL COMPRESSION

Linus Friedrich[1], Pawel Lyssakow[1,2], Garth Pearce[2], Martin Ruess[3], Chiara Bisagni[4], Kai-Uwe Schröder[1]

[1]Institute of Structural Mechanics and Lightweight Design, RTWH Aachen University
Wüllnerstrae 7, 52062 Aachen, Germany
e-mail: Linus.Friedrich@sla.rwth-aachen.de, Pawel.Lyssakow@rwth-aachen.de,
Kai-Uwe.Schroeder@sla.rwth-aachen.de

[2]School of Mechanical and Manufacturing Engineering, UNSW Sydney
Sydney NSW 2052, Australia
e-mail: g.pearce@unsw.edu.au

[3]School of Engineering, University of Glasgow
Ranking Building, Oakfield Avenue Glasgow G12 8LT, United Kingdom
e-mail: martin.ruess@glasgow.ac.uk

[4]Aerospace Structures and Computational Mechanics, Delft University of Technology
Kluyverweg 1, 2629HS Delft, The Netherlands
e-mail: C.Bisagni@tudelft.nl

Keywords: Shell buckling, imperfection sensitivity, composite shells.

Abstract. *Shell structures are used as primary structures of space launch vehicles. These structures are thin-walled and are thus prone to buckling when loaded in compression. Because of the imperfection sensitivity of these structures, small deviations of the real shell from the theoretically perfect shell may result in a tremendous decrease in load carrying capacity. For this reason, geometrical imperfections need to be taken into account. When designing unstiffened composite shells, the laminate stacking sequence influences both, the buckling load of the geometrically perfect shell and the imperfection sensitivity of the shell. Consequently, to derive laminate stacking sequences that maximize the buckling load of real shell structures, geometrical imperfections need to be taken into account already in an early design phase. In this paper, two laminate stacking sequences that were derived to maximize the buckling load of the geometrically perfect and imperfect shell structure are studied using stochastic methods. To this end, combination of non-rotational symmetric imperfections derived from measured data and variations of the ply orientation are studied in a stochastic analysis on basis of Monte Carlo simulations. The results of this study will be used to evaluate the influence of the stacking sequence as one of the essential properties dominating the structural response of geometrically imperfect laminate composite shell structures.*

1 INTRODUCTION

The structural design of imperfection sensitive shells is one of the problems constantly addressed in the field of mechanics. Due to the imperfection sensitivity of thin walled shells, deriving reliable design loads is a demanding task and it is mainly focused on how imperfections can be taken into account. To this end, several approaches were developed as, for example, summarized by Hoff [1] and Bushnell [2]. Recent approaches to account for geometrical imperfections are summarized by Castro [3] and Friedrich [4].

When designing imperfection sensitive isotropic shell structures, the design variable is the shell wall thickness and the shell's dimensions, which are defined on a system level. The load carrying capacity of laminated composite shell structures is additionally influenced by the laminate stacking sequence, see Figure 1. The laminate stacking sequence influences:

- the buckling load of the geometrically perfect shell and the shell's imperfection sensitivity, which results in a theoretical buckling load prediction
- the imperfection pattern of a manufactured shell and thus, the real buckling load is influenced by the shell's laminate stacking sequence

Consequently, it can be stated that the optimization of a laminate stacking sequence for a shell structure with regard to the load carrying capacity of the real shell is associated with many uncertainties. This drawback eliminates the possibility of performing structural optimizations, especially in an early structural design phase as such optimizations are time consuming and computationally costly and do not necessarily yield useful results because of the high number of uncertainties involved.

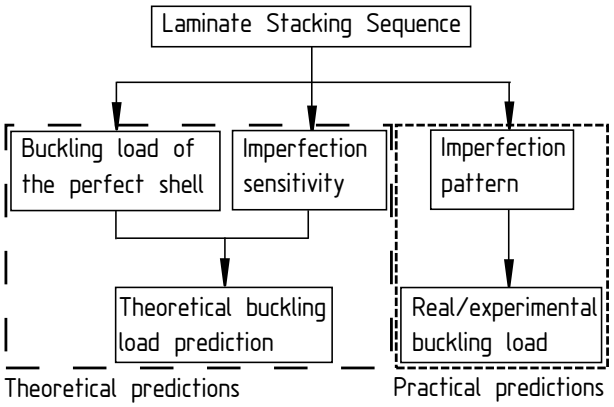


Figure 1: Influence of the laminate stacking sequence of composite shell structures on the structural response

Due to the high number of uncertainties, simplifications with regard to the imperfection pattern and boundary conditions need to be accepted to maximize the buckling load of imperfection sensitive composite shell structures. Based on these simplifications laminate stacking sequences can be identified, which are thought to maximize the buckling load of real shell structures. Shell structures derived based on two different approaches are further studied in this paper. One, a shell structure identified by Zimmermann in the 1990s [5] and two, a shell structure recently identified by Friedrich et. al. [6]. This paper is structured as follows: the study structures taken into account are introduced in section two. Subsequently, results determined using stochastic analyses are presented in section three and the paper closes with a summary and outlook in section four.

Table 1: Overview of material properties [5]

Properties	Composite shells
E_{11}	123 551 MPa
E_{22}	8 708 MPa
G_{12}	5 695 MPa
ν_{12}	0.32

2 INTRODUCTION OF STUDY STRUCTURES

Two different approaches to derive the study structures here discussed were taken into account. One, the laminate stacking sequence were derived to maximize the buckling load of geometrically perfect shell structures. The shell structures have been derived by Zimmermann in the 1990s [5, 7] and the resulting shells were denoted Z-shells. Two, the laminate stacking sequence were derived to maximize the buckling load of geometrically imperfect shells with rotational symmetric imperfections. The shell structures have been derived by Friedrich et. al. [6] and the resulting shells were denoted FL-shells.

The material properties used in Zimmermann [5] and Friedrich [6] are summarized in Table 1. The geometrical properties are defined as follows: the inner radius of the shell was 250 mm and the length of the shell was 510 mm. The length to radius ratio is about two and thus, the influence of the length of the shell and the type of boundary conditions do not affect the buckling load of the shell structure as shown in [8].

The results of deriving laminate stacking sequences that maximize the buckling load of a geometrically perfect and geometrically imperfect shells are summarized in Table 2 for those shells with two and five angle plies, that is $\mathrm{n_p} = 2$ and $\mathrm{n_p} = 5$. Additionally, the laminate stacking sequences influence on the buckling load of the geometrically perfect and geometrically imperfect shell with rotational symmetric imperfections are given in Figures 2 and 3 for two angle plies and an imperfection magnitude of 0.2 times the shell wall thickness that is $\mathrm{w_0/t} = 0.2$. The buckling loads of the $\mathrm{n_p} = 2$-shell are presented in detail since the buckling loads could be represented with reference to the ply orientation of the inner and outer ply and this allowed a suitable graphical evaluation of results.

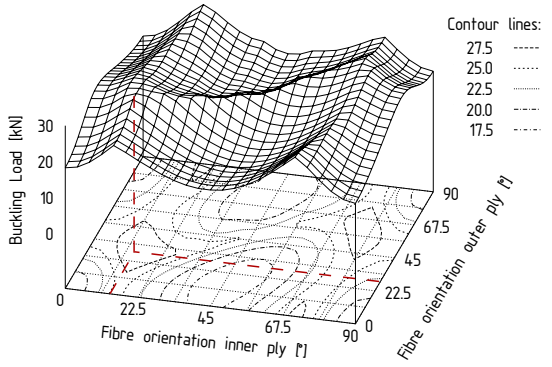


Figure 2: Buckling load of the perfect shell structure - $\mathrm{n_p} = 2$ according to Friedrich[6]

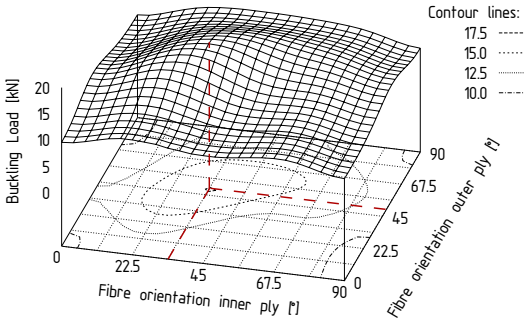


Figure 3: Buckling load of the imperfect shell structure - $\mathrm{w_0/t} = 0.2$ - $\mathrm{n_p} = 2$ according to Friedrich[6]

Table 2: Maximum critical buckling loads for geometrical perfect and imperfect shell structures with w_0/t= 0.2 according to Friedrich et. al. [6]

n_p		Shell	Stacking sequence	$F_{cr,perf}$ [kN]	$F_{cr,imp}$ [kN]
	perf.*		$\pm 15°/\pm 26°$	29.85	13.70
2	imp.**	FL2	$\pm 34°/\pm 49°$	21.66	17.56
				-27.45%	+28.14%
	perf.*	Z17	$\pm 30°/90°_2/\pm 23°/\pm 38°/\pm 53°$	286.74	129.18
5	imp.**	FL5	$\pm 38°/\pm 68°/90°_2/90°_2/\pm 38°$	197.70	139.81
				-31.05%	+8.14%

*stacking sequence derived to maximize the buckling load of the perfect shell
**stacking sequence derived to maximize the buckling load of the imperfect shell

The following findings were made when evaluating Figures 2 and 3:

- the laminate stacking sequences maximizing the buckling load of the geometrically imperfect shell were fundamentally different to those derived when maximizing the buckling load of the perfect shell in Zimmermann [5]. This finding was in contrast to the findings made by Khot [9] where the same laminate stacking sequence of a $[0°, \theta°, -\theta°]$ maximized the buckling load of a geometrically perfect and imperfect shell.
- the buckling load of the geometrically perfect shells exhibited many local maxima, whereas the imperfect shells only exhibited one global maximum buckling load. In addition, the maxima for the imperfect shell was far more robust, that is less sensitive to small changes in ply angles.

Similar findings as those summarized above could be made for the shell structures studied with three to five angle plies, that is for $n_p = 3, 4, 5$. The laminate stacking sequences maximizing the buckling load of the geometrically perfect and imperfect shell differed for all number of angle ply laminates. The results summarized in Table 2 indicated that the buckling load of the FL5-shell of the geometrical imperfect structure with rotational symmetric imperfections and an imperfection magnitude of $w_0/t = 0.2$ was 8% higher than the corresponding buckling load of the Z17-shells. Since the shell with five angle plies maximizing the buckling load of the geometrically perfect shell, the Z17-shell is often used as reference structure, the structural performance of the Z17-shell and the FL5-shell are studied further in this paper. To this end, geometrical imperfections other than simplified rotational symmetric imperfections, which were used to identify the FL-shells, are taken into account in combination with deviations of the laminate stacking sequence of their nominal value.

3 STOCHASTIC ANALYSES

Geometrically nonlinear finite element computations were performed to compare the structural performance of the Z17- and FL5-shells further. In addition to geometrical imperfections, deviations of the stacking sequence lay up were taken into account, which may occur during the manufacturing process. Since neither the geometrical imperfections nor deviations of the laminate stacking sequence from the nominal stacking sequence were known, both of the influencing parameters were varied in a stochastic manner. The geometrical imperfections in a shell were idealized as non-rotational symmetric imperfections, based on Schillinger's method

of separation [10] and adapted to laminated composite shell structures as done in [11]. To this end, based on measured imperfection patterns of composite laminated shell structures studied in a DLR-ESA study [12] and using the method of separation, stochastic imperfection patterns were defined and used to model the geometrically imperfect shell structure. The laminate stacking sequence orientations were varied in 1°-steps within the range $\pm 5^\circ$ around the nominal stacking sequence assuming a uniform distribution. The nonlinear structural response curves were determined using dynamic implicit analyses in a displacement controlled manner and the commercial software package Ansys [13]. Linear shell elements were used and 171 elements in circumferential and 63 elements in axial direction were used to model the shell structure.

The structural response of 2000 different shell structures with varying combinations of geometrical imperfections and laminate stacking sequence imperfections are studied for the Z17- and FL5-shell by evaluating the resulting load displacement curves due to an implicit dynamic computation . The nonlinear buckling loads are plotted over the number of computations performed, c.f. Figure 4. For both the Z17- and FL5-shell, 2000 nonlinear analyses were performed for each shell configuration and evaluated subsequently.

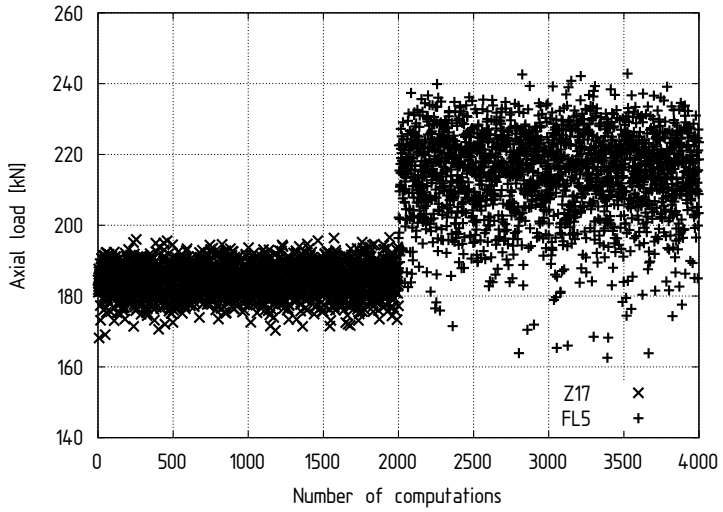


Figure 4: Stochastic studies - buckling load vs. number of computations - geometrical and lay-up imperfections

Since the scatter of the buckling loads of the Z17-shell were very low compared to the scatter observed for the FL5-shell, numerical analyses were performed using an explicit dynamic solver and the commercial software package Abaqus [14] to further evaluate the results. Because of the high resulting numerical effort, recently only 40 configurations were studied for the Z17- and FL5-shell each. The comparison of buckling loads of the Z17- and FL5-shell due to explicit dynamic and implicit dynamic analyses are summarized in Figure 5. The results given in Figure 5 are preliminary results, the resulting standard deviations and the corresponding design loads are not converged and thus in future studies additional computations need to be performed.

The mean values of the buckling loads, $\mathrm{F}_{\mathrm{mean}}$, the standard deviations σ, the resulting buckling loads for 95.5 $\%$ and 99.7 $\%$ probability, $\mathrm{F}_{95.5\%}$ and $\mathrm{F}_{99.7\%}$, are summarized in Table 3 for the stochastic analyses performed using implicit and explicit solvers. The buckling loads $\mathrm{F}_{95.5\%}$ and $\mathrm{F}_{99.7\%}$ are thereby determined by reducing $\mathrm{F}_{\mathrm{mean}}$ with two times or three times the standard deviation, σ.

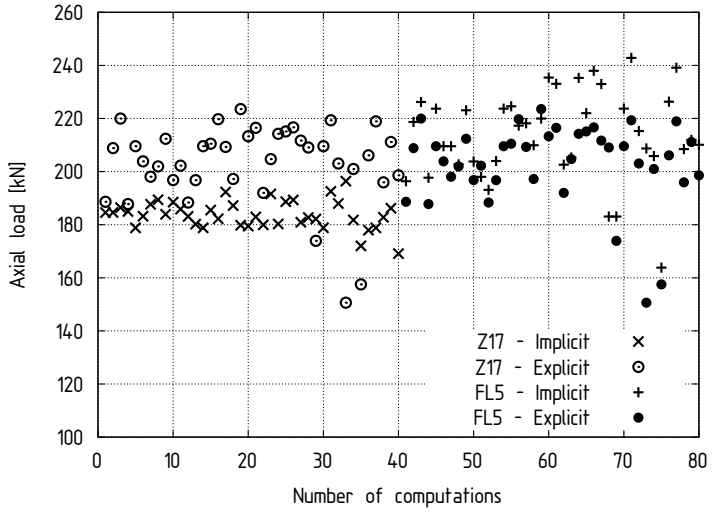


Figure 5: Stochastic studies - buckling load vs. number of computations - geometrical and lay-up imperfections - explicit vs. implicit dynamic analyses using Abaqus [14] and Ansys [13]

Table 3: Evaluation of stochastic studies performed using a combination of geometric imperfections and stacking sequence lay-up imperfections

Type of analysis	No. of comp.	Z17				FL5			
		F_{mean} [kN]	σ [kN]	$F_{95.5\%}$ [kN]	$F_{99.7\%}$ [kN]	F_{mean} [kN]	σ [kN]	$F_{95.5\%}$ [kN]	$F_{99.7\%}$ [kN]
dynamic implicit	2000	184.41	4.47	175.47	171.00	214.23	12.12	189.99	177.87
dynamic explicit	40	204.66	9.29	186.08	176.79	203.10	15.56	171.98	156.43

The following findings were made when evaluating the results given in Figures 4 and 5 and those results summarized in Table 3:

- the mean value of the FL5-shell's buckling load was higher than the mean value of the Z17-shell's buckling load for the dynamic implicit analyses
- the standard deviation of the FL5-shell's buckling load was higher than the standard deviation of the Z17-shell's buckling load that is the scatter of the FL5-shell's buckling loads was higher than for the Z17-shell
- the preliminary results due to the explicit dynamic analyses indicated that the standard deviation of the Z17-shell increased more significantly than the standard deviation of the FL5 shell when changing from implicit to explicit dynamic analyses
- the design load for a 99.7% probability level of the FL5 shell were 4% higher than the design load for a 99.7% probability level of the Z17 shell when using implicit dynamic analyses and the commercial finite element package Ansys [13]
- the design load for a 99.7% probability level of the FL5 shell were 11.5% lower than the design load for a 99.7% probability level of the Z17 shell when using explicit dynamic analyses and the commercial finite element package Abaqus [14]

In Friedrich et. al. [6], it was shown that determining the nonlinear buckling load of the Z17-shell was computationally demanding. Due to the rather small bending stiffness, the shell revealed a distinct rotational symmetric pre-buckling state such that determining the nonlinear buckling load became difficult. This finding was confirmed by the studies here performed using implicit and explicit dynamic solving algorithms and using the commercial finite element software packages Abaqus [14] and Ansys [13]. Furthermore, it was shown that evaluating the structural performance of the two composite shell structures, Z17- and FL5-shells, is a demanding task, which is influenced by the actual physical behaviour and numerical difficulties when determining the structural response. The final numerical evaluation of the structural performance of the Z17- and FL5-shells could not be completed in this paper; but, additional studies are needed to allow for a final numerical evaluation.

4 SUMMARY AND OUTLOOK

In this paper, the structural performance of two shell structures, the Z17- and FL5-shell was studied for geometrical imperfections other than rotational symmetric imperfections and stacking sequence lay-up imperfections. Dynamic implicit and explicit analyses were performed and the results were evaluated with regard to the statistical buckling load predictions. The discrepancy between implicit dynamic and explicit dynamic results discussed above needs to be readdressed and a solving algorithm leading to robust buckling load predictions needs to be chosen for future studies. Thus, a final numerical assessment on the structural performance of the Z17- and FL5-shell may be realized.

Additionally, it would be desirable to compare the structural performance of the Z17- and FL5-shell in an experiment. To this end, these two shell structures need to be manufactured using the same manufacturing process and the same mandrel. Thus, additional insight on the structural behaviour of the Z17- and FL5-shell can be gained and serve as basis for further discussions.

REFERENCES

[1] N. J. Hoff. *Perplexing Behavior of Thin Circular Cylindrical Shells in Axial Compression*. AFOSR-66-0155. Department of Aeronautics and Astronautics, Stanford University, 1966.

[2] David Bushnell. "Buckling of Shells-Pitfall for Designers". In: *AIAA Journal* 19.9 (1981). DOI: 10.2514/3.60058. URL: http://dx.doi.org/10.2514/3.60058.

[3] S. Castro et al. "Exploring the constancy of the global buckling load after a critical geometric imperfection level in thin-walled cylindrical shells for less conservative knockdown factors". In: *Thin-Walled Structures* 72 (2013), pp. 76–87.

[4] L. Friedrich, T.-A. Schmid-Fuertes, and K.-U. Schröder. "Comparison of theoretical approaches to account for geometrical imperfections of unstiffened isotropic thin walled cylindrical shell structures under axial compression". In: *Thin-Walled Structures* 92.0 (2015), pp. 1 –9. ISSN: 0263-8231. DOI: http://dx.doi.org/10.1016/j.tws.2015.02.019.

[5] R. Zimmermann. "Optimierung axial gedrückter CFK-Zylinderschalen". PhD thesis. Universität-Gesamthochschule Siegen, 1991.

[6] L. Friedrich et al. "Stacking sequence influence on imperfection sensitivity of cylindrical composite shells under axial compression". In: *Composite Structures* 134 (2015), pp. 750

–761. ISSN: 0263-8223. DOI: http://dx.doi.org/10.1016/j.compstruct.2015.08.132.

[7] R. Zimmermann. "Buckling Research for imperfection Tolerant Fiber Composite Structures". In: *Spacecraft Structures, Materials and Mechanical Engineering, Proceedings of the Conference held by ESA, CNES and DARA in Noordwijk, 27-29 March 1996*.

[8] L. Friedrich and Hans.-G. Reimerdes. "Imperfection Sensitivity of Circular Cylindrical Shells of Varying length Subjected to Axial Compression". In: *54th AIAA Structures, Structural Dynamics, and Materials Conference*. 2013. DOI: 10.2514/6.2013-1771.

[9] N. S. Khot and V. B. Venkayya. *Effect of fiber orientation on initial postbuckling behaviour and imperfection sensitivity of composite cylindrical shells*. Tech. rep. Technical Report AFFDL-TR-70-125. Air Force Flight Dynamics Laboratory, 1970.

[10] D. Schillinger and V. Papadopoulos. "Accurate estimation of evolutionary power spectra for strongly narrow-band random fields". In: *Computer Methods in Applied Mechanics and Engineering* 199.17–20 (2010), pp. 947 –960. DOI: http://dx.doi.org/10.1016/j.cma.2009.11.008.

[11] Jendi Kepple et al. "Stochastic analysis of imperfection sensitive unstiffened composite cylinders using realistic imperfection models". In: *Composite Structures* 126.0 (2015), pp. 159 –173. ISSN: 0263-8223. DOI: http://dx.doi.org/10.1016/j.compstruct.2015.02.063.

[12] R. Degenhardt, L. Bethge, and L. Kärger. *Probabilistics aspects of buckling knock down factors - tests and analysis, final report*. Tech. rep. ESA contract 19709/06/NL/IA, 2007.

[13] *Ansys - Software Package*. 2015.

[14] *ABAQUS 6.13 - Software Package*. Dassault Systems. 2013.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

STRENGTH AND WEIGHT EQUIVALENT SUBSTITUTION OF LARGE SANDWICH PANELS BY MONOLITHIC CFRP STRUCTURES

Martin Meindlhumer[1,2] **and Martin Schagerl**[2]

[1]Johannes Kepler University Linz
Institute of Constructional Lightweight Design
Linz, Austria
e-mail: martin.meindlhumer_1@jku.at, martin.schagerl@jku.at

[2] FACC Operations GmbH
Research and Development
Ried im Innkreis, Austria
e-mail: martin.meindlhumer@facc.com

Keywords: Aircraft Structures, Composite Structures, Sandwich Structures, Control Surfaces, Alternative Design Philosophies, Shell Structures, Buckling

Abstract. *A general trend in the aircraft industry is to replace sandwich structures made from fiber-polymer-composites and low-density honeycomb cores by monolithic structures entirely made from fiber-polymer-composites. The fields of applications addressed here are primary and secondary control surfaces of civil aircraft wings and stabilizers. However, they can be directly extended to other applications. The main motivation for such a substitution is to reduce manufacturing efforts. For honeycomb sandwich designs mainly resin pre-impregnated fiber semi-finished products are applied and often the structures are built and cured in an autoclave in one or several steps. This is cost intensive. Thus, resin infusion technology is an attractive alternative. Dry fibers are saturated with resin during the curing process by applying vacuum. Here, just an oven is required. Another reason is that sandwich structures used on aircraft are exposed to large varying environmental conditions. Condensation and absorption of water can cause an increase in the total mass and once the water is frozen, serious damages may occur. Other advantages include ease in detection of structural damages and manufacturing defects and reduction in repair efforts. Control surface structures are loaded by pressure distributions over their surfaces. The sandwich design is optimal to withstand bending moments, shear forces and torsional moments. The challenge is to keep the mass, when the design is changed to monolithic.*

This contribution presents two design alternatives, one developed on top of the other, which allow the manufacturing of monolithic highly integrated laminated structures within one layup and curing step. Additionally, aerodynamic properties are guaranteed on the top and the bottom of the control surface. A reference structure is derived from a state-of-the-art spoiler, which is equipped with a honeycomb core. This structure is then compared with the alternative design proposed.

1 INTRODUCTION

A general trend in the aircraft industry is to replace sandwich structures by monolithic structures. In the context to this work a sandwich structure is built by a lower and upper skin made from layers of fiber-reinforced-polymers and a honeycomb core in between. A monolithic structure is entirely made from composite layers. The fields of applications addressed here are primary and secondary control surfaces of civil aircraft wings and stabilizers, but they can be extended to other applications directly. There are several reasons motivating such a substitution. One is simply to reduce the manufacturing efforts. Honeycomb cores require the use of resin pre-impregnated fiber products. The sandwich structure has to be cured in one or several steps in an autoclave, where defined pressures and temperatures are applied in a nitrogen-filled environment. This process is cost intensive. Another reason is their behavior during their operating life. Sandwich structures as used on an aircraft are exposed to large varying environmental conditions, which may cause negative implications. One example is the condensation of water, which can be absorbed by the aramid honeycomb core. Despite the mass increase, the absorbed water may freeze and cause serious damage to the sandwich structure. Monolithic structures will allow the application of alternative, more cost efficient manufacturing process technologies like resin infusion. Here, dry fiber products are saturated with resin during the curing process by applying vacuum. This approach is cheaper, because just an oven is required. An increased level of automation is given. In addition, the service life is better controllable in respect to maintenance and repair. Other advantages include ease in detection of structural damages and manufacturing defects and reduction in repair efforts.

Control surface structures are loaded by pressure distributions acting on their surfaces. The sandwich design is, from the structural point of view, optimal to withstand bending moments, shear forces and torsional moments. The challenge is to keep the weight, when the design is changed to monolithic. In addition, aerodynamic properties, which are needed on the top and the bottom surfaces, increase the demand for design solutions which are simpler to manufacture. This contribution presents and discusses two alternative design approaches, which allow the fabrication of monolithic highly integrated laminated structures within one layup and curing step. Compared to [1], the approaches are developed under analytic considerations on the basis of a simplified geometry derived from an aircraft spoiler, which is equipped with a honeycomb core. First, this sandwich reference structure is analyzed for static displacements and occurring stresses. Its weight is calculated. Second, two design alternatives are presented and analyzed. The latter one is derived from the drawbacks of the first one. Finally, the results are justified with the analysis data of the sandwich reference structure and a conclusion is drawn.

2 AIRCRAFT SPOILER AS CASE EXAMPLE

An aircraft spoiler is a typical representative of control surfaces where the sandwich design is state of the art. Spoilers are often multifunctional. Immediately after landing, or in an event of an aborted take-off, they are deflected in an almost upright position. As a result of the interrupted airflow over the flaps, the wing loses a large part of its lift, which increases the normal force on the tires and makes breaking more effective. In addition, they create considerable drag and these combined effects increase the deceleration up to 20%. Outboard spoilers may be used in flight when an appreciable increment of drag is required to obtain a high rate of descent or improved speed stability with a constant angle of decent[2]. In the present work the design and functionality of an aircraft wing spoiler is transferred to a simplified analysis model more suitable for the purpose of the current demonstration. However, the work presented in the

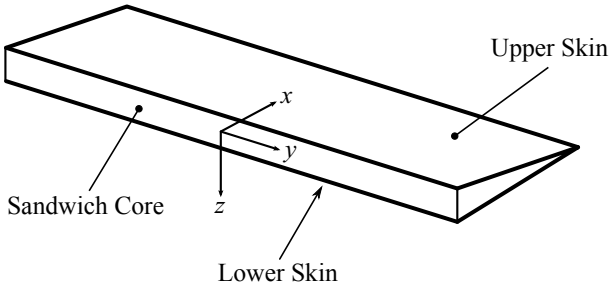


Figure 1: Simplified geometry of an aircraft control surface.

E_{11}	E_{22}	ν_{12}	ρ	CPT	R^t_{11}	R^c_{11}	R^t_{22}	R^c_{22}	R_{12}
154	8.5	0.35	1.58	0.127	2610	1450	55	285	105
GPa	GPa	-	$\mathrm{g/cm^3}$	mm	MPa	MPa	MPa	MPa	MPa

Table 1: Material data of the unidirectional carbon fiber epoxy resin tape.

following can be extended straightforwardly to other control surface structures and to various applications.

2.1 A simplified sandwich structure for reference

The reference design is based on the overall dimensions, material data, loads and boundary conditions of a state-of-the-art aircraft spoiler. Thus, a simplified problem, but of similar scale is investigated. In Fig. 1 the simplified geometry, their components and its position in the Cartesian coordinate system is presented. Its overall geometric dimensions are $2400 \times 800 \times 120$ mm. The base surface of the triangular prism is coplanar to the yz-plane of the defined coordinate system. The structure is symmetrical about the xy-plane and the xz-plane. Unless otherwise stated, the structure is presented in the shown position in the following. The laminate of the lower and upper skin is produced by unidirectional tapes of epoxy resin-impregnated carbon fibers stacked in the sequence of $(45°/\text{-}45°/90°)_s$, where the $0°$-direction points in the x-direction. An aramid honeycomb is selected for the sandwich core. Its ribbon direction is aligned with the y-axis. The corresponding shear strength is higher in this direction compared to the perpendicular one. The material properties and strength allowable of the carbon epoxy tape and the honeycomb core are provided in Tab. 1 and 2, respectively. The axes of the local material coordinate systems are named as 11 (in fiber direction for the carbon plies or ribbon direction of the honeycomb core), 22 (in-plane transverse fiber direction or ribbon direction) and 33 (out-of-plane transverse fiber direction or ribbon direction). Accordingly, shear directions are defined with the indices 12, 23 and 13. The superscripts c and t indicate compression and tension, respectively. The core and the skins are bonded together with film adhesive. However, the structural behavior of the adhesive joint is not considered in this study, as a perfect binding at material interfaces being assumed. For mass calculation, the adhesive density is given as $250\ \mathrm{g/m^2}$.

Despite the generality in practical application of the presented design approaches, the boundary conditions are particularly specified. Here, air loads acting on the upper skin are transferred to the wing structure via three stiff metal brackets located at the center and each corner. In contrast to air suction-up loads in retracted position, pressure loads occur when the spoiler is in extended position immediately after landing, which is the most critical loading situation. As

E_{11}	E_{22}	E_{33}	G_{12}	G_{23}	G_{13}	ρ	S_{33}^c	S_{13}	S_{23}
1	1	138	1	19	31	0.048	1.24	0.98	0.5
MPa	MPa	MPa	MPa	MPa	MPa	$\mathrm{g/cm^3}$	MPa	MPa	MPa

Table 2: Material data of the aramid honeycomb core.

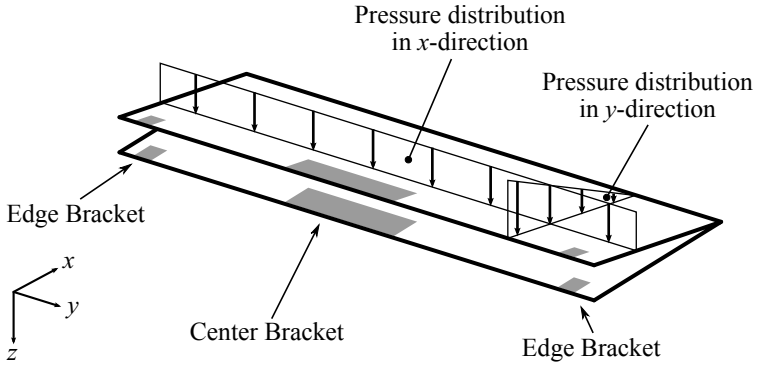


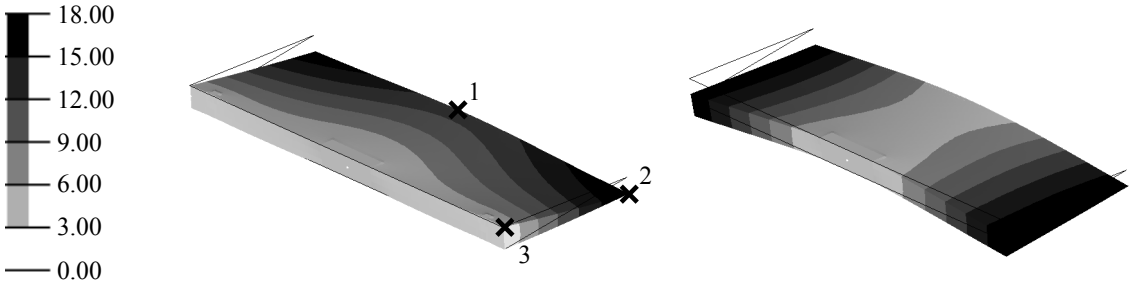
Figure 2: Schematic sketch of the applied pressure field and the bracket locations, where the structure is constrained.

shown in Fig. 2, the pressure is triangular distributed in x-direction with its maximum at the leading edge ($x = 0$) and rectangular in y-direction. The triangular distribution originates from spoiler extension. The rectangular one comes from the assumption, that the neighboring spoilers are in extended mode as well. Two load cases are considered. In the first case, the structure is intact. All degrees of freedom of the nodes at the two locations named center bracket in Fig. 2 are constraint. The nodes at the edge brackets are constraint in z-direction only. A maximum pressure of $0.0345\ \mathrm{N/mm^2}$, which represents the ultimate load, is applied. In the second case, one or both of the edge brackets failed. Therefore, the constraints of the nodes at the edge bracket locations are released. In consideration of sub-component failure only the limit loads are applied, which are 1.5 times lower than ultimate loads. Thus, the maximum acting pressure is $0.023\ \mathrm{N/mm^2}$. Details of the definition of limit load and ultimate load can be found in the EASA Certification specifications for large airplanes in chapter 25.301 and 25.302 [3].

The Finite Element discretization of the lower skin, upper skin and the alternative design features proposed in the following sections is utilized with standard first-order shell elements. The mesh of the honeycomb core in the reference sandwich design is built with first-order hexahedral elements. The average edge length of the volume and shell elements is $20\ \mathrm{mm}$. For pre- and post-processing the commercial software HyperWorks 13.0 (Altair Engineering Inc., Troy, MI, USA) is applied. The FEM simulations are carried out using the commercial solver OptiStruct (Altair Engineering Inc., Troy, MI, USA).

2.2 The structural behavior of the sandwich design

The sandwich structure described above is investigated regarding its deformation and regarding the stresses in the plies and the honeycomb core. Reserve factors are calculated from the materials resistance values (strength allowables) and finally, the structural mass is calculated. The absolute displacement plots for load cases 1 and 2 are shown in Fig. 3. To compare this design to the following ones, three, for the global displacement characteristic nodes are selected. Their absolute displacement values can be found in Tab. 4. The maximum in-plane ply stresses



a) Load Case 1 b) Load Case 2

Figure 3: Absolute displacement plot of the sandwich reference structure for a) load case 1 and b) load case 2 in [mm], scale factor: 10, markings show the 3 characteristic nodes for further evaluation.

occurring for load cases 1 and 2 are summarized in Tab. 5. The stresses in the honeycomb core shall be considered as well. Here, the maximum stress values are

$$\sigma_{33}^{c} = 0.2\text{MPa} \qquad \sigma_{13} = 0.4\text{MPa} \qquad \sigma_{23} = 0.3\text{MPa}$$
$$\sigma_{33}^{c} = 0.1\text{MPa} \qquad \sigma_{13} = 0.4\text{MPa} \qquad \sigma_{23} = 0.2\text{MPa}$$

for load cases 1 and 2, respectively, where the elements next to the constraint nodes are not considered. Reserve factors are calculated to show the overall static strength of the structure. For the carbon plies the maximum stress first ply failure criteria

$$RF_{ij}^{(c,t,\text{-})} = \frac{R_{ij}^{(c,t,\text{-})}}{\sigma_{ij}^{(c,t,\text{-})}} \tag{1}$$

is applied [4] for the maximum in-plane stresses summarized in Tab. 5. The reserve factors are given in Tab. 6. The material allowables are taken from Tab. 1. The systematic application of (1) yields the reserve factors for the honeycomb for load cases 1 and 2, respectively

$$RF_{33}^{c} = 6.2 \qquad RF_{13} = 3.3 \qquad RF_{23} = 1.3$$
$$RF_{33}^{c} = 12.4 \qquad RF_{13} = 4.9 \qquad RF_{23} = 1.3$$

Finally, the total mass of the sandwich reference design calculated from the individual mass densities of the components is given in Tab. 7. The data generated from the reference design is used to be compared with those of the alternative designs developed in the following section.

3 DESIGN ALTERNATIVES AND THEIR STRUCTURAL ANALYSES

The result of the above investigated sandwich structure reveals that the dominating global reactions to the given loading situations are bending around the x-axis and bending around the y-axis (see Fig. 1). The honeycomb core has the function of keeping the distance between the bottom and top laminates. However, a further main task is to resist the acting shear stresses with a minimum of structural mass. The substitution of the honeycomb core requires design features that are able to carry the shear stresses. In the following, two alternative designs are presented, which might be able to substitute the honeycomb core. In general, the same carbon epoxy ply material and film adhesive (if needed) are applied. The standard layup of the lower and upper skin is kept, but extended by additional layers as necessary. The load cases and boundary conditions are identical to that of the sandwich structure.

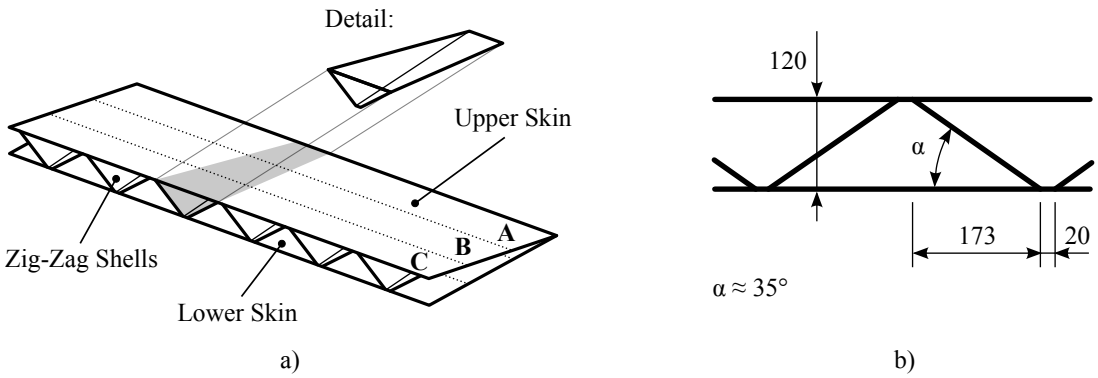


Figure 4: Monolithic box with a) zig-zag stiffeners and b) its local dimensions in [mm].

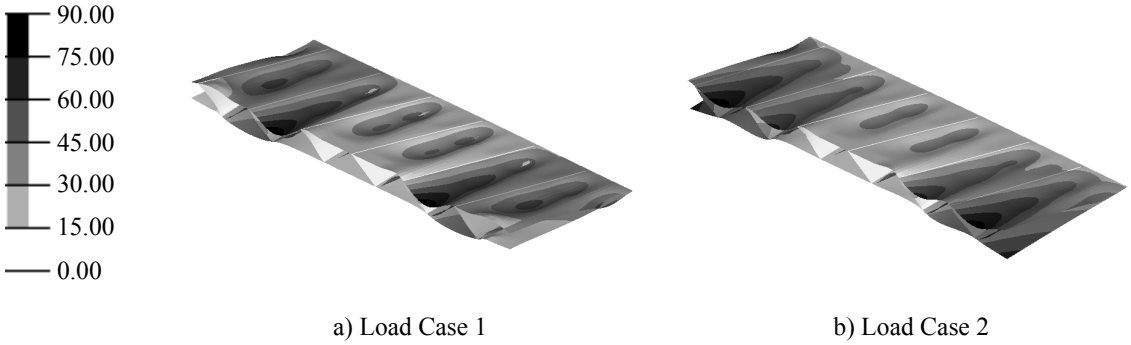
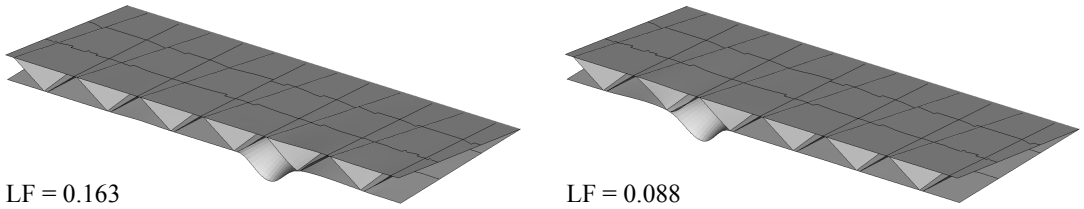


Figure 5: Absolute deformation plot of the zig-zag structure for a) load case 1 and b) load case 2 in $[\mathrm{mm}]$, scale factor: 1.

3.1 Monolithic box with zig-zag stiffeners

The first concept is inspired by structures made from trusses. In such constructions, diagonal trusses are used to produce frames with high resistance to shear forces. Here, the function of a diagonal truss is adopted with plane stiffeners arranged in inclined positions to form a zig-zag pattern as presented in Fig. 4a. A structure designed in this way can be manufactured in one component and one curing process. The overall dimensions are identical to the sandwich structure. The detail dimensions of the zig-zag pattern can be found in Fig. 4b. With $\alpha \approx 35°$ twelve ribs are added as indicated. A distance of $20\ \mathrm{mm}$ between the run-out of one rib and the onset of the next rib is defined to account for design radii needed for manufacturing purpose. The structure is divided along the x-direction (Fig. 1) in three equidistant zones with different ply layup sequences labeled **A**, **B** and **C** (Fig. 4a). The zone-based ply definition can be found in Tab. 3.

The finite element analysis for this design shows large displacements (see Fig. 5), especially at the upper skin and low buckling load factors at the lower skin (see Fig. 6). Regarding the global displacement, listed in Tab. 4, the values for the characteristic nodes 1 and 2 (Fig. 3) are approximately equal to the simplified sandwich structure. The large displacements at node 3 seem to be a result of the missing stiffness against flexural shear at the open end of the structure. Moreover, the local displacements of the upper skin sections between the inclined ribs caused by the externally acting air pressure are unacceptably high. Again, a main portion is caused by the open design at the leading edge. The lower skin sections (parts of the lower



a) Load Case 1 b) Load Case 2

Figure 6: Buckling shapes of the zig-zag structure for a) load case 1 and b) load case 2, scale factor: 100.

skin confined between two inclined ribs) are loaded by in-plane compression and in-plane shear. Thus, this structural elements are prone to buckling and due to the free edge, the risk of buckling is significantly increased. Therefore, the first buckling mode occurs at the leading edge location at very low buckling load factors of 0.163 and 0.088 for load case 1 and load case 2, respectively. These factors multiplied with the limit load show the load level at which buckling occur and should be larger than one. One option to increase the buckling load is to apply additional layers of carbon. However, a high number of plies is needed, because of the currently small load factor. Also other buckling modes, at other locations might become critical. Thus, this approach results in a non-efficient, i.e. not weight optimal structure. The same happens, when the upper skin thickness is increased locally until the displacements caused by the air pressure are reduced to an acceptable limit. Therefore, global design adjustments are necessary to achieve a weight-optimized result, which is actually competitive to the sandwich structure.

3.2 Monolithic torque box with ribs

The obvious solution is to close the open structure. Thereby, the buckling load is increased significantly and the displacements of the local upper skin sections are reduced as the flexural and shear stiffness are increased. This design update can be realized by inserting a C-spar as presented in Fig. 7a to the cost of a separate component needed to be manufactured. However, the complexity of the C-spar is relatively low and the assembly procedure is simple compared to the benefits provided. The C-spar contributes essentially to the global stiffness of the structure, whereas the impact on the mass is low, because of the decreased length of the inclined ribs. The increase in the global stiffness behavior is used to further improve the design regarding buckling and local deformation. The introduced C-spar allows to change the ribs from inclined ones to simply vertical ones as presented in Fig. 7. That means the width of the local upper and lower skin sections are nearly reduced by half in comparison to the zig-zag structure presented in Fig. 4, which has a major impact on buckling and local displacements. The reduced surface of the ribs decreases the resulting mass further.

Finally, the carbon ply layup of the structure is optimized. The standard layup defined by the sandwich structure is applied again. However, it is adjusted with additional plies locally according to the occurring buckling modes. The number of plies and their orientations depend on the principal normal stress. The ply layup of the structure is categorized by zones, which are shown in Fig. 8. The layup sequences of the individual labeled zones are provided in Tab. 3.

As shown in Fig. 8a the layup of zones on the lower skin is the most complex one. These zones are with higher number of plies, resulting in a larger laminate thickness than those on the upper skin. The highest number of plies is needed in the center bracket area (see Fig. 2). The number of plies decreases towards the trailing edge (from $x = 0$ to $x = 800\,\mathrm{mm}$), the

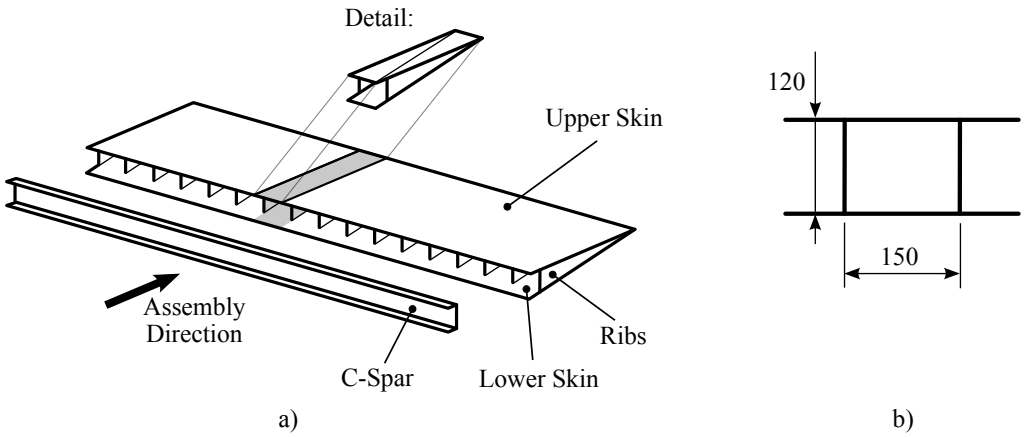


Figure 7: Monolithic torque box with a) ribs and b) its local dimensions in [mm].

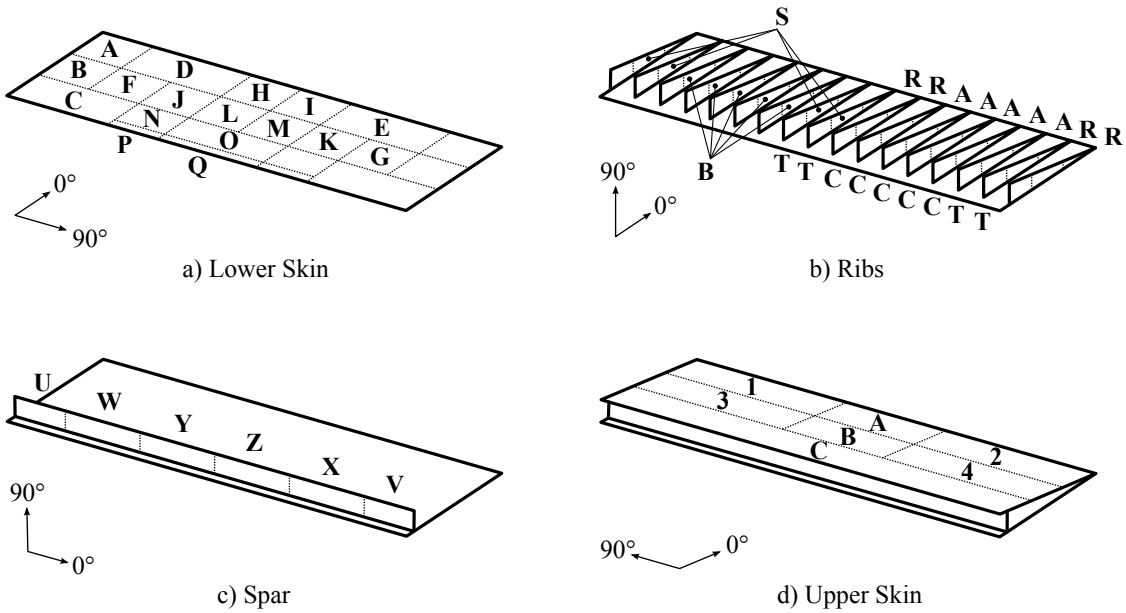


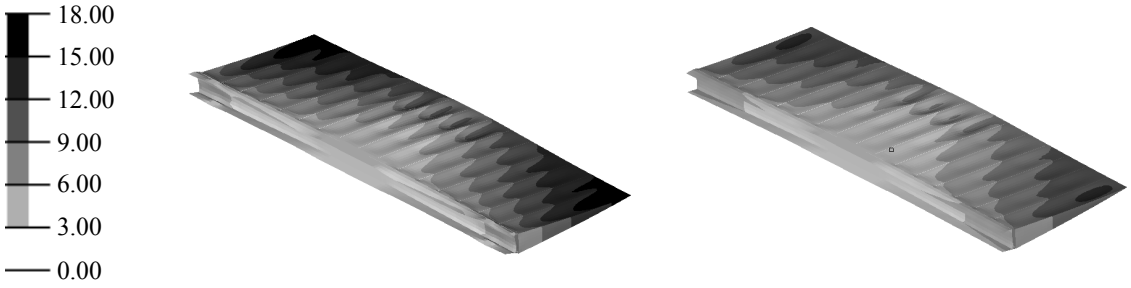
Figure 8: Zone based definition for ply layup sequence of the torque box structure. Not labeled zones are equal to the opposite ones.

Zone	Qty.	Ply Layup
A	6	(45/-45/90/90/-45/45)
B	10	(45/-45/90/90/90/90/90/90/-45/45)
C	14	(45/-45/90/90/90/90/90/90/90/90/90/-45/45)
D	8	(45/-45/90/45/45/90/-45/45)
E	8	(45/-45/90/-45/-45/90/-45/45)
F	11	(45/-45/90/90/90/45/90/90/90/-45/45)
G	11	(45/-45/90/90/90/-45/90/90/90/-45/45)
H	9	(45/-45/90/45/0/45/90/-45/45)
I	9	(45/-45/90/-45/0/-45/90/-45/45)
J	12	(45/-45/90/90/90/90/45/90/90/90/-45/45)
K	12	(45/-45/90/90/90/90/-45/90/90/90/-45/45)
L	13	(45/-45/90/90/90/90/45/0/90/90/90/-45/45)
M	13	(45/-45/90/90/90/90/-45/0/90/90/90/-45/45)
N	16	(45/-45/90/90/90/90/90/90/90/90/90/90/90/90/-45/45)
O	17	(45/-45/90/90/90/90/90/90/0/90/90/90/90/90/90/-45/45)
P	24	(45/90/90/90/90/90/-45/45/-45/90/90/90/90/90/90/90/90/90/90/90/90/-45/45)
Q	24	(45/90/90/90/90/90/-45/45/-45/90/90/90/90/90/90/0/90/90/90/90/90/90/-45/45)
R	4	(45/-45/-45/45)
S	8	(45/-45/90/90/90/90/-45/45)
T	12	(45/-45/90/90/90/90/90/90/90/90/-45/45)
U	6	(45/-45/-45/-45/-45/45)
V	6	(45/-45/45/45/-45/45)
W	8	(45/-45/-45/-45/-45/-45/-45/45)
X	8	(45/-45/45/45/45/45/-45/45)
Y	10	(45/-45/-45/-45/-45/-45/-45/-45/-45/45)
Z	10	(45/-45/45/45/45/45/45/45/-45/45)
1	8	(45/-45/90/-45/-45/90/-45/45)
2	8	(45/-45/90/45/45/90/-45/45)
3	11	(45/-45/90/90/90/-45/90/90/90/-45/45)
4	11	(45/-45/90/90/90/45/90/90/90/-45/45)

Table 3: Ply layup sequence for the individual zones.

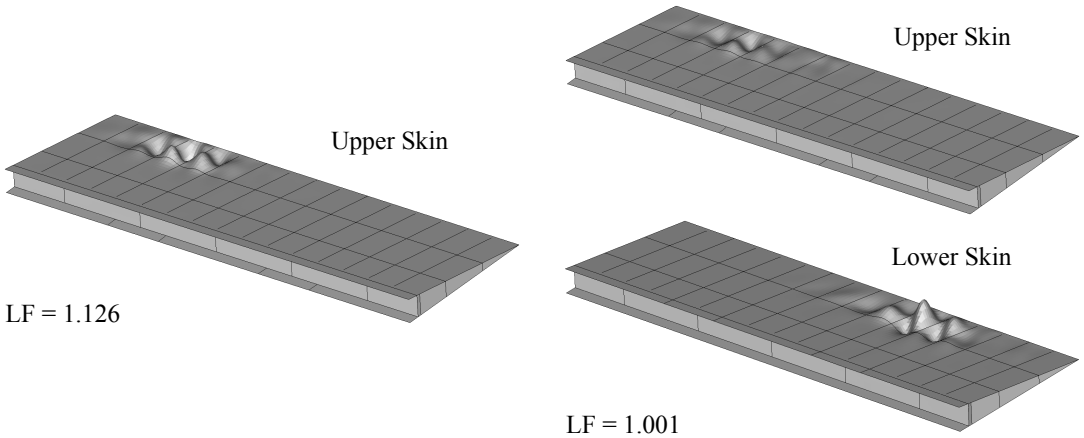
outboard edge and the inboard edge. Except the standard layup the ply sequence is mirrored along the xz-plane of the defined coordinate system given in Fig. 1. The structural mass of the final design is provided in Tab. 7 and is 520 g less than the sandwich structure.

The aforementioned torque box structure is analyzed with respect to displacement, buckling, and static strength. The absolute displacements under applied load cases are shown in Fig. 9 with a scale factor of 5. The displacements of the characteristic nodes can be found in Tab. 4. In general, the displacements are considered to be acceptable. Figures 10a and b present the first buckling mode for load cases 1 and 2, respectively. For both load cases the reserve factors are above 1. The reserve for load case 1 is 1.126 and is thus large enough. The reserve factor 1.001 for load case 2 yields a small margin. However, load case 2 occurs due to failed brackets and for such failure load cases onset of buckling below limit load might be acceptable (when certain conditions, which shall not be discussed here, are fulfilled). For the analysis of the stresses the same procedure as for the sandwich structure is applied. In Tab. 5 the maximum occurring in-plane ply stresses are summarized. For those, the reserve factors for first ply failure according to the maximum stress criteria in (1) are calculated and provided in Tab. 6. As one can see all reserve factors are greater than 1.



a) Load Case 1 b) Load Case 2

Figure 9: Absolute displacement plot of the torque box structure for a) load case 1 and b) load case 2 in $[\mathrm{mm}]$, scale factor: 5.



a) Load Case 1 b) Load Case 2

Figure 10: Buckling shapes of the torque box structure for a) load case 1 and b) load case 2, scale factor: 100.

4 DISCUSSION OF RESULTS

Tables 4 to 7 compare the properties of the presented designs. An overview about the global stiffness behavior is given in Tab. 4 on behalf of the three characteristic nodes selected in Fig. 3. The absolute node displacements of the torque box structure are only larger at node 3 for load case 1 and at node 1 for load case 2. However, it seems reasonable to envelope both load cases and consider the maximum displacement of each node only. Therefore, the torque box structure is clearly the better design. The differences in the displacements between the torque box design and the sandwich structure are considerable large. As mentioned in section 3.1, the global stiffness of the zig-zag structure is, except at node 3, approximately in the same range as the sandwich structure, making itself a competitive design in terms of the global stiffness. With respect to the occurring stresses and the resulting reserve factors, the values from the sandwich design and the torque box design are provided only. The stresses in the zig-zag structure are not further discussed here because of the unacceptable large deflections. Table 5 presents the maximum occurring in-plane ply stresses for load cases 1 and 2, respectively. The stress values are approximately the same for load case 1, whereas the stress values of the torque box structure for load case 2 are less than half of those from the sandwich structure. This result is remarkable.

	Node 1	Node 2	Node 3
Load Case 1			
Sandwich Structure	10.41	17.84	1.52
Zig-Zag Structure	10.57	18.59	24.39
Torque Box Structure	8.75	15.85	2.37
Load Case 2			
Sandwich Structure	1.00	17.94	17.47
Zig-Zag Structure	6.56	13.78	28.21
Torque Box Structure	5.50	11.65	5.45

Table 4: Absolute displacement of the alternative designs at the three characteristic nodes, [mm].

	σ_{11}^t	σ_{11}^c	σ_{22}^t	σ_{22}^c	σ_{12}
Load Case 1					
Sandwich Structure	492.5	519.9	38.7	40.4	25.3
Torque Box Structure	546.4	518.4	39.7	40.0	23.6
Load Case 2					
Sandwich Structure	795.7	821.7	34.0	35.1	32.2
Torque Box Structure	362.4	259.9	20.3	14.4	12.6

Table 5: Maximum in plane ply stresses for the alternative designs for load cases 1 and 2 in [MPa].

A conclusion with respect to the strength of the sandwich and torque box structure is made in Tab. 6 for load cases 1 and 2. Here, the resulting reserve factors in accordance with the maximum stress requirement given in (1) are provided. Finally, the structural mass of each design is calculated. The results and their details are stated in Tab. 7. As one can see, the mass of the torque box structure is 520g less than the mass of the sandwich structure. For the zig-zag structure a mass reserve of 1.01kg to the sandwich structure exists, which could be used for structural improvements.

5 CONCLUSION

In this work a path towards the substitution of a honeycomb core in a sandwich structure by a structure entirely made from carbon plies was presented. Two major external (i.e. not geometrical) requirements on the structure were considered and finally realized. These were on the one hand, the integration of a lower skin for aerodynamic reasons and on the other hand, the manufacturing in a single layup and curing process (with exception of an assembly step for a pre-produced C-spar afterwards).

As expected, the critical design requirement moved from material strength to structural stability. The honeycomb core, which supports the lower and upper skin as a distributed foundation

	RF_{11}^t	RF_{11}^c	RF_{22}^t	RF_{22}^c	RF_{12}
Load Case 1					
Sandwich Structure	5.3	2.8	1.4	7.1	4.2
Torque Box Structure	4.8	2.8	1.4	7.1	4.4
Load Case 2					
Sandwich Structure	3.3	1.8	1.6	8.1	3.3
Torque Box Structure	7.2	5.6	2.7	19.8	8.3

Table 6: Minimum reserve factor for the alternative designs for load cases 1 and 2.

	Carbon	Core	Adhesive	Total
Sandwich Structure	4.64	5.53	0.94	11.11
Zig-Zag Structure	10.10	-	-	10.10
Torque Box Structure	10.52	-	0,07	10.59

Table 7: Mass calculation of the design alternatives, [kg].

was replaced by local skin supports in the form of applied ribs and a C-spar. Consequently, the skins became critical for buckling in the sections between the supports. In addition, the need to close the structure were justified, because this measure increased the buckling load factor and reduced the local skin section deformations essentially. The mass equivalent (or even better, the reduced mass) substitution of the sandwich structure was in principal realized after introduction of closing the geometry to act as a torque box.

The presented designs are introduced on basic analytical considerations in a step-by-step approach. First, the global layout was defined by uniformly distributed ribs. Then, the ply numbers, orientations and layup sequences were adjusted according to iteratively calculated results of deformation and buckling analyses. Finally a stress analysis was performed to assure the compliance with the material allowables. However, the possibility of irregular distributed stiffeners together with a locally optimized laminate thickness was not addressed in this article and is part of future investigations.

REFERENCES

[1] M. Meindlhumer, M. Schagerl and M. Fleischmann, *Towards a monolithic design of large aircraft wing spoilers using numerical topology and laminate optimization*. In: Rodrigues et al. (Eds.): Engineering Optimization IV, 651-656, Taylor and Francis Group, London, 2015.

[2] M. C. Y. Niu, *Airframe structural design - Practical design information and data on aircraft structures, 2nd Edition*. Hong Kong Comilit Press Ltd., 2002.

[3] CS-25 Amendment 3, *Certification specifications for large aeroplanes*. EASA, 2007.

[4] E. J. Barbero, *Introduction to composite materials design, 2nd Edition*. Taylor and Francis Group LLC, 2011.

ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

ROBUST INTEGRAL COMPOSITE AIRCRAFT STRUCTURES

Tamas Havar[1]

[1] Airbus Group Innovations
Willi Messerschmitt Str. 1, 82024 Taufkirchen
e-mail: tamas.havar@airbus.com

Keywords: Composite, Design, Sizing Aircraft, Load Introduction, Robust.

Abstract. *In several research projects highly integrated composite aircraft structures have been design, with a strong emphasis on low cost manufacturing. At the same time the robustness of the part had to be increased in order to lower direct maintenance costs for the end customers. Several integral aircraft part designs are shown, which allow the usage of low cost composite manufacturing procedures, but at the same time guarantee sufficient performance against the occurring loads and external damages. The focus is on the integral composite load introductions that withstand high loadings, even when small cracks have occurred due to external damages. The cracks can be stopped in areas with smaller internal strains or by specific ply drops and junctions. This has been shown by extensive analysis of the critical areas as well as the whole part and proven by coupon, subcomponent and component tests. This allows a light weight but also robust design in combination with low manufacturing and assembly.*

1 INTRODUCTION

Current Aircraft Structures are being made more and more of composite materials. These composite materials are very light combined with high strength and have very good fatigue characteristics. On the contrary their damage behavior is critical leading to low damage allowable for the material. This limits the weight potential and leads to additional higher manufacturing costs by adding toughening interleave materials are thicker structures. The aim is to reduce manufacturing and assembly costs for future design by using out of autoclave processes and a higher integration of subparts in the large structure. For these highly integrated part designs even with included load introduction, the phenomena of crack propagation becomes even more critical, since a small crack can propagate through the whole structure leading to an early failure of the complete structure. Therefore new designs of integral load introduction in combination with sizing methods were developed to provide low cost, low weight but robust structures for future aircrafts.

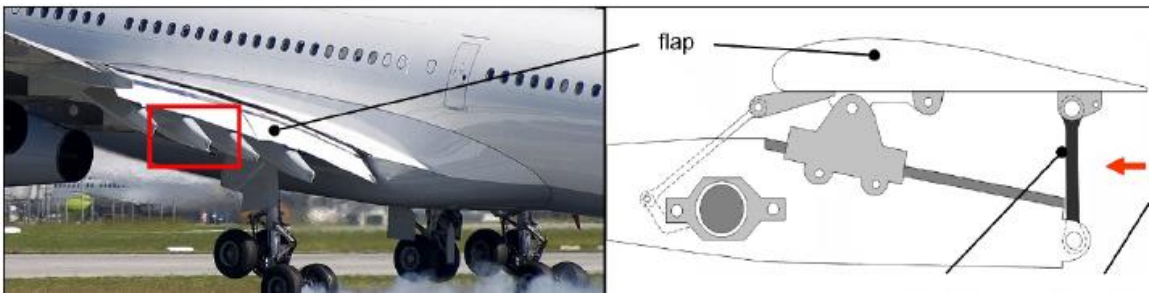


Figure 1: High Lift System of Commercial Aircraft

2 MODULAR COMPOSITE LOAD INTRODUCTION

In cooperation between different AIRBUS divisions, a new advanced composite load introduction rib for high lift devices of future long range aircraft has been developed to minimize weight and manufacturing costs.

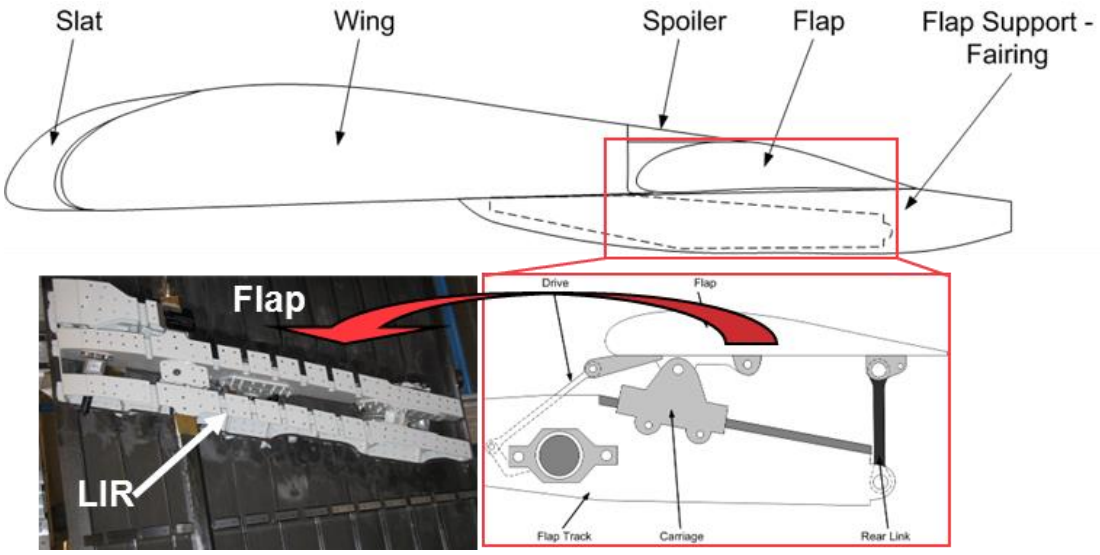


Figure 2: Typical wing cross section with flap and load introduction rib

The focus of the new integrated design of the load introduction rib is on the reduction of the complexity and to simplify the preform manufacturing and therefore the cost. The complex loading of the load introduction structure requires a detailed numeric analysis for an accurate calculation of all critical stresses. The static analysis of the new composite load introduction rib and drive fittings show sufficient strength. Based on prior investigations, the composite lugs show satisfying damage tolerance behavior leading to the conclusion that a second load path is not necessary.

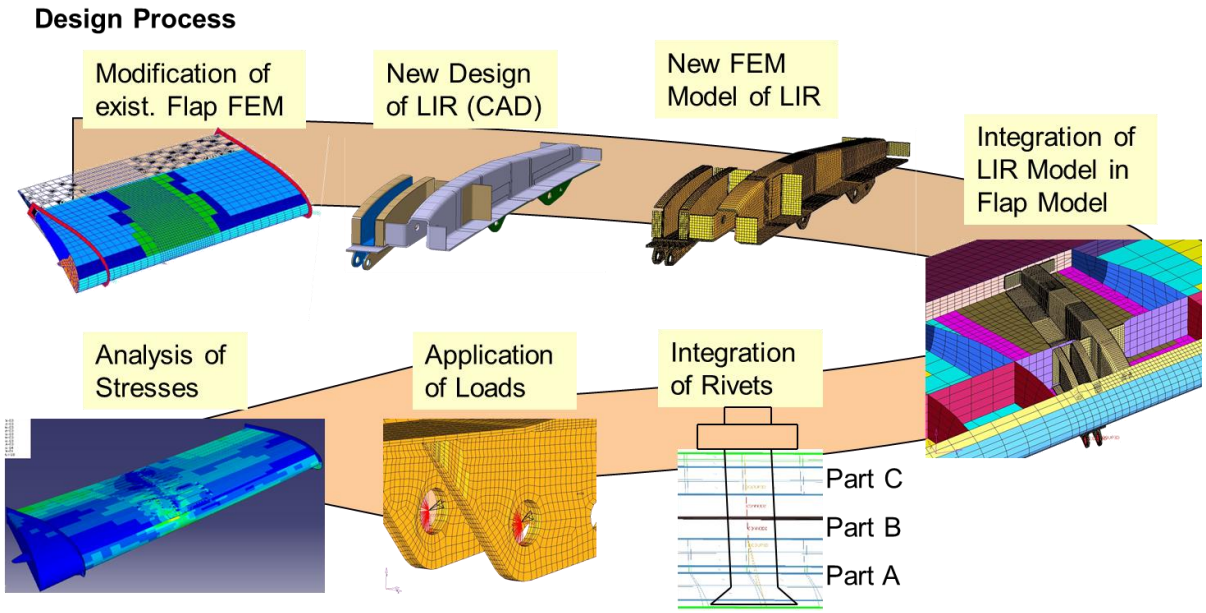


Figure 3: Design process for new composite load introduction rib

The design was verified by subcomponent and component manufacturing by Airbus Helicopters. Knowledge from the prototyping went back into the design to improve the design and to verify to low cost approach. Afterwards the load introduction rib was assembled into an existing high lift device and validated in a full scale component test.



Figure 4: Design process for new composite load introduction rib

3 GENERAL LOAD INTRODUCTION PROBLEM

The partly integral high lift load introduction has shown that current metallic brackets and fittings used for introducing external forces into large structures can be done with composite design. For future aircraft a more integral design is favorable reducing assembly time and effort. This is especially critical for high production rates for singe aisle aircrafts.

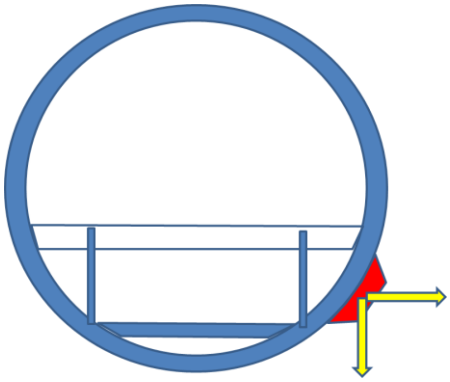


Figure 5: Complete Integral load introductions at aircraft fuselage

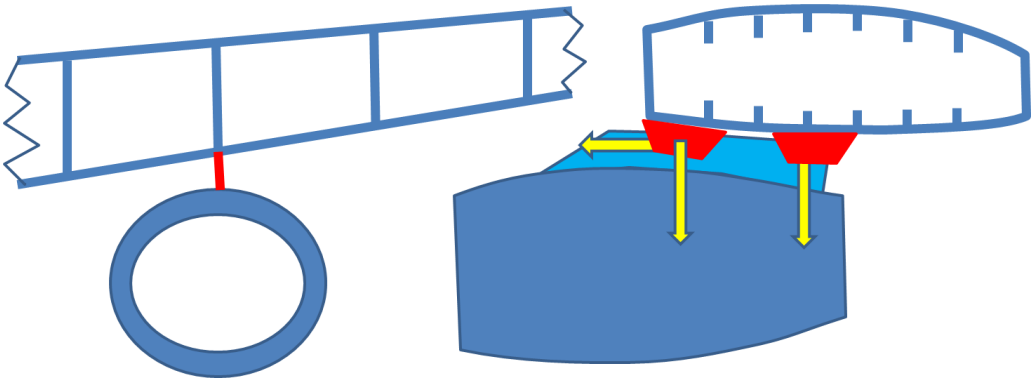


Figure 6: Complete Integral load introductions at aircraft wing/engine junction

For composite structures this is even more difficult than for metallic parts. New integral load introduction designs for composite structures were developed. Usually high bypass loads in the skins/panels have be transferred, while out of plane loads are need to be introduced, best done into frames and stiffeners on the inside.

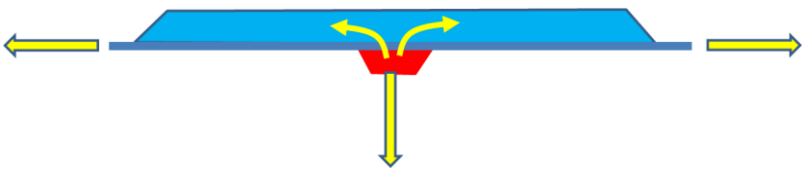


Figure 7: Panel/skin junction with external load bracket

This is also referred as X cross section problem at composite structures. Some composite designs are on the market (3D weaves, Crucify Forms…), but either with very limited performance or difficult manufacturing processes.

4 INTEGRATED LOAD INTRODUCTION

4.1 Composite Cross Section Design

Several composite cross section designs have been developed. In order to mount external load introduction, so called fins were developed. In general a cut in the lower skin has be made in order to penetrate with the fin, which is positioned interleave between the frames/stiffeners on the inside. Just to mount the fin to the skin/panel would lead to very high our of plane loads on the skin, leading to poor performance. This design allows transferring very high loads through the skin into the backing stiffening structures.

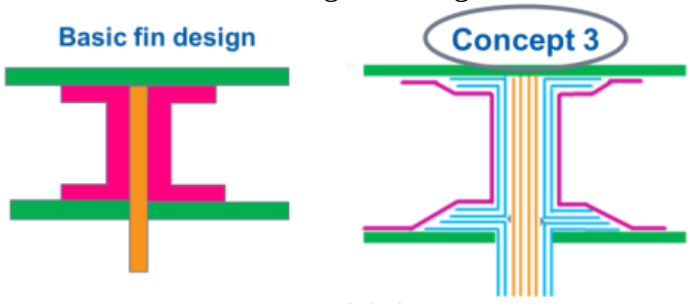


Figure 8: Basic cross section (fin) design

The different concepts were roughly sized with FEA, but only 1 showed high performance and good damage tolerant capability. A high damage tolerance capability is very important to highly integrated structures, hence cracks/delamination caused by external impacts, could grow rapidly leading to early failure of the integral load introduction. Therefore design with no growths or slow growths for damages were chosen

4.2 Detailed FE Analysis

The most promising design shown in Fig. 8, was analysed with non linear effects to determine the maximum load carrying capability. A representative cross section was chosen with a 3 point bending load case. The edges of the skin/panel were fixed, while the fin (load introduction) was loaded perpendicular to the skin with tension and compression. The stiffening back up structure of the skin/panel, which are usually frames, spars and ribs, were also included. These usually carry the out of plane loads of the skin and here the external loads on the fin.

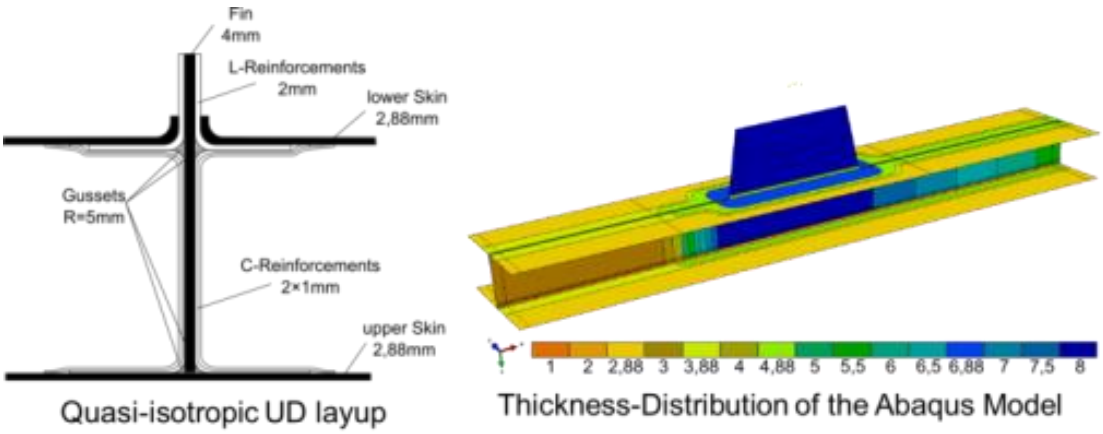


Figure 9: Representative fin analysis with non linear FEA

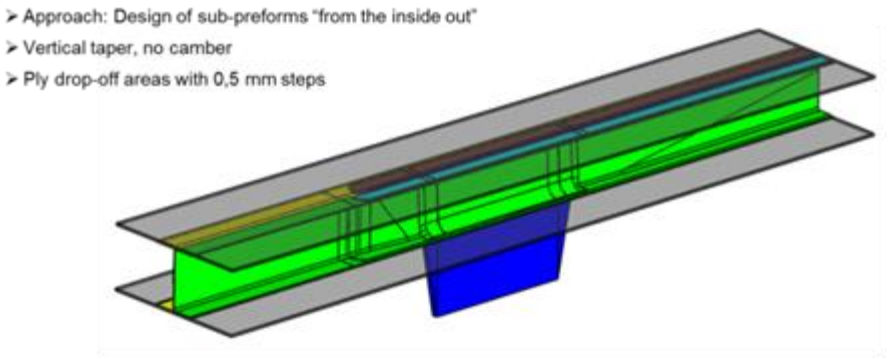


Figure 10: Representative composite skin cross section design with integral fin

At the non linear FE analysis, the transfer loads of the fin to the stiffening elements (transversal interlaminar shear) were of most interest. These are usually tested with Mode I and Mode II test for different materials to determine the interlaminar material allowable. These allowable values were used to determine the reserve factors against the maximum static loads occurring at the cross sections. The FEA has shown that are these stresses are not very critical due to a large connection area between the fin and the backup structures (frame, ribs, spars…).

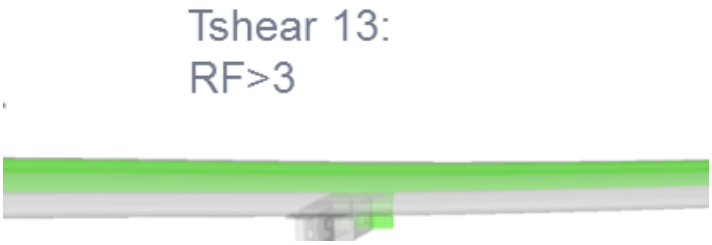


Figure 11: Transversal shear stresses in fin to frame junction

5 VERIFICATION AND VALIDATION

5.1 Verification

In order achieve the high cost reduction goals, out of autoclave manufacturing materials and processes were used as basis. The representative fin design was manufactured CFRP non crimped fabric and an out of autoclave VAP® process.

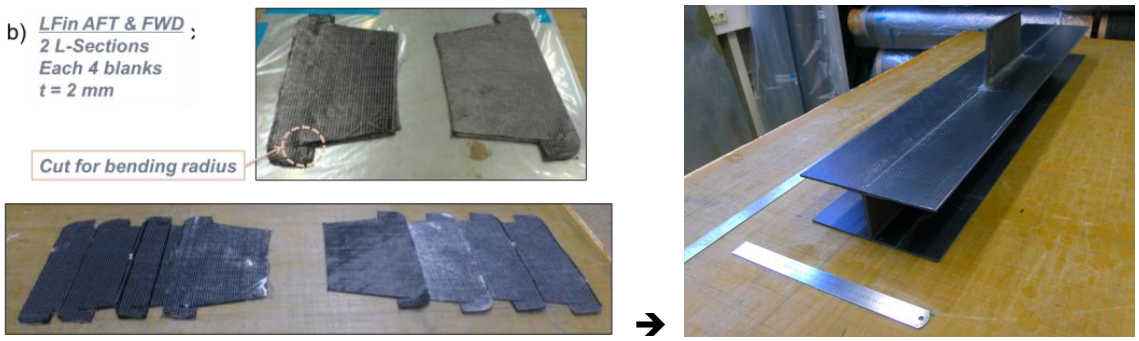


Figure 12: Manufacturing of representative composite skin cross section design with integral fin

Afterwards several prototypes were statically and fatigue tested. During the test, buckling of the skin was visible at the maximum static load, which is usual for very high loaded structures. This was later on reduced with additional stiffeners on the side.



Figure 13: Testing of representative composite skin cross section design with integral fin

In order to visualize the critical fin run out area, where cracks could occure, additional white paint was applied. At maximum static load small cracks did occur at the run out area. This was party caused by manufacturing effects (resin rich areas) and by the high corner loading. This did not lead to failure, though; since the cracks were only at the surface. Fracture did finally occur at the backup stiffener structure were it was not reinforced close to clamping. Here an in plane shear fracture could be seen (Fig. 14). Therefore the fin has proven itself to be very damage tolerant even with cracks at the surface, the fin did not fracture, but rather the skin/frame backup structure outside of the load introduction area.



Figure 14: Fracture of representative composite skin cross section design with integral fin

5.2 Validation

The successful testing of the load introduction elements (fins) was analyzed with nonlinear FEA. Here the buckling phenomena could also be shown. In the post buckling analysis an increased in plane loading of the stiffening backup structure could be shown, which lead to the final fracture above ultimate load. The FEA showed very good conformance with the strain gauges in the test [Fig. 15]. Therefore the design was successfully verified in the test with very good validation to the non linear FEA.

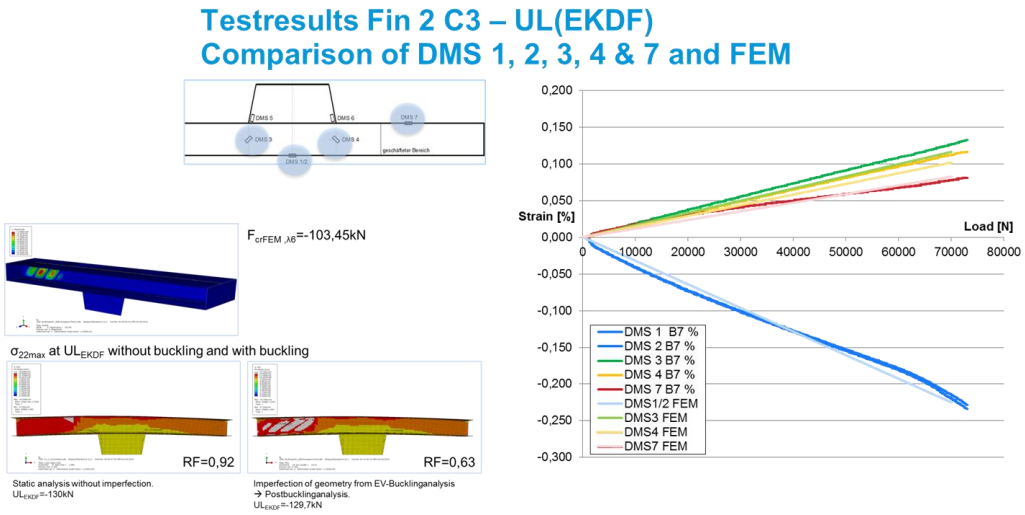


Figure 15: Validation of representative composite skin cross section design with integral fin

6 CONCLUSIONS

This papers shows the design, verification and validation of partly integral and fully integral composite load introduction for future aircraft. Both design have been proven to be very damage tolerant, leading to a robust design and therefore low maintenance costs for the airliners. The manufacturing of the load introduction design has proven the high cost reduction potential of over 20% in regard to current bolted design. A weight reduction of over 10% could also be achieved for the new load introductions.

7 REFERENCES

[1] S. Heimbs, H. Lang, T. Havar, *Central European Journal of Engineering, Vol. 2, 4th Edition*. Springer, 2012.

[2] Havar, T. Werchner C.; Analysis of Innovative Composite Aircraft Structures, ECCOMAS Conference 2012.

ECCOMAS Congress 2016

VII European Congress on Computational Methods in Applied Sciences and Engineering

ECCOMAS Congress 2016

Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering

M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (Eds)

First Edition, September 2016

ISBN: 978-618-82844-0-1



Institute of Structural Analysis and Antiseismic Research
National Technical University of Athens, Greece