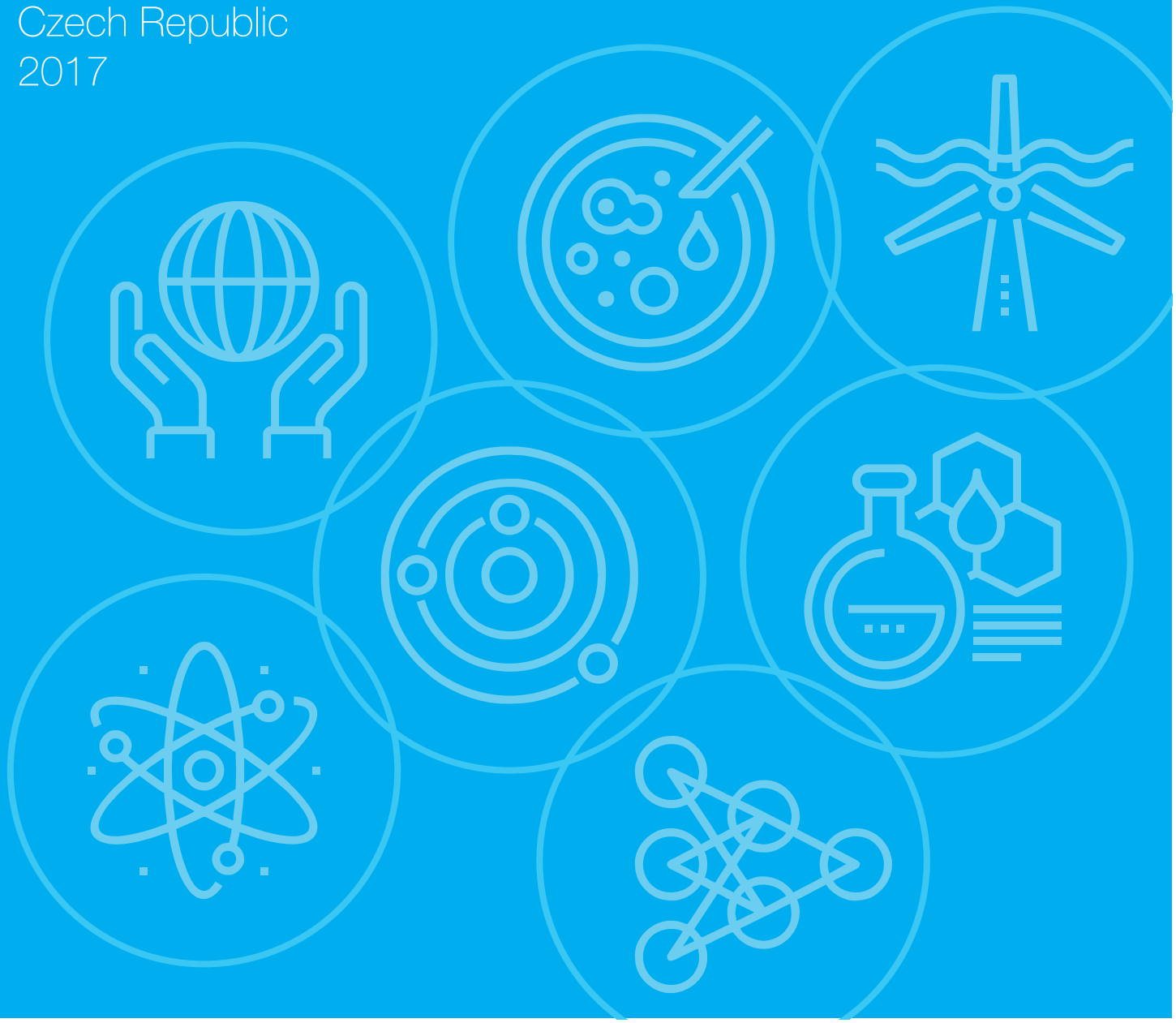
IT4Innovations
national01$#&0
supercomputing
center@#01%101

SUPERCOMPUTING IN SCIENCE AND ENGINEERING

IT4Innovations National Supercomputing Center
Czech Republic
2017



VŠB – Technical University of Ostrava

IT4Innovations
national
supercomputing
center

Supercomputing in Science and Engineering

Editors: Karina Pešatová, Barbora Poláková, John Cawley
Ostrava, 2017
Published by VŠB – Technical University of Ostrava
Design and typesetting: Milan Smetana, Emílie Smetanová
Page count: 224
Edition: 1st
Printed by: Knowlimits, s.r.o.
Copies: 300
Not for sale

ISBN 978-80-248-4037-6

SUPERCOMPUTING IN SCIENCE AND ENGINEERING

IT4Innovations National Supercomputing Center
Czech Republic
2017

Editors: Karina Pešatová, Barbora Poláková, John Cawley
Reviewers: Martin Palkovič, Branislav Jansík, Vít Vondrák, Tomáš Kozubek, Jan Martinovič, René Kalus, Tomáš Brzobohatý

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

This publication was supported by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project „IT4Innovations National Supercomputing Center – LM2015070“

Table of contents

Director's introduction 9
Martin Palkovič

The infrastructure of IT4Innovations 10

01 ASTROPHYSICS

Computational Planetology: Timing of Water Plume Eruptions on Enceladus Explained by Interior Viscosity Structure 16
Marie Běhounková

Star Formation in Gaseous Layers 19
František Dinnbier

Accretion and Jet Launching in Young Stellar Objects 22
Ivana Orlitová

Origin of the Second Stellar Generation in Globular Clusters 24
Richard Wünsch

02 COMPUTATIONAL SCIENCES

Verification of MPI Applications Defined in Kaira 30
Marek Běhálek

Benchmarking of the CFD Codes on Salomon Cluster 33
Tomáš Brzobohatý

Stochastic Estimation of the Diagonal of the Inverse (SEDI) 36
Greg Hernandez

Permon Toolbox Development 40
David Horák

Tsunami Simulation with Parallel Adaptive Mesh Refinement on the Xeon Phi Nodes of Salomon 43
Michael Bader

Parallelisation of the Pragtic Software for an Automated Fatigue Damage Calculation 46
David Horák

Convolutional Neural Networks for Image Restoration 49
Michal Hradiš

Rendering in Blender Cycles Using MPI and Intel® Xeon Phi™ 52
Milan Jaroš

Evaluation of Parallel I/O Strategies for Distributed Sparse Matrices 56
Daniel Langr

Parallel Boundary Element Method for Sound-hard Scattering 59
Michal Merta

Development of a Parallel Algorithm for Fuzzy Association Rules Mining 62
Viktor Pavliska

Task Programming and Asynchronous Communication for Irregular Computations 65
Eric Petit

Sparse Block Matrix Linear Algebra (SBMLA) 68
Lukáš Polok

ESPRESO – Accelerated Solver Library for Intel Xeon Phi Systems 70
Lubomír Říha, Tomáš Brzobohatý

Computational Methods for Solving Stochastic Equilibrium Models 73
Olaf Schenk, Simon Scheidegger

MATSOL – Development, Testing and Support Libraries 77
Oldřich Vlach

BEM4I – Vectorised Boundary Element Library for Many-core Systems 80
Michal Merta

03 EARTH AND ENVIRONMENTAL SCIENCES

HPC Framework for Rainfall-Runoff Uncertainty Modelling 86
Štěpán Kuchař

Urban Canopy Climate Forcing on a Regional Scale Modelled by the RegCM and WRF Models 90
Peter Huszár

Parallelization of Flood Model Calibration 93
Štěpán Kuchař

Hydrological Model Remote Execution and HPC as a Service 97
Jan Martinovič, Štěpán Kuchař, Václav Svatoň, Vít Vondrák

Large-scale Geosimulations Using GEM and Trilinos 100
Jiří Starý

Adjoint Tomography of Surface Waves: Choice of Regularisation Based on Synthetic Tests 104
Ľubica Valentová

Parallel Modeling of Ocean Flows and their Magnetic Signatures 108
Jakub Velímský

04 ENGINEERING AND COMPUTATIONAL FLUID DYNAMICS

Axially and Radially Cooled Brake Disc with Cover 114
Marek Gebauer

Solving Non-Linear Structural Dynamics Problems on a Supercomputer 117
Tomáš Karásek

HPC for Modelling Strongly Nonlinear Processes in Mechanics, Parts I.-III. 119
Pavel Maršálek, Petr Horyl, Tomáš Karásek

CFD Proxy – Strong Scaling in CFD 122
Christian Simmendinger

Scalable Solvers for Subsurface Flow Simulations 125
Jakub Šístek

Multilevel Domain Decomposition Solvers for Incompressible Flows 128
Jakub Šístek

ACLA – Active Load Alleviation 131
Petr Vrchota

05 CHEMISTRY AND MATERIAL SCIENCES

Benchmarking of New Accurate Multi-reference Coupled Cluster Methods 136
Jiří Brabec

Atomic-scale Design of Opticaly Active Nanostructured Amorphous Silica 139
Antonio Cammarata, Luisa Spallino, Paolo Nicolini, Benjamin Irving

Quantum Simulations of Doped Helium Nanodroplets I. Photoabsorption Spectra of Charged Helium Clusters 141
Rajko Ćosić

Molecular Docking on High Performance Computing Clusters 144
Rafael Doležal

The SQM/COSMO Scoring Function Reliably Describes Noncovalent Interactions in Protein-ligand Complexes 147
Pavel Hobza

Theoretical Study of the State-selected Reactions of Electronically Excited Oxygen Ions with Methane Relevant for Titan's Atmosphere 150
Jan Hrušák

Conformational Transitions and Membrane Binding of the Neuronal Calcium Sensor Recoverin 153
Pavel Jungwirth

Modelling of Elementary Processes in Cold Rare-gas Plasmas 156
René Kalus

Enzymatic Reaction Mechanisms Investigated by Car-Parrinello ab Initio Dynamics 159
Jaroslav Koča

Detecting Materials in Coal-polyurethane Geocomposite CT-scans 162
Marek Pecha, Martin Čermák, David Horák

Periodic DFT Studies of Zeolite Based Catalysts I. 165
Štěpán Sklenák

Periodic DFT Studies of Zeolite Based Catalysts II. 168
Štěpán Sklenák

Periodic DFT Studies of Zeolite Based Catalysts III. 171
Štěpán Sklenák

Quantum-mechanical Simulations of Interfaces in Advanced Materials 174
Mojmír Šob

Exploring the Effect of Loading Conditions on the Ideal Tensile Strength 177
Mojmír Šob

Hydrophobic Impurities in Water Clusters 180
Aleš Vítek

06 LIFE SCIENCES AND BIOLOGY

Estimation of Empirical Distribution Function for Testing of Modality of 3D Particle Sizes Based on 2D Observations 186
Michal Burda

A non-Newtonian Model of Blood Capturing Segregation of Erythrocytes – First Step to Proper Modelling of the Blood Coagulation Process 189
Marek Čapek

FEM Model of the Soft Tissues – Comparison with Ultrasound Elastography 192
Jan Grepl

Convergence Testing of a K-space Pseudospectral Scheme for Transcranial Time-reversal Focusing 195
Jiří Jaroš, Bradley E. Treeby

Structural Investigations of the Human Mitochondrial Lon Protease 198
Lubomír Kováčik

Enhancement of the IT4IBlender Tool for Parallel Image Segmentation of Computer Tomography and Magnetic Resonance Images 202
Tomáš Karásek

Modelling Beam Distortion During Focused Ultrasound Surgery in the Prostate Using k-Wave 205
Jiří Jaroš

07 PHYSICS

Quantum Collisions of Electrons with Hydrogen Atoms 210
Jakub Benda, Karel Houfek

TEBDOL – Parallel Simulation of Ultracold Atoms in Optical Lattices 213
Miroslav Urbánek, Pavel Soldán

One-dimensional and Two-dimensional Magnetic Systems 216
Dominik Legut

Comprehensive Modelling Long Time Evolution of Suddenly Ionised Rare Gas Clusters Through Multiscale Model and Full Dynamics Calculations 218
Ivan Janeček

Dear readers,

we proudly present to you this selection of bleeding edge scientific and engineering projects that have been made possible through the use of the large research e-infrastructure at IT4Innovations National Supercomputing Center. I strongly believe our supercomputing facility is a crucial asset, helping to improve the excellence of Czech science.

The IT4Innovations National Supercomputing Center was established at the end of 2011 with the clear goal to implement and operate large research e-infrastructure in High Performance Computing (HPC) in the Czech Republic. Our first cluster, Anselm, became operational in May 2013. Anselm was the first, and for two years the only large High Performance Computing system publicly available for Czech scientists and researchers nationwide complying with the principles of open access. In September 2015 the supercomputer Salomon was inaugurated and became available. At the time of its completion it was the 14th fastest supercomputer in Europe, and rated 40th in the world, according to the TOP500 list. The computational resources of the e-infrastructure are allocated via open access competitions that are announced three times a year.

IT4Innovations National Supercomputing Center is not just a computational cycle provider, we also offer high quality HPC related expertise and research, world class training and unique MSc. and Ph.D. study programmes in the HPC domain. Our membership in the pan-European infrastructure PRACE allows Czech scientists access to the largest HPC systems in Europe, should even larger scale systems than nationally available be required for particular research tasks. Our membership in ETP4HPC allows us to influence the HPC Strategic Research Agenda, the document that outlines a roadmap for the achievement of exa-scale capabilities by the European High Performance Computing ecosystem. So that we may better serve our users, we continually strive to improve all aspects of the services we offer.

I hope that we maintain your goodwill towards the IT4Innovations National Supercomputing Center, and that our e-infrastructure will continue to help you make breakthroughs in your research in the future.

Yours faithfully,

Martin Palkovič Digitally signed by Martin Palkovič
Date: 2017.04.26 11:13:45 +02'00'

Martin Palkovič
Managing director

THE INFRASTRUCTURE OF IT4INNOVATIONS

IT4Innovations operates the largest supercomputing facility in the Czech Republic, currently two X86 clusters named Salomon and Anselm. The supercomputer infrastructure was installed in several stages. The first cluster, Anselm, was installed in temporary mobile units in May 2013. The Salomon supercomputer became operational in September 2015 and is one of the 100 most powerful supercomputers in the world. Both clusters are now permanently located in the new IT4Innovations building, which has a 500 m^2 server room.

The construction of the IT4Innovations building started in December 2012, the administrative section being completed in March 2014. The physical infrastructure supporting the supercomputers is fully redundant in power and cooling, using direct hot water cooling (2 x 600 kW), cold water cooling (200, 400 and 600 kW) and heat recovery (up to 380 kW) for building heating to lower the operational costs. The backup power is provided through diesel rotary UPSes, each with a capacity of 2.5 MVA. The Internet connection is 2x40 Gb/s connected to two different endpoints of CESNET network.

The Anselm Cluster

The Anselm cluster consists of 209 computational nodes. Each node is equipped with 16 cores (two eight-core Intel Sandy Bridge processors). There are four types of compute nodes: 180 regular compute nodes without any accelerator, with 2.4 GHz CPUs and 64 GB of RAM; 23 compute nodes with GPU accelerators (NVIDIA Tesla Kepler K20), with 2.3 GHz CPUs and 96 GB of RAM; 4 compute nodes with MIC accelerators (Intel Xeon Phi 5110P), with 2.3 GHz CPUs and 96 GB of RAM; and 2 fat nodes with larger RAM (2.4 GHz CPUs, 512 GB RAM). InfiniBand (QDR) and Ethernet networks interconnect the nodes.

The theoretical peak performance of the Anselm cluster is 94 TFLOP/s with a maximal aggregated LINPACK performance of 73 TFLOP/s. All computing nodes share 320 TiB /home disk storage to store user files. The 146 TiB shared /scratch storage is available for scratch data. These file systems are provided by Lustre parallel file system. There is also a local hard drive (330 GiB available) on all compute nodes.

The Salomon Cluster

The Salomon cluster consists of 1008 computational nodes. Each node is equipped with 24 cores (two twelve-core Intel Haswell processors). All compute and login nodes are interconnected through a 7D Enhanced hypercube Infiniband network (see Figure 1). There are two types of compute nodes: 576 compute nodes without any accelerator and 432 compute nodes with MIC accelerators (two Intel Xeon Phi 7120P per node). Each node is equipped with two 2.5 GHz processors and 128 GB of RAM.

The theoretical peak performance of Salomon exceeds 2 PFLOP/s with an aggregated LINPACK performance of over 1.5 PFLOP/s earning the system 67th place in the TOP500 list of most powerful supercomputers (November 2016). All computing nodes share 500 TiB /home disk storage to store user files. The 1.6 PiB shared storage is available for scratch and project data. In addition to the cluster nodes an SMP/NUMA machine is incorporated into the system with 112 CPU cores and 3.25 TB of RAM.

The operating system on Anselm is Redhat Linux and on the supercomputer Salomon it is CentOS. State of the art programs and tools to develop, profile and debug HPC codes, such as Intel Parallel Studio XE Cluster Edition, Allinea Forge and RougeWave Totalview are also available. The clusters provide several implementations of the MPI library. A large amount of application software, both open source and from ISVs is available, including numerical libraries and software packages from many domains of science.

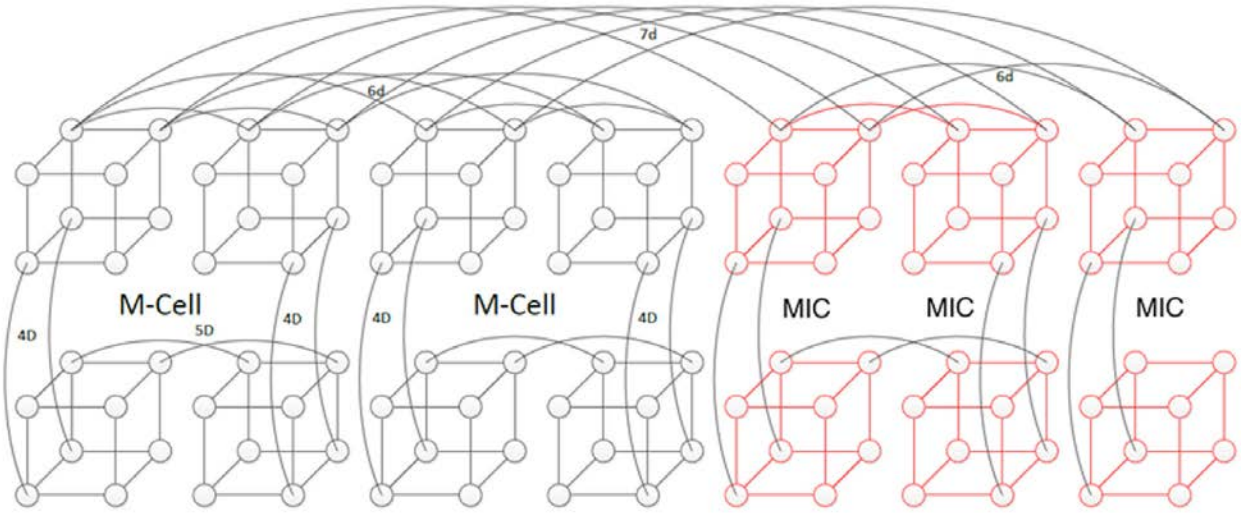


Figure 1.
A simplified view of the 7D Enhanced Hypercube topology of the Salomon supercomputer.
M-Cell denotes nodes without acceleration, MIC denotes the accelerated nodes interconnection.

	ANSELM CLUSTER	SALOMON CLUSTER
Production	Spring 2013	Summer 2015
		67th at TOP500 world supercomputers
Peak performance	94 Tflop/s	2011 Tflop/s
Operating system	RedHat Linux 64bit 6.x	RedHat Linux 64bit 6.x
		CentOS 64bit 6.x
Nodes	209	1008
CPU	2x Intel SandyBridge 8 core 2.3/2.4 GHz	2x Intel Haswell 12 core 2.5 GHz
	3344 cores in total	24192 cores in total
RAM	64 GB/96 GB/512 GB	128 GB/3.25 TB (uv node)
GPU accel.	23x NVidia Tesla Kepler K20	N/A
MIC accel.	4x Intel Xeon Phi 5110P	864x Intel Xeon Phi 7120P
		61 cores each
		52704 core in total
Storage	320 TiB /home (2 GB/s speed)	500 TiB /home (6 GB/s speed)
	146 TiB /scratch (6 GB/s speed)	1638 TiB /scratch (30 GB/s speed)
Interconnection	Infiniband QDR 40 Gb/s	Infiniband FDR 56 Gb/s

Table 1.
Technical data

Acknowledgements

The IT4Innovations National Supercomputing Center was established mainly thanks to European funds, namely the Operational Programme Research and Development for Innovation, priority axis 1 – European Centers of Excellence fund, which is managed by the Czech Ministry of Education, Youth and Sports in the Czech Republic. We gratefully acknowledge the continued support of the Ministry of Education, Youth and Sports and their project, Large infrastructures for Research, Experimental Development and Innovation, for financing the operations of IT4Innovations.

01 | ASTROPHYSICS

COMPUTATIONAL PLANETOLOGY: TIMING OF WATER PLUME ERUPTIONS ON ENCELADUS EXPLAINED BY INTERIOR VISCOSITY STRUCTURE

Research institution:
Charles University in Prague

Principal investigator:
Marie Běhounková

Researchers:
Klára Kalousová, Michaela Walterová

Project IDs:
OPEN-3-13, OPEN-4-16

Introduction

Numerical simulation is an important tool used for interpreting observations of moons and planets. The planetary processes are on broad time scales going from a few days up to hundreds of million years. As a part of the project we have concentrated on both long-timescale (including sub-solidus, two-phase convection) and short-timescale processes (tidal deformation). We have particularly focused on the tidal deformation in the ice shell of Saturn's moon Enceladus.

Enceladus is one of the most enigmatic bodies in the Solar System and it is intensively studied by mission Cassini. Enceladus is a small body and therefore it should be rather cold and inactive. Yet, it contains a deep internal ocean and it is one of the most active bodies in the Solar system due to its ongoing geysering activity connected to four prominent faults in the south polar region. Enceladus is orbiting Saturn on a synchronous and slightly eccentric orbit. It is therefore (periodically) deformed by tides during one orbital period. The deformation of the ice shell is sensitive to the mechanical properties and its internal structure. Measurements of the deformation are therefore a possible key to Enceladus' internal structure. The magnitude of the deformation is, however, unknown. Nevertheless, by measuring the geysering activity, it has shown that its timing is influenced by the stress regime along the stripes [5,7]. The observed activity is, however, somewhat delayed compared to the case with a purely elastic shell and a global ocean. The observed delay may thus provide a constraint on the internal structure.

Results and Methods

We therefore tested a hypothesis that the tensile tidal stress along the tiger stripes is related to the observed activity (brightness) of the geysers. We first computed the viscoelastic deformation as a function of time/true anomaly (i.e. orbital position) in the entire shell. In order to evaluate the stress and deformation due to tides, we consider small deformations in a hydrostatically pre-stressed spherical incompressible viscoelastic body. The time evolution of stress and displacement in such a body is governed by the mass and momentum conservation equations. The spatial discretisation employs spherical harmonic function in horizontal directions and finite difference scheme in the vertical direction [11]. We integrated the set of equations in the time domain until we reached a converge solution [1]. This approach allows the introduction of the lateral variation of viscosity as well as the combination of boundary conditions on the bottom boundary. We employed the Andrade rheology as it is designed to describe the viscoelastic response for periods corresponding to tidal forcing [4,2]. We then evaluated the time evolution of normal stress along four prominent faults (see Figure 1).

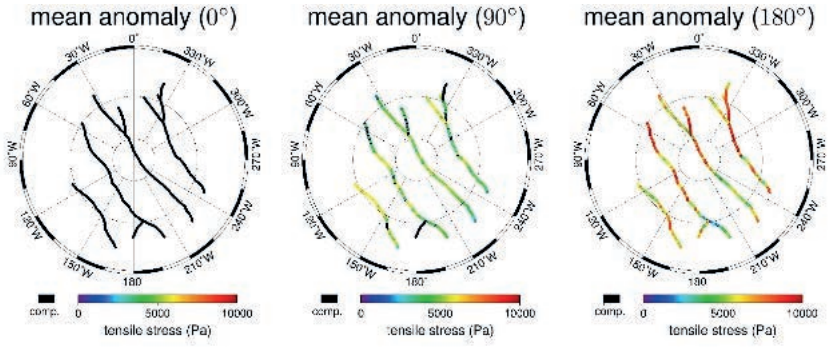


Figure 1.
Evolution of tensile stress along the orbit, periapse (left), mean anomaly 90° (centre) and apoapse (right), compression (black) – low activity, tension (color) – high activity.

We demonstrated that the tidal deformation and the delay of the predicted activity is primarily influenced by the ice shell thickness L, the extent of a possibly localised deep ocean with an angular width D varying between 0° (no ocean) and 360° (global ocean), the low viscosity zone mimicking a convective region above the internal ocean with viscosity η and the thickness of the lithosphere L. Through a systematic search of the parametric space, we predicted theoretical curves of eruption activity for more than 1,000 interior models (see sensitivity to parameters in Figure 2).

The results show that a thin lithosphere (< 5 km) and a warm ice mantle having a minimal viscosity between 1 and $5{\cdot}10^{13}$ Pa.s in the region above the ocean are the main requirements to explain the observed activity. Successful models with a global ocean can be found, but only for an ice shell thicker than 50 km. For ice shell thicknesses L < 50 km, successful models can only be found for regional seas with angular widths between 90° and 110°. The energy source for the observed activity is highly probable tidal heating. The observed heat loss

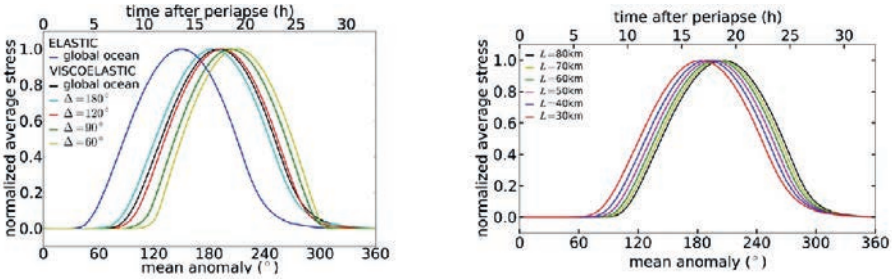


Figure 2.
Theoretical (predicted) activity, sensitivity of theoretical curves to viscoelastic properties and extent of internal ocean (left) and ice shell thickness (right).

We then computed the misfit between the theoretical curves and two available datasets [5,7]. As a first step in the best-fit model selection, we defined a model as statistically acceptable if it is better than a model with no time dependence on a confidence level of 95% using F-test [7]. We then classified the acceptable models according to their misfit values as ranking within the 5%-, 10%- or 20%-best models (Figure 3).

is estimated to be up to 20 GW through the south polar region [6]. We therefore predicted tidal heating in the ice shell for all investigated models. We found that tidal heating strongly varies with all studied parameters (Figure 4). The global tidal heating does not exceed 5 GW for models predicting the geysering activity. Nevertheless, the maximum heating for our models can be as high as several tens of gigawatt, but only for cases with

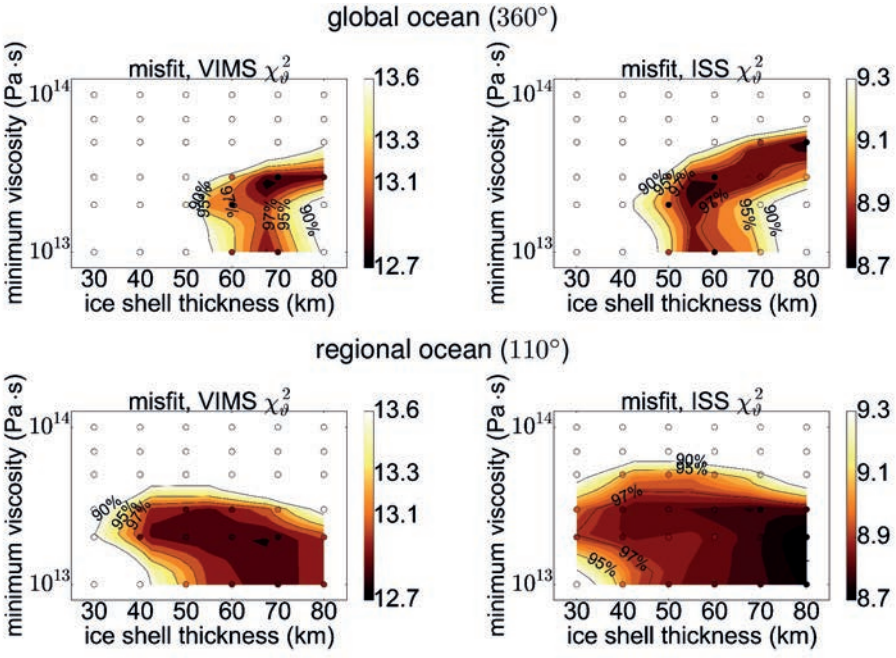


Figure 3.
Misfit between theoretical curve and data, left: VIMS [5] and right ISS [7] data sets, top global ocean, bottom regional ocean

extremely low minimum viscosity η = 10^{13} Pa.s and an ocean width larger than 180°. We conclude that the current tidal heating in the ice shell is neither in equilibrium or the only source of the energy.

On-going Research / Outlook

For the numerical model used here, the stress and deformation is not influenced by the faults. Therefore, the next step is the incorporation of the faults. Additionally, new data based on the libration [10] show that the ice shell thickness is possibly lower than expected and predicted by our model. In the future, we will focus on the inconsistency between the libration measurement and timing of plume.

Conclusion

The observed delay in eruption activity hints at the presence of only a small localised internal ocean (Δ < 120°) and a thin ice shell or global internal ocean and thick ice shell. We further show that the observed activity is consistent with a thin lithosphere over a warm ice mantle in the south polar region. However, tidal heating for such models is low and cannot explain the present day heat loss in Enceladus's southern polar region. Enceladus is therefore neither in equilibrium [8,3] or are its deep heat sources of importance [9].

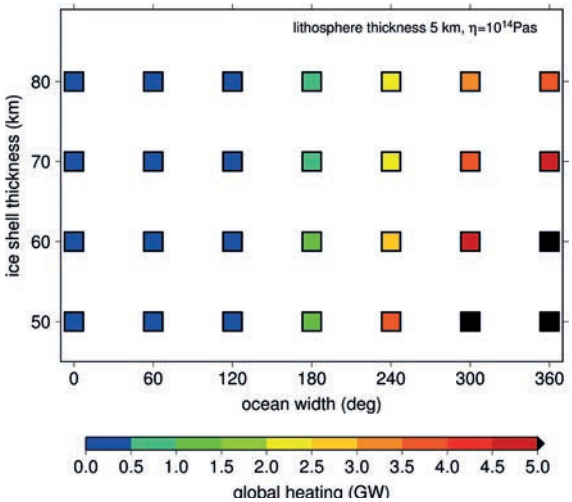


Figure 4.
Global tidal heating as a function of ocean width and ice shell thickness, minimum viscosity 10^{14} Pa.s.

References

[1] Běhounková M.; G. Tobie, G. Choblet, and O. Čadek. Coupling mantle convection and tidal dissipation: Applications to Enceladus and Earth-like planets. J. Geophys. Res. 2010, 115:9011–. Doi: 10.1029/2009JE003564

[2] Běhounková M.; Tobie, G., Choblet, G., Čadek, O., Porco, C., Nimmo, F. Timing of water plume eruptions on Enceladus explained by interior viscosity structure, Nature Geoscience 8. 2015, 601-604. Doi: 10.1038/ngeo2475

[3] Běhounková M.; G. Tobie, G. Choblet, and O.Čadek. Tidally Induced Thermal Runaways on Extrasolar Earths: Impact on Habitability. ApJ. 2011, 728:89–. Doi: 10.1088/0004-637X/728/2/89

[4] Castillo-Rogez J.C.; M. Efroimsky, and V. Lainey. The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. J. Geophys. Res. 2011, 116(E9). Doi: 10.1029/2010JE003664

[5] Hedman M. M.; C. M. Gosmeyer, P. D. Nicholson, C. Sotin, R. H. Brown, R. N. Clark, K. H. Baines, B. J. Buratti, and M. R. Showalter. An observed correlation between plume activity and tidal stresses on Enceladus. Nature. August 2013. 500:182-184. ISSN 0028-0836

[6] Howett C. J. A.; J. R. Spencer, J. Pearl, and M. Segura. High heat flow from Enceladus' south polar region measured using 10-600 cm^{-1} Cassini/CIRS data. J. Geophys. Res. 2011, 116(E3). Doi: 10.1029/2010JE003718

[7] Nimmo F.; C. Porco, and C. Mitchell. Tidally Modulated Eruptions on Enceladus: Cassini ISS Observations and Models. AJ. 2014, 148(3):46. Doi: 10.1088/0004-6256/148/3/46

[8] O'Neill C. and F. Nimmo. The role of episodic overturn in generating the surface geology and heat flow on Enceladus. Nature Geoscience. 2010, 3(2): 88-91. Doi: 10.1038/ngeo731

[9] Roberts J. H. The fluffy core of Enceladus. Icarus. 2015, 258: 54-66. Doi: 10.1016/j.icarus.2015.05.033

[10] Thomas P. C.; R. Tajeddine, M. S. Tiscareno, J. A. Burns, J. Joseph, T. J. Loredo, P. Helfenstein, and C. Porco. Enceladus's measured physical libration requires a global subsurface ocean. Icarus. 2016, 264: 37-47. Doi: 10.1016/j.icarus.2015.08.037

[11] Tobie G.; O. Čadek, and C. Sotin. Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus. Icarus. 2008,196(2): 642-652. Doi: 10.1016/j.icarus.2008.03.008

Publications

[1] Běhounková M.; Tobie, G., Choblet, G., Čadek, O., Porco, C., Nimmo, F. Timing of water plume eruptions on Enceladus explained by interior viscosity structure, Nature Geoscience 8. 2015, 601-604. Doi:10.1038/ngeo2475

[2] Běhounková M.; G. Tobie, G. Choblet, and O. Čadek. Tidal stress and jets activity on Enceladus, European Planetary Science Congress 2014, Portugal, EPSC2014-406, oral presentation.

[3] Běhounková M. and O. Čadek. Tidal dissipation in heterogeneous bodies: Maxwell vs Andrade rheology, European Planetary Science Congress 2014, Portugal, EPSC2014-409, poster.

STAR FORMATION IN GASEOUS LAYERS

Research institution:
Astronomical Institute of the Czech Academy of Sciences

Principal investigator:
František Dinnbier

Researcher:
Richard Wünsch

Project ID:
OPEN-4-3

Introduction

Star formation occurs in the coldest and densest parts of giant molecular clouds. In current Galaxy, stars usually form in clusters, which typically contain from less than 100 to more than tens of thousands stars. Newly born stars cover a wide range of masses; from 0.08 solar mass (Msun) to approximately 130 Msun. Massive stars (stars more massive than 8 Msun) are of particular interest mainly for the following reasons:

(i) by releasing their gas, which is processed by thermonuclear burning, they enrich the interstellar medium with heavier elements, changing the chemical content of the Galaxy.

(ii) they have a strong feedback on surrounding interstellar medium (ionising radiation, stellar winds and supernovae). The feedback presumably influences future star formation in the area. Feedback may be either negative (i.e. the medium is heated up and dispersed) or positive (the medium is swept up into a shell, which cools down and forms a new generation of stars).

We focus on issue (ii) with the possibility of positive feedback. In this case, the crucial step is the fragmentation of the swept up shell. The mass spectrum of fragments transforms into the mass spectrum of stars. If the shell forms massive fragments, which may produce massive stars, these stars will sweep up a new shell. By this process, one star forming event triggers another so that star formation is maintained and can propagate.

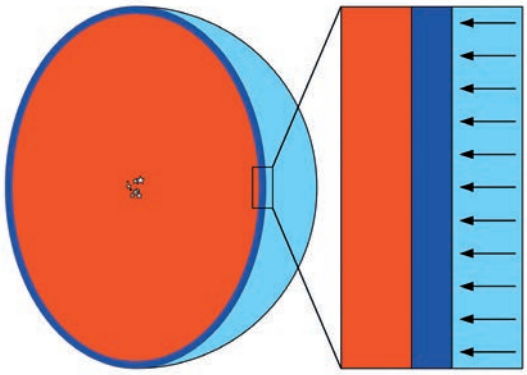


Figure 1.
Model of a shell powered by massive stars

On the contrary, if the shell forms only low mass fragments, which are not able to produce massive stars, these stars will not sweep up a new shell, and star formation terminates.

We note that the star forming process itself contains many elementary physical processes spanning huge length- and time-scales, and as a whole is far from being fully understood. To get the answer about a particular aspect, one has to resort to approximations.

We investigate fragmentation of a swept up shell. Since the fragment sizes are significantly smaller than the shell radius, we neglect its curvature by approximating a small patch on the shell surface by a layer (see Figure 1). The model layer is characterised by three parameters: its surface density S, sound speed c and external pressure P_{ext} acting on its surfaces. Only one dimensionless quantity, $A, A = 1/\sqrt{(1+2*P_{ext}/(\pi G S^2)}$, can be formed from these parameters, so our parameter space is only one dimensional. In the presented simulations we fix surface density and sound speed, and change parameter A by varying only the external pressure.

At the beginning, the simulated layer is seeded with small amplitude perturbations. We investigate its evolution in linear regime (i.e. when the amplitudes are still small so the problem can be described with the dispersion relation), and the following non-linear regime when the amplitudes become large so that the layer breaks into fragments.

Methods

We use MPI-parallel hydrodynamic code Flash4.0 [1]. The code is a standard state-of-the-art tool for modelling astrophysical fluid phenomena. The code uses adaptive mesh refinement and enables us to simulate many relevant astrophysical processes (self-gravity including periodic and mixed boundary conditions, gravitational collapse to a protostar, ionising radiation, ...). On Anselm, the code with the adopted choice of physical processes scales almost linearly at least up to 256 cores.

Results

In the simplest case, we study the fragmentation of layers confined with thermal pressure from both surfaces. Our set of simulations span a wide range of degree of pressure confinement.

We find that there are two qualitatively different kinds of fragmentation depending on the relative importance of the external pressure. If self-gravity of the layer is more important than the confining pressure, the emerging fragments are already so massive that once they form, they collapse monolithically to protostars (see Figure 2). On the contrary, if the confining pressure dominates self-gravity, the layer breaks into a large number of gravitationally stable fragments. The fragments then continuously grow due to coalescence until they assemble enough mass to become gravitationally unstable and collapse to protostars (see Figure 3).

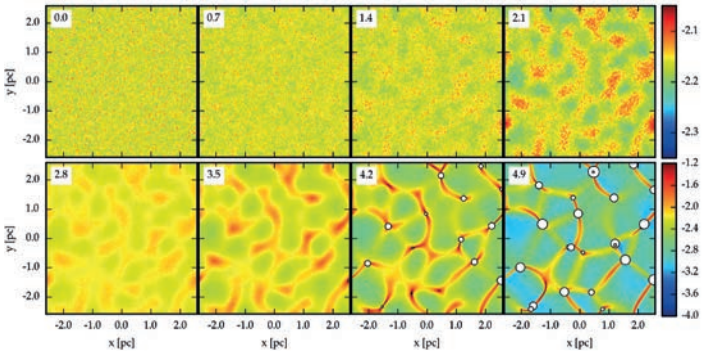


Figure 2.
Evolution of surface density for the self-gravity dominated layer (monolithical collapse). Time in millions of years is indicated in the upper left corner. The layer breaks into gradually slendering filaments with protostars (circles) at their junctions.

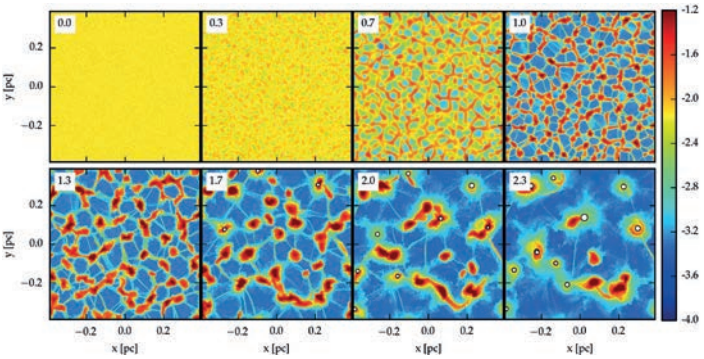


Figure 3.
Evolution of surface density for the external pressure dominated layer

Apart from this qualitative description, we determine the fragmenting time (i.e. the time when the majority of the gas is already in gravitationally bound objects). From dimensional analysis it follows that the layer fragments at time $t=f(A)c/(GS)$, where the only unknown is the functional form of f. We calibrate this function on our simulations.

Although layer fragmentation is a non-linear process, the dispersion relation, which itself describes only the linear part of fragmentation, is widely used for estimating the final fragmenting time. Comparing the estimates with simulations, we assess the error done by this approximation. Analytical estimates are reliable unless the layer is dominated by the external pressure. In this case, the real fragmenting time can be, by a factor of a few, longer since the estimates do not take into account the delay caused by coalescence.

To make the model more realistic, we transform to the reference frame which is in the rest with the wall of the shell. In this frame, a small patch in the wall (i.e. layer) is accreting the ambient medium from the top as the shell expands, and is backed by thermal pressure from the bottom. We perform simulations with three different Mach numbers, 5, 10 and 20 to mimic various evolutionary stages as the shell slows down.

The accreting simulations show analogical behaviour to their non-accreting counterparts. If the accretion is at a high rate so that self-gravity due to the elevated surface density soon overcomes the external pressure, and the majority of fragmentation is accomplished in the self-gravity dominated case, evolution of fragments is similar to that of monolithic collapse. In the opposite case, when the fragmentation occurs entirely in the pressure dominated case, fragments grow due to coalescence. As the shell expands, its surface density increases and the external pressure decreases. Figure 4 shows the evolution of these quantities for three typical values for young star forming regions. Assuming that the shell with instantaneous P_{ext} and S

fragments at the same rate as a layer with these parameters, and using results from our simulations, we estimate the instant of shell fragmentation. The fragmenting time is a function of the density of the ambient medium, temperature of the shell and ionising luminosity of the powering stars. We obtain the mass of fragments at the fragmenting time. This simplified picture indicates that for typical densities in the interstellar medium, and shell temperatures, the fragments are of too low mass for star formation to propagate. This conclusion is in direct opposition to that of [2] because they used analytical estimates without corrections to non-linear behaviour (e.g. coalescence).

On-going research / Outlook

We used significantly simplified models to estimate properties of fragmenting shells. In future work, we may include more realistic thermodynamics, and interaction of the shell with ionising radiation produced by central stars. We also plan to include accretion of highly inhomogeneous medium onto the layer. We expect that the inhomogeneous medium will set typical mass- and length-scales, therefore significantly changing the initial properties of fragments.

The simplified models are also useful for understanding and interpretation of more realistic and complicated simulations. For example, such as those we calculated under another project OPEN-6-21. We are currently analysing the simulations more in-depth.

Conclusion

The present project addresses two issues: (i) how does the external pressure influence fragmentation of a layer, and (ii) what are typical masses of fragments and the fragmenting timescale for an expanding shell.

We find that the role of the external pressure distinguishes between two qualitatively different kinds of fragmentation. If the external pressure is unimportant, the fragments monolithically collapse without merging or further fragmentation. If the external pressure is important, the fragmentation occurs in two steps: the layer first breaks into stable fragments, and the fragments continuously coalesce until they assemble enough mass to collapse.

Based on our idealised model for shell expansion, we conclude that massive stars are not likely to form in swept up shells unless the ambient medium is of very low density, or the shell is at an unusually high temperature.

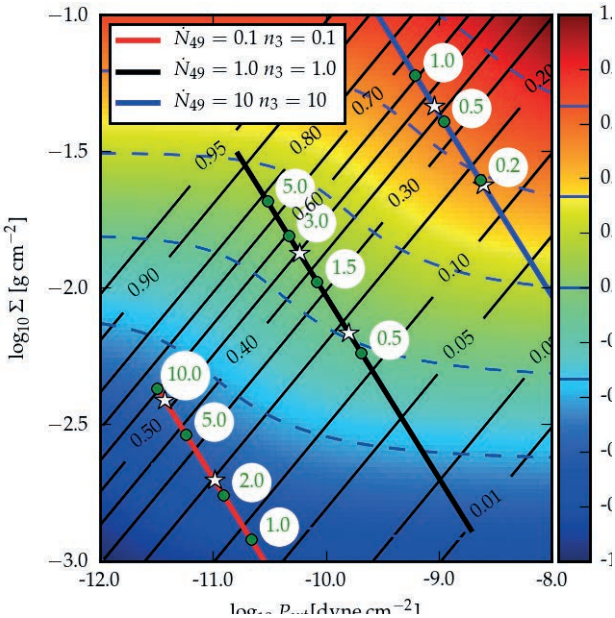


Figure 4.
Evolutionary paths of shells blown by stars producing 10^{48} (red), 10^{49} (black) and 10^{50} (blue) hydrogen ionising photons per second. The shells expand into a homogeneous medium of particle density 10^2, 10^3 and 10^4 cm^{-3}. Time in millions of years is shown by numbered green dots. The beginning and end of fragmentationis shown by white asterisks.

References

[1] Fryxell, B., et al. An Adaptive Mesh Hydrodynamics Code for Modeling Astrophysical Thermonuclear Flashes. The Astrophysical Journal Supplement Series. 2000, 131(1), 273-334. Doi: 10.1086/317361

[2] Whitworth, A. P., Bhattal, A. S., Chapman, S. J., Disney, M. J., Turner, J. A. Fragmentation of shocked interstellar gas layers. Astronomy and Astrophysics. 1994, 290, 421-427.

Publications

[1] Dinnbier, F.; Wunsch, R.; Whitworth, P.; Palous, J.: Fragmentation of vertically stratified gaseous layers: monolithic or coalescence driven collapse (submitted to MNRAS)

[2] Dinnbier, F. Doctoral thesis (in preparation)

ACCRETION AND JET LAUNCHING IN YOUNG STELLAR OBJECTS

Research institution:
Astronomical Institute of the Czech Academy of Sciences

Principal investigator:
Ivana Orlitová

Researchers:
Ghassen Rezgui, Hamed Marzougui, Taieb Lili, Jiří Svoboda

Project partner:
University of Tunis El Manar

Project IDs:
DD-16-8, OPEN-7-29

Introduction

Astrophysical objects, such as stars in their early evolutiona stages (so called Young Stellar Objects), neutron stars, blac holes and active galactic nuclei often present highly collimate outflows of particles, called jets. All of these systems are cha acterised by accretion of surrounding matter onto their gravi centre. Accretion takes place through a disk, formed by du and plasma (ionised gas) orbiting the central object. The jets a launched from inner parts of the disk, and propagate to distan es much larger than the sizes of the systems launching them. this way, the accretion-ejection processes can have an impo tant impact on the environment to large distances, and on th formation and evolution of astrophysical objects. Understandir of accretion-ejection phenomena has grown enormously in th past decade, and physical processes underlying the phenon enological description of the problem are being elucidated. The role of magnetic fields has become evident (e.g. Murphy et al. 2010; Sheikhnezami et al. 2012).

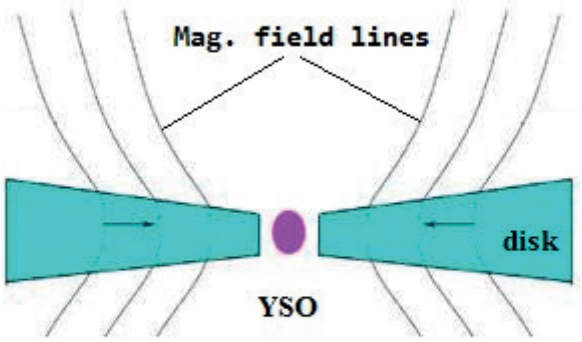


Figure 1.
Section of the accretion disk in rotation around a young stellar object (YSO). The large scale magnetic field threads the disk, connecting then the accretion flow.

Results and Methods

The detailed mechanisms of the jet launching have not yet been fully understood. Our project aims at exploring the role of non-ideal, dissipative magneto-hydrodynamic processes on the accretion disk structure and the jet, which is now an active field of research in various types of systems (e.g. Simon et al. 2015; Bai 2015). We focus on viscous, resistive accretion disks around young stellar objects to explore 1) thermal conduction, which is a mode of heat transfer caused by a difference in temperature between two regions and realised without global displacement of material; and 2) the Hall effect, which induces a drift of charged particles due to their differential motion in a magnetic field. Recent theoretical studies show that both of these effects, originally neglected, can significantly alter the accretion flow, accretion-disk structure, launching of the jet, and its collimation.

We have used the IT4Innovations facilities to carry out an extensive simulation of the accretion disk and jet, using the public, state-of-the-art, magneto-hydrodynamic code PLUTO, which solves mixed hyperbolic/parabolic systems of partial differential equations targeting high Mach number flows in astrophysical fluid dynamics. The code, written in C++ and parallelised, was designed with a modular and flexible structure, whereby different numerical algorithms can be separately combined to solve systems of conservation laws using the finite volume or finite difference approach based on Godunov-type schemes. The discretised equations were solved on a structured adaptive mesh, with the use of the Chombo library for parallel computations over block-structured, adaptively refined grids. The code was successfully tested on the Anselm cluster under the project DD-16-8, and the full simulation was successfully run on Salomon under the project OPEN-7-29.

We have solved time-dependent, viscous, resistive, magneto-hydrodynamic equations, taking into account the disk and jet evolution simultaneously. The aim is to model, under symmetry assumptions, the launching of a jet from a magnetised accretion disk (Figure 1). Since jets and outflows are observed as bipolar streams, we assume that the system is symmetric with respect to the equatorial plane. The computational domain spans a rectangular grid region applying a purely uniform spacing in

the radial direction and a uniform spacing vertically (Figure 2). It covers (512*1,536) uniform grid cells on a physical domain of (r * z) = (40 * 120) ri. The boundary conditions are axial symmetry on the rotation axis (r=0) and equatorial symmetry for the disk mid-plane (z =0). We apply an outflow boundary condition (zero gradient) to the ghost cells at the upper z and r boundaries.
As the choice of boundary conditions on the outer sides of the domain must be appropriate, it is very important to note that the gravitational potential has a singularity at the origin. This very inner part of the domain called sink has the capability of absorption of the fluid from the disk, thus allowing the accretion to the central object. To avoid this singularity we cut out several cells in the bottom left corner of the grid from the computational domain (two cells high in the direction and 14 cells wide in the direction). The sink is described as an internal boundary condition in PLUTO code (see Murphy et al. 2010). A small percentage of fluid is injected by the top of this inner part into the grid at the escape velocity, by imposing a density 1.1 times the local initial density. This maintains an axis dense enough to ensure that Lorentz force cannot produce unphysically low densities on the axis. We also note that to be sure that this internal boundary does not allow artificial jets, we impose in this region of the computational domain the non-positivity of the poloidal velocity.

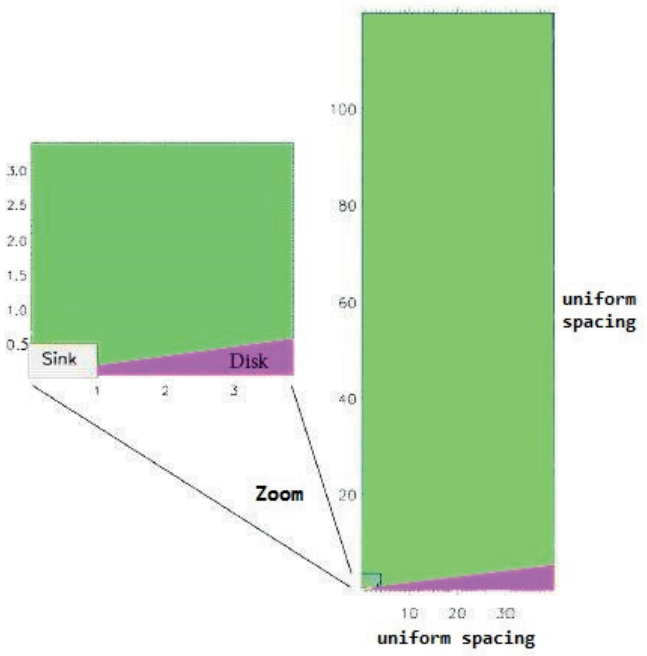


Figure 2.
*Computational domain covering a purely uniform spacing in radial and vertical directions. The grid cells is (512*1,536) resulting in resolution of 0.078 on a physical domain of (r * z) = (40 * 120) ri. Zoom on the internal boundary (sink) is also shown (see text for more details).*

Outlook

Our simulation has recently been completed and we are currently in the stage of data analysis and visualisation. We will focus on the thermal conduction effects, with and without Hall diffusion, on the structure of the disk itself, on the accreted mass fraction that will be loaded into the outflow, and especially on the collimation and acceleration of the jet. We will address the thermal conduction influences on the accretion flow and on the matter ejection from the accretion disk. We will study the Hall diffusion impact on the accretion disk structure and on the jet launching mechanism in the presence of the thermal conduction. Since thermal conduction acts so as to oppose the formation of the temperature gradient that causes it, we expect it to modify the temperature and density profile for accretion flows. We expect changes in the disk structure, in the components of the current density, the accretion and ejection rates, the jet velocity and collimation degree.

Conclusion

The computational power of IT4Innovations has allowed us to efficiently perform a new magneto-hydrodynamic simulation of accretion disk structure in young stellar objects, including non-ideal effects in the presence of a magnetic field. Our results will contribute to the understanding of the jet launching in these objects.

References

[1] Bai, Xue-Ning, et al. Hall Effect Controlled Gas Dynamics in Protoplanetary Disks. II. Full 3D Simulations toward the Outer Disk. The Astrophysical Journal. 2015, 798, 84. Doi: 10.1088/0004-637X/798/2/84
[2] Murphy, G.C., et al. Large scale magnetic fields in viscous resistive accretion disks. Astronomy & Astrophysics. 2010, 512, A82. Doi: 10.1051/0004-6361/200912633
[3] Sheikhnezami, S., et al. Bipolar Jets Launched from Magnetically Diffusive Accretion Disks. I. Ejection Efficiency versus Field Strength and Diffusivity. The Astrophysical Journal. 2012, 757, 65. Doi: 10.1088/0004-637X/757/1/65
[4] Simon, J.B., et al. Magnetically driven accretion in protoplanetary discs. Monthly Notices of the Royal Astronomical Society. 2015, 454, 1117. Doi: 10.1093/mnras/stv2070

ORIGIN OF THE SECOND STELLAR GENERATION IN GLOBULAR CLUSTERS

Research institution:
Astronomical Institute of the Czech Academy of Sciences

Principal investigator:
Richard Wünsch

Researchers:
Tereza Jeřábková, František Dinnbier, Pavel Jáchym, Rhys Taylora

Project partners:
Charles University in Prague, National Institute for Astrophysics, Optics and Electronics, Mexico

Project ID:
OPEN-5-26

Introduction

Globular clusters (GCs) are very massive spherically symmetric objects consisting of 105 – 106 stars concentrated within a radius of several parsecs. There are approximately 150 GCs in our Galaxy, distributed in a spherical galactic halo. Their ages are comparable to the age of the Galaxy itself, i.e. the age of the Universe, and the way how they were formed is one of the largest mysteries in the theory of star formation. Until recently it has been believed that they are simple systems consisting of a single generation of stars with the same age and chemical composition. However, about a decade ago it has been found by very accurate photometric observations that most, if not all globular clusters consist of two or more stellar populations. Moreover, recent spectroscopic observations show that the second and subsequent generations of stars include products of high temperature hydrogen burning, which takes place in certain types of stars.


Globular cluster M15. Credit: NASA/HST

Therefore, the different populations have been interpreted as subsequent stellar generations with later ones formed out of winds or outflows of the previous generation stars. In Tenorio-Tagle et al. (2007) and Wünsch et al. (2008) we suggest a mechanism in which the later stellar generations are formed by the thermal instability and a subsequent gravitational collapse of stellar winds and supernova ejecta shocked in mutual collisions.

Our aim is to test this mechanism using 3D radiation-hydrodynamic simulations. Specifically, we run three types of simulations: (i) high-resolution radiation-hydrodynamic simulations of the first 3.5 million years of the cluster life-time exploring how the second generation stars form out of stellar winds; (ii) grid of medium-resolution hydrodynamic-only simulations of clusters older than 3.5 million years that include exploding supernovae; and (iii) first tests of high-resolution cluster simulations treating massive stars as individual sources (contrary to the previous models where the energy and mass are inserted smoothly into the cluster volume).

Results and Methods

We use the MPI-parallel adaptive mesh refinement hydrodynamic code FLASH4 (Fryxell et al., 2000). Hydrodynamic equations are solved by a standard Riemann solver using the piecewise parabolic method. Effects of ionising radiation are calculated with our own TreeRay algorithm combining the method of reverse ray-tracing with tree. Gravitational forces are calculated by our tree-based solver sharing the most CPU intensive part with TreeRay. Very dense cold gravitationally unstable gas is converted into sink particles, simulating the star formation process.

Objective (i). Initially, we ran a series of low resolution (1283 grid cells) test simulations of fast cooling winds in a star cluster to identify the physically most interesting parameters. Then, we have calculated three main production runs of a cluster with mass 107 Solar masses and radius 3 pc at high-resolution with the finest grid corresponding to 5123. The three production runs consumed most of the 700,000 core-hours devoted to the objective (i). The results confirm that the fast stellar wind can be captured by a combination of the thermal instability and gravitational infall into the centre of the cluster.

The three production runs differ in the so called heating efficiency of gas η_{he} – a quantity determining what fraction of the mechanical energy of individual stellar winds is converted into the thermal energy of the hot gas at shocks when stellar winds collide. We have found that η_{he} is a critical parameter on which properties of the second stellar generation depend. This is illustrated in Figure 1 showing that if the heating efficiency is low ($\eta_{he}=0.05$, top panel), second generation stars are formed only in the very centre of the cluster. On the other hand, the high heating efficiency ($\eta_{he}=0.3$, bottom panel) leads to secondary star formation dispersed throughout the whole cluster. This has an important implication for the GC formation theory, because most models need to assume that a fraction of first generation stars escape from the cluster due to tidal interactions with the galaxy, and it is only possible if the two stellar generations are spatially separated.

Furthermore, we have calculated the expected emission in the hydrogen recombination line H30α from the three production runs (see Figure 2). This line was chosen because it traces the warm gas predicted by our simulations and its frequency (231.9 GHz) allows simultaneous observation of the molecular gas which is assumed to accompany the star formation process. We assume that the cluster is at distance of interacting galaxies NGC4038/9 (20 Mpc) which is the closest known place where such massive young clusters exist. The observations are assumed to be taken by the Atacama Large Millimeter Array (ALMA) with forty 12 m antennae. Results obtained within this project will be used to support a proposal asking for an observing time on ALMA to test the fast cooling winds scenario. Results of objective (i) are described in Wünsch et al. (2017, submitted).

Objective (ii). We ran a grid of medium-resolution (2563) simulations of interacting supernovae in clusters older than 3.5 million years (see Figure 3). These runs are hydrodynamic-only, i.e. radiation, cooling and gravity are not included. The aim is to determine conditions necessary for the emergence of the smooth star cluster wind out of interacting supernova ejecta. For that purpose, we explore how the hydrodynamic flow properties depend on star cluster parameters as, for instance, the supernova rate, star cluster radius and supernova ejecta mass. It has been determined that the critical parameter is the number

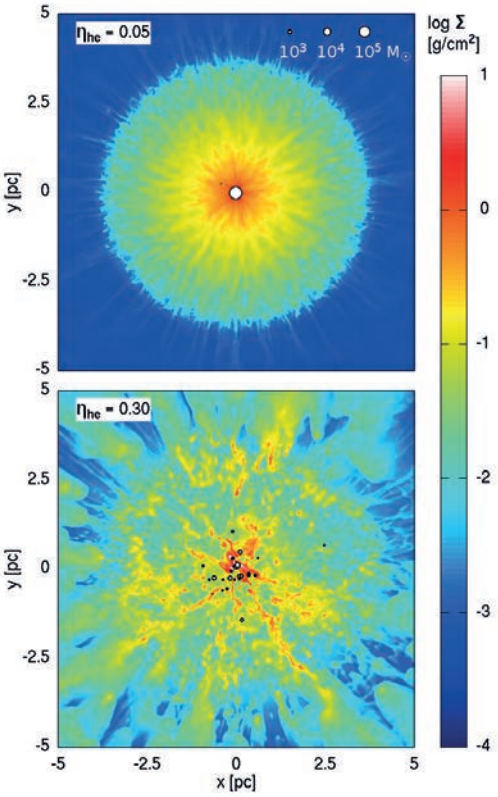


Figure 1.
Gas column density in simulations of 3M yr old cluster with mass 107 Solar masses and radius 3 pc. The two shown models differ only in the heating efficiency being η_{he}=0.05 for the top panel and η_{he}=0.3 for the bottom one. Forming sub-clusters of second generation stars are shown by white circles.

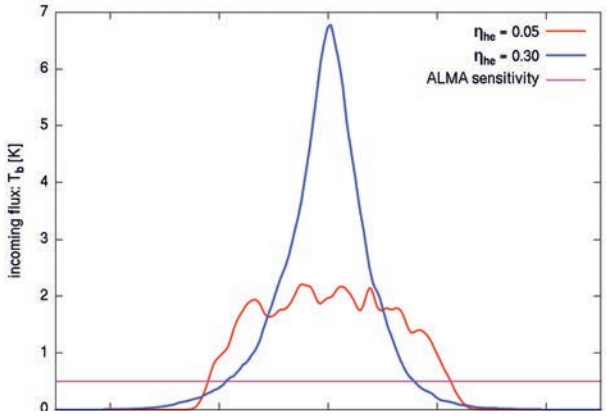
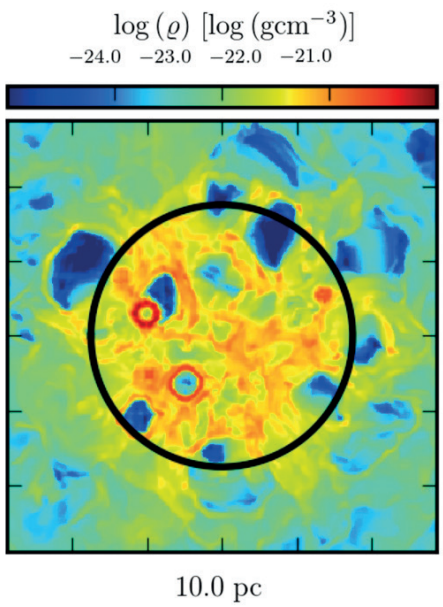


Figure 2.
Synthetic observations of hydrogen recombination line H30α for the two simulation as in Figure 1. The horizontal magenta line at 0.5 K denotes the expected sensitivity of ALMA with forty 12 m antennae for super star clusters in interacting galaxies NGC4038/9 with exposure time of 18 minutes.

of interacting supernovae Π_{SN} and the cluster wind is formed for $\Pi_{SN} \gtrsim 102$. Objective (ii) consumed approximately 200,000 core hours and the results have been published as a part of the diploma thesis of T. Jeřábková; a publication by Jeřábková et al. (2017) is in preparation.

Objective (iii). Using high-resolution (7683) hydrodynamic-only simulations, we were able to confirm that interaction of individual stellar winds leads to the formation of the star cluster wind, with properties similar to the analytical solution by Chevalier & Clegg (1985). This is in agreement with results obtained previously by Cantó et al. (2000). It allows us to prepare the code for future runs with more physical effects included that will allow us to explore the secondary star formation in detail. Objective (iii) consumed approximately 100,000 core hours.



On-going Research / Outlook

In future, we plan to extend our work in two directions. Firstly, we plan to self-consistently calculate whether (and what fraction of) supernova ejecta can be captured by a similar mechanism as stellar winds. This could explain the origin of so called anomalous GC that exhibit variation in the iron abundance. For that, we will need to implement cooling of the hot gas due to emission from the dust produced by supernovae. Secondly, we plan to more accurately simulate secondary star formation using zoom-in simulations resolving individual stars in the star cluster centre.

Conclusion

Using high-resolution radiation-hydrodynamic simulations we confirm that a significant fraction of first generation stellar winds can be captured inside the cluster and feed secondary star formation there.

We have found that the so called heating efficiency of the gas is a critical parameter determining whether the secondary star formation occurs only in the star cluster centre (low heating efficiency) or everywhere throughout the cluster (high heating efficiency).

We have predicted profiles of radio hydrogen recombination line $H30\alpha$ expected to be obtained by observing super star clusters in interacting galaxies NGC4038/9 by ALMA. This allows us to test the suggested model by observations.

Running a grid of medium resolution hydrodynamic simulations we have determined that interacting supernovae should lead to the emergence of the star cluster wind if the number of simultaneously interacting supernovae (Π_{SN}) is large enough.

Figure 3.
Gas density in the mid-plane of the simulation with multiple interacting supernovae

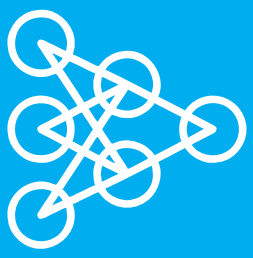
References

[1] Chevalier, R. A., Clegg, A. W., 1985, Nature, 317, 44. Doi: 10.1038/317044a0
[2] Cantó, J., Raga, A. C., & Rodríguez, L. F. 2000, ApJ, 536, 896. Doi: 10.1086/308983
[3] Fryxel, B., Olson, K, Ricker, P., et al., 2000. ApJ, 131, 273-334. Doi: 10.1086/317361
[4] Tenorio-Tagle, G.; Wünsch, R., Silich, S., Palouš, J., 2007, ApJ, 658, 1196. Doi: 10.1086/511671
[5] Wünsch, R., Tenorio-Tagle, G., Palouš., J., Silich, S., 2008, ApJ, 683, 683. Doi: 10.1086/589967

Publications

[1] Wünsch Richard; Palouš, Jan; Tenorio-Tagle, Guillermo; Ehlerová, Soňa; 2017; Astrophysical Journal, submitted
[2] Jeřábková, Tereza; Wünsch, Richard; Palouš, Jan; Tenorio-Tagle, Guillermo; 2017; Astrophysical Journal; in preparation
[3] Jeřábková, Tereza; 2016; Diploma thesis; Faculty of Mathematics and Physics, Charles University

02 | COMPUTATIONAL SCIENCES



VERIFICATION OF MPI APPLICATIONS DEFINED IN KAIRA

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Marek Běhálek

Researchers:
Ondřej Meca, Stanislav Böhm

Project ID:
IT4I-6-4

Introduction

Message Passing Interface (MPI) is a de-facto standard for parallel computations in the environment of distributed memory. It is commonly used in high performance computing (HPC). In our research, we are developing a tool called Kaira [1,2]. It is a development environment for MPI programs. Applications developed in KAIRA are created as combinations of sequential C++ codes and visual programs (based on modified Coloured Petri nets [3]) that capture parallel behaviours and distributed memory aspects. The tool can automatically generate stand-alone MPI applications from this description; but it is not an automatic parallelisation tool as the user is fully responsible for modelling the parallel aspects.

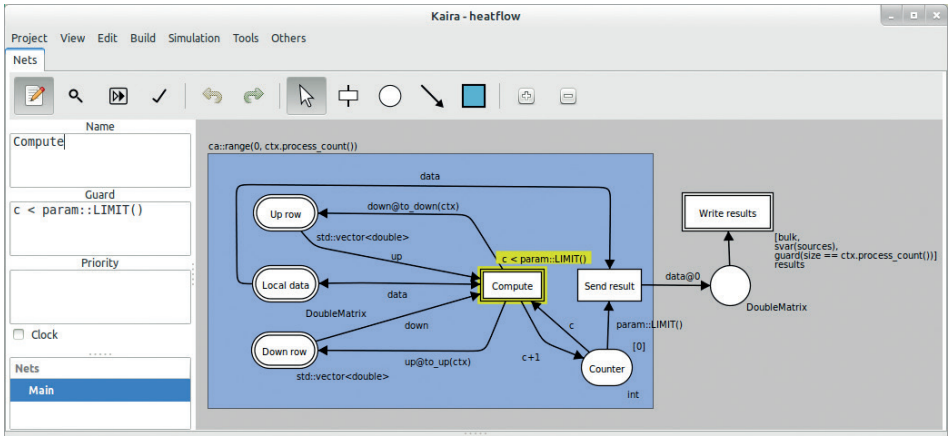


Figure 1. A screenshot of a visual program (heatflow example from Kaira's distribution)

The implemented visual programming model is also a unifying element in our tool. All supportive development activities like debugging, performance analysis, performance prediction and verification are controlled and presented in the same conceptual frame, so results from different analyses should be easily comparable and reusable.

The overall goal of our research is to reduce the complexity in programming MPI applications. Even if MPI is relatively simple to use, it represents a quite low-level interface and direct development in C++ with MPI can be laborious and time-consuming. Furthermore, the complexity of creating distributed applications also lies in other supportive activities like debugging and profiling. Even an experienced programmer can spend a lot of time learning a new set of complex tools. Hence for an experienced MPI programmer, our tool can serve as a fast prototyping tool with easily accessible features like profiling, performance analyses and verification. For a non-expert in this area that just wants to occasionally use a supercomputer without a huge time investment, it represents a stand-alone development environment with a unified control over typical activities occurring during development.

The IT4I-6-4 project focuses on verification features of the Kaira tool. Although there are various tools used during the development of MPI applications, very few tools provide verification features. There are tools monitoring a run of an MPI application and detecting errors like MPI misuse or an occurrence of a deadlock. However, MPI introduces non-deterministic behaviours, and thus two runs of an application may have different outcomes for the same input. To cover all possible behaviours, a systematic scan of the (relevant) state space is usually performed.

Now, we briefly describe verification tools for MPI applications. The first verification tool in this area is MPI-SPIN. It is an extension of the SPIN model checker, which can check the properties of models specified by the PROMELA modelling language. MPI-SPIN is no longer actively developed. The main creator started

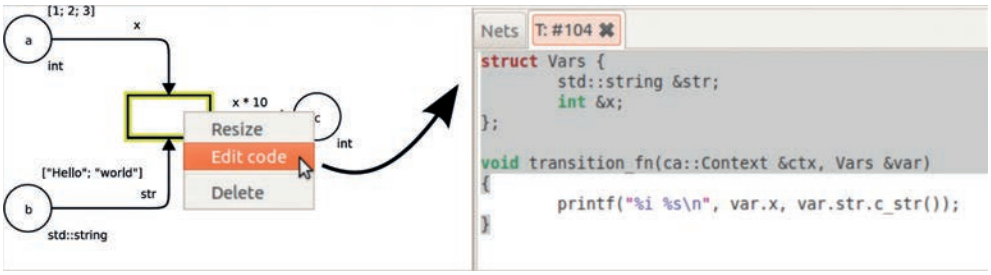


Figure 2.
Code editing in a transition

to work on a new verification tool named TOOLKIT FOR ACCURATE SCIENTIFIC SOFTWARE (TASS). The input for TASS is a C code (C++ is not supported). The advantage of TASS consists in providing symbolic execution. Symbolic execution is able to discover bugs that are input dependent. Unfortunately, TASS only supports a limited subset of the C standard library and MPI; e.g. it does not support non-blocking operations. The tools mentioned thus far create an explicit representation of the state space, which is a complicated task in the case of C/C++ programs. To simplify the problem, the methods of stateless analysis have been introduced. It is implemented by the IN-SITU PARTIAL ORDER (ISP) tool. DISTRIBUTED ANALYZER OF MPI (DAMPI) is a successor of ISP; it verifies distributed memory systems, and it is also based on stateless analysis. MOPPER is built upon ISP; it detects deadlocks by encoding event traces into propositional formulas and a reduction to SAT (satisfiability of Boolean formulas). Another related work is a new, as yet unpublished, verification tool called AISLINN (verif.cs.vsb.cz/aislinn), which is developed by one of the authors of this paper. We can also mention DIVINE [12] as an explicit state space analyser working on C/C++ programs. It is able to verify C/C++ threaded applications, however the current version cannot be directly applied to MPI programs.

Results and Methods

The verification features are implemented like a new module in Kaira. It checks the absence of deadlocks and computation cycles, and verifies the uniqueness of termination (i.e. computing the same result by all runs). Such knowledge can be very useful in our target domain – verification of scientific computing in a distributed environment. For example, considering a task like matrix multiplication, the result (computed matrix) should be independent on the runtime details like the ordering MPI of messages or the number of processes. Moreover, there should be no cycle – the computation should be continuously progressing.

To reduce the state space a new method based on Partial Order Reduction (POR) was implemented. It is especially de-

Program	MPI processes	KAIRA 1 (with POR) time	state	edges	KAIRA 2 (without POR) time	state	edges	ISP	AISLINN	MPI-SPIN
W (5 intervals)	3	0.005s	123	150	0.006s	227	388	3.39s	0.76s	0.034s
W (50 intervals)	3	0.193s	20.9k	27.6k	0.310s	39.7k	74.2k	>1h	1320s	0.34s
W (500 intervals)	3	13.44s	1.9M	2.5M	28.8s	3.9M	7.4M	>1h	>1h	2.48s
W (5 intervals)	5	0.013s	526	700	0.064s	2k	5.4k	17.9s	5.53s	9.88s
W (10 intervals)	5	0.98s	101k	146k	6.52s	411k	1.1M	>1h	434s	54.1s
W (20 intervals)	5	60s	5.7M	8.8M	>1h	>23M	>69M	>1h	>1h	157s
W (10 intervals)	10	210s	12M	16.5M	>1h	>31M	>81M	>1h	324s	>1h
HF	4	0.182s	1206	1205	94s	256k	418k	4.16s	1.17s	69.3s
HF	16	0.724s	4818	4817	>1h	>5.4M	>43M	2.2s	1.9s	>1h
HF	64	9.76s	19266	19265	out of memory			30.1s	6.03s	>1h
MM	4	0.004s	32	31	0116s	1863	7345	0.11s	0.7s	3.1s
MM	16	0.023s	212	211	>1h	>5.4M	>43M	2.2s	1.9s	>1h
MM	64	0.655s	1604	1603	out of memory			10.7s	7.2s	>1h
JS (4 jobs)	3	10.2s	321k	1.4M	17.04s	779k	3.6M	error	>1h	error
JS (7 jobs]	3	346s	9.1M	42M	>1h	>20M	>100M	error	>1h	error
JS (3 jobs)	4	933s	10.9M	70M	>1h	>29M	>180M	error	>1h	error

Table 1.
Execution results of verification of selected experiments

signed to work with MPI and the visual program from Kaira. The new implemented method was published in [4]. We can remark that it has turned out useful to use a modification of the standard independence relation, due to our use of systems with action priorities. The POR techniques allow to explore only some enabled actions when processing a particular state; a crucial requirement is that the omitted actions cannot lead to enabling any unexplored action that is dependent on (i.e. possibly conflicting with) the explored ones. In our case we guarantee this by a straightforward exploration of a simple monotonic condition-event over-approximation of the real system; we thus replace time-consuming computations with sequences of trivial operations.
This simple idea yields satisfactory results. The experiments show that the method is sufficiently precise to significantly reduce the state space, and it presents a good trade-off between the complexity of the analysis and the achieved reduction. To demonstrate the efficiency of the method, we have compared it with other verification tools on common computational patterns appearing in HPC. The following programs are used in the experiment (they are all a part of Kaira's distribution):

- Program Workers (W) – represents a standard centralised parallelisation with the manager-workers pattern. Processes use MPI blocking communication to exchange data.
- Program Heatflow (HF) – computes spreading of heat on a cylinder by an iterative process. This cylinder is divided among processes. Hence, each process exchanges border parts of the computed area with two neighbour processes. MPI non-blocking communication is used to exchange these border parts. The program was verified with the grid size 64x64 and 100 iterations.
- Program Matrix multiplication (MM) – implements the well-known Cannon's algorithm for a distributed matrix multiplication using systolic arrays. Verification was performed with rectangular input matrices and both were divided to sub-matrices with sizes 3x3.
- Program Job stealing (JS) – represents a decentralised load-balancing algorithm. All processes are computing jobs. When a process is finished, it asks its neighbours for work. Jobs are sent by MPI non-blocking communication and a pending job request is checked by the MPI Iprobe function (the function's usage caused a runtime error for ISP and MPI SPIN).

Achieved results are summarised in Table 1. For more details about the performed experiments see [4].

On-going Research / Outlook

In future work we want to focus on deeper analyses of applications and execute verification itself in the distributed memory environment. Right now, it verifies distributed applications implemented in Kaira, but the verification itself is only executed on shared memory systems.

Conclusion

Verification of MPI programs with common communication patterns lead to significant verification times for quite small instances, as was shown by the benchmarks. The proposed solution of using a higher level presentation of (process communication in) programs allows us to implement a more efficient approach, even though the abstract model was not designed for verification purposes. This gives another reason for using our model for prototyping MPI applications. For example, Kaira was successfully used to parallelise the Discrete Self-Organising Migrating Algorithm. This case study is published in [5].
Using KAIRA, a programmer can utilise an environment that supports the development in various directions. Now the environment is enriched by the efficient state-space analysis, which allows verification of larger instances than the current tools verifying directly MPI codes. This feature helps users to recognise potential flaws in the design while they have fully functional versions of their MPI applications.

References

[1] Böhm, S., Běhálek, M., Meca, O. and M. Šurkovský. Visual programming of MPI applications: debugging, performance analysis, and performance prediction. Comput. Sci. Inf. Syst. 2014, 11(4), 1315-1336. Doi: 10.2298/CSIS131204052B
[2] Böhm, S., Běhálek, M., Meca, O. and M. Šurkovský, M. Kaira: development environment for MPI applications. In: Application and theory of petri nets and concurrency. Lecture Notes in Computer Science. Springer, 2014, 8489, 385-394. Doi:10.1007/978-3-319-07734-5_22
[3] Jensen, K. and L.M. Kristensen. Coloured Petri Nets – Modelling and Validation of Concurrent Systems. New York: Springer, 2009. ISBN 3642002846

Publications

[4] Meca, O., Böhm, S., Běhálek, M. and P. Jančar. An Approach to Verification of MPI Applications Defined in a High-Level Model. In: ACSD 2016, Torun, Poland. 19-24 June 2016, to appear.
[5] Běhálek, M., Gajdoš, P. and D. Davendra. Utilisation of Parallel Computing for Discrete Self-organising Migration Algorithm. Chapter in Self-Organising Migrating Algorithm. Studies in Computational Intelligence. 2016, 626, 121-145. Doi: 10.1007/978-3-319-28161-2_6

Project website: kaira.cs.vsb.cz

BENCHMARKING OF THE CFD CODES ON SALOMON CLUSTER

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Tomáš Brzobohatý

Researcher:
Tomáš Blejchař

Project ID:
IT4I-9-5

Introduction

In the last decade, together with growing available computational power, an increase in the complexity of engineering problems to be solved by industry could be observed. In the beginning, the increase in the number of cores per CPU and CPU frequency available in personal computers and workstations was sufficient to keep up with industrial demand. However, the latest requirements from industry in terms of the increasing complexity of problems as well as stress on reducing the computational time significantly push more towards the use of clusters and supercomputers with thousands and even hundreds of thousands of cores to be able to satisfy their needs. Particularly in the area of Computational Fluid Dynamics (CFD), the increase in problem sizes and their complexity drives the development of highly parallel codes in both commercial and non-commercial communities.

One of the main reasons why companies seek the services of HPC centres is that those systems allow them to solve their problems with required accuracy and in an acceptable time. To solve the aerodynamics of a car without a supercomputer, i.e. using a common workstation, a relatively coarse numerical model with approximately 10 million unknowns should be used. Using our HPC system, a much more detailed model with approximately 1.3 billion unknowns could be solved.

The second equally important reason is that using supercomputers could dramatically decrease the required computational time.

Results and Methods

The main objective of this internal project was porting open source software for CFD simulations to the IT4Innovations infrastructure. The researchers at IT4Innovations use a wide portfolio of open source codes for CFD simulation, such as OpenFOAM, Code_Saturne, Palabos, and Elmer, and offer expertise and experience in using these codes.

The application of Open Source Field Operation and Manipulation C++ libraries (OpenFOAM) for solving engineering problems on parallel platforms is presented. OpenFOAM [1] is a free, open source CFD software package developed by OpenCFD Ltd. at the ESI Group and distributed by the OpenFOAM Foundation. It has a large user base across most areas of engineering and science, from both commercial and academic organisations.

The objective of this article is to present the scalability of OpenFOAM on IT4Innovations parallel platforms solving actual engineering problems of fluid dynamics.

We tested the parallel scalability of OpenFOAM with various setups of linear solvers implemented in OpenFOAM (PCG, GAMG…) for several engineering benchmarks to find the optimal solver settings for our clusters. Within the IPCC [2] (Intel Parallel Computing Center) project, we implemented a different version of the PCG algorithm with non-blocking operations, pipelined PCG, for increasing parallel scalability of the conjugate gradient method. The speedup of pipelined PCG related to standard implementation of PCG is depicted in Table 1.

Domains	1,536	1,920	2,688	3,072
Pipelined PCG	1.4	1.3	1.7	2.0
simpleFOAM	1.03	1.03	1.16	1.28

Table 1.
Speedup of pipelined PCG and whole simpleFOAM solver

MPI processes	1536	3072	6144	12288
OF_2.4_IPCCPatch	3.6 [sec]	12 [sec]	60 [sec]	Out of time [4 h]
OF_2.4	96 [sec]	521 [sec]	2942	Out of time [4 h]
Speed UP	26.64	43.31	49	~

Table 2.
Influence of rewriting collective operations to OpenFOAM startup

Figure 1.
Hydraulic valve

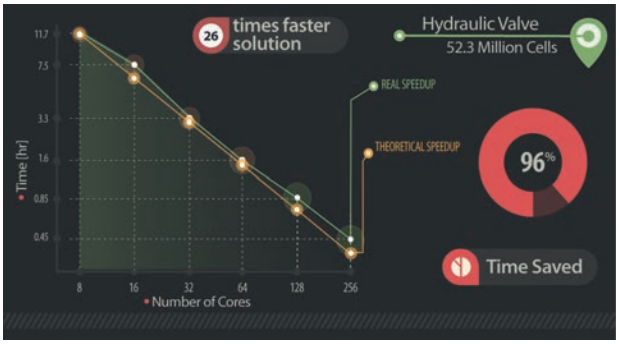


Figure 2.
Parallel scalability – hydraulic valve

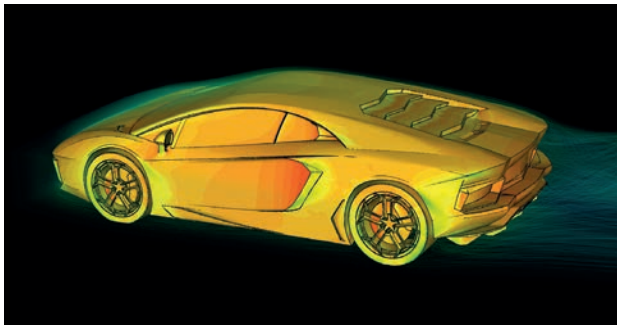


Figure 3.
Sports car external aerodynamic

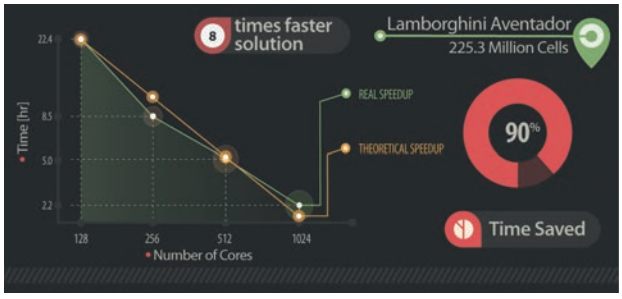


Figure 4.
Parallel scalability – sports car

Thanks to IPCC, start-up of OpenFOAM running on up to 6 thousand cores is possible using a patch for collective operations [3] in the OpenFOAM communication layer. The upper limit for parallel scalability without the patch is only around 1 thousand cores for benchmarks with about 1 hundred million cells (Table 2).

Parallel scalability tests

Now, we present four real world examples of the CFD simulation for testing parallel scalability of the OpenFOAM solver in the IT4Innovations infrastructure:

Hydraulic valve

As the first example, we present a steady-state simulation in the hydraulic valve (Figure 1). For this problem, a numerical model consisting of 52 million cells was created. The mesh was generated by ANSYS Workbench. A 1,000-time step iteration of the standard simpleFoam solver with the k-epsilon turbulent model was used. The computational time for this model using a different number of cores is depicted in Figure 2. From this graph the reduction of the computational time could be observed. For 8 cores, which is the equivalent for common workstations, 12 hours of computational time were needed to solve this problem. The computational time was reduced to 0.45 hours when 256 cores were used.

Sports car aerodynamics

As the second benchmark for testing parallel scalability, we present a problem of external aerodynamics around a sports car body (Figure 3). The model consisting of 225 million cells was generated by the snappyHexMesh utility; a 2,000-time step iteration of the standard simpleFoam solver with the k-omegaSST turbulent model was used. The result of parallel scalability up to 1,024 cores is depicted in Figure 4. The computational time was reduced from 22.4 hours (128 cores) to 2.2 hours when 1,024 cores were used.

Car lamp with radiation

As the next example, we present solution of a car lamp with radiation (Figure 5). The model consisting of 17.3 million cells was generated by ANSYS Workbench and contains six regions. For solving the problem, we use a 20,000-time step iteration of the chtMultiRegionSimpleFoam solver with the fvDOM radiation model. The result of parallel scalability up to 256 cores is depicted in Figure 6. The computational time was reduced from about one day to one hour.

Atmospheric boundary layer

The last benchmark is the steady-state atmospheric boundary layer simulation in an urban area (Figure 7). The numerical model consisting of 14 million cells was generated by the snappyHexMesh utility. A 2,000-time step iteration of the simpleFoam solver with the k-epsilon turbulent model was used. The result of parallel scalability up to 128 cores is depicted in Figure 8.

Conclusion

In this article, we presented parallel scalability of OpenFOAM solvers on several CFD benchmarks using the IT4Innovations infrastructure and have shown some possibilities to significantly reduce solver start-up and increase parallel scalability using a new implementation of the pipelined conjugate gradient method.

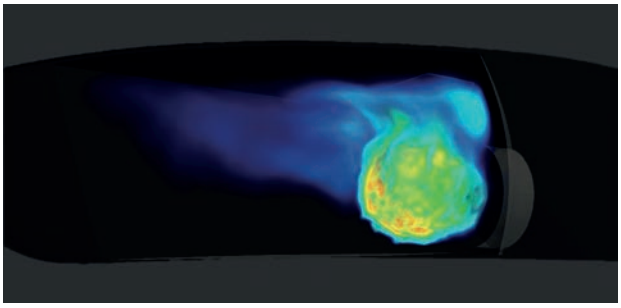


Figure 5.
Car lamp model

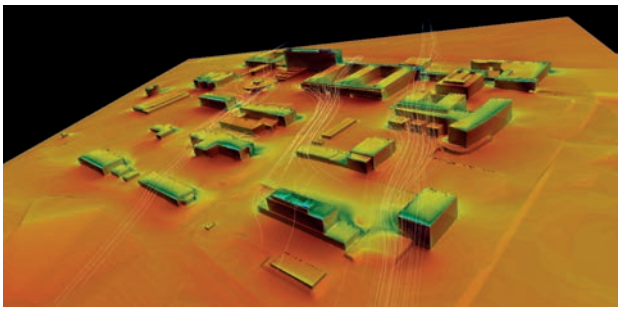


Figure 7.
VSB campus ABL simulation

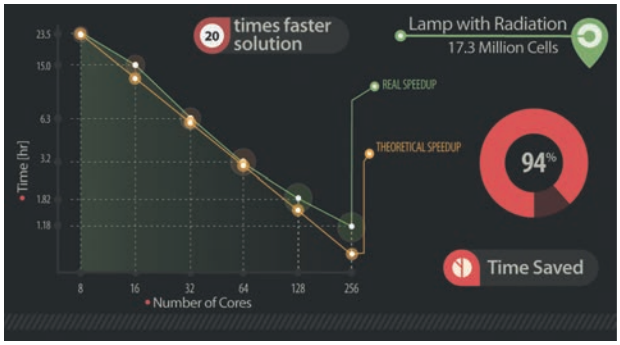


Figure 6.
Parallel scalability – car lamp with radiation

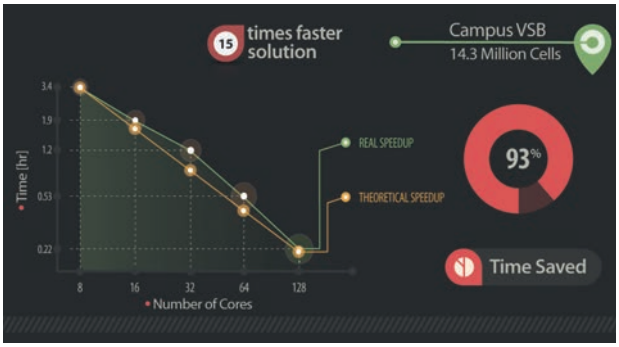


Figure 8.
Parallel scalability – ABL

References

[1] Open Source Field Operation and Manipulation (OpenFOAM) C++ libraries. Available at <www.openfoam.org>.

[2] Intel® Parallel Computing Centers (Intel® PCC). Available at <https://software.intel.com/en-us/ipcc>.

[3] Agrawal, N., Edwards, P., Pandey, A., Klemm, M., Ojha, R. and Rihab Abdul Razak. Performance Evaluation of OpenFOAM* with MPI-3 RMA Routines on Intel® Xeon® Processors and Intel® Xeon Phi™ Coprocessors. In: Proceedings of the 22nd European MPI Users' Group Meeting on ZZZ – EuroMPI '15. New York, New York, USA: ACM Press, 2015, pp. 1-2. Doi: 10.1145/2802658.2802676

Publications

[1] Riha, L., Brzobohaty, T. and T. Karasek. Accelerated OpenFOAM Solvers. Proceedings of the Fourth International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering. Civil-Comp Press, Stirlingshire, UK; 2015, Paper 32. Doi: 10.4203/ccp.107.32

[2] Brzobohaty, T., Riha, L., Karasek, T. and T. Kozubek. Performance evaluation of OpenFOAM on many-core architectures. AIP Conference Proceedings. 2015, 1648, art. no. 830004. Doi: 10.1063/1.4913030

STOCHASTIC ESTIMATION OF THE DIAGONAL OF THE INVERSE (SEDI)

Research institution:
Intel Corporation Exascale Labs, Paris

Principal investigator:
Greg Hernandez (INTEL)

Researcher:
Thomas Guillet (prev. INTEL)

Project ID:
DD-15-6

Introduction

Summary

SEDI (for Stochastic Estimation of the Inverse of the Diagonal) is a new scalable application for selective sparse matrix inversion developed in the context of sparse linear solvers in the framework of the EXA2CT European project.

The basic question is how to compute the elements of the diagonal of the inverse of a large matrix? If this matrix is large (which is the actual trend as problem sizes are increasing), its inversion will require huge amounts of computational resources (processors, memory).

To by-pass this issue, we adopt a stochastic approach (which is an estimation) in order to avoid the matrix inversion and thus reduce the problem complexity and speed-up the process of getting the values of the elements of the diagonal of the inverse of the matrix.

Details

The huge sizes of covariance matrices in data-intensive applications and their dense nature incur an immense (cubic) cost that demands exascale computing resources using matrix inversion. Thus, analysing terabytes of data, which translate to millions of data samples, already requires exaflops of computations, stressing current computing resources to their very limit and rendering practical analysis scenarios for the near future completely intractable, even using state of the art parallel dense linear algebra packages.

We focused on reducing the order of complexity from cubic to quadratic. We moved away from traditional matrix factorisation techniques and adopted stochastic estimation and iterative algorithms. In parallel, we designed our techniques to allow deployment on massively parallel architectures and at the same time achieve high processor performance. The outline of our approach is as follows:

- We turned to stochastic estimation. We retrieved the diagonal entries of the inverse of a matrix by means of a minimum bias stochastic estimator. We showed that by means of a few, with respect to the matrix size, matrix–vector products of the inverse covariance matrix with a set of carefully selected vectors, we can quickly get a quite accurate estimation of the diagonal. It is crucial to point out that for the matrix–vector products with the inverse covariance matrix, by no means did we invert the covariance matrix but rather approximately solved linear systems with it.
- We solved the linear systems by utilising an iterative refinement scheme, which is not based on matrix factorisations. Thus, instead of relying on the Cholesky decomposition of the covariance matrix, we designed an iterative refinement scheme that utilises a small constant number of matrix–vector products with the covariance matrix. The consequence is a reduction of computational complexity from cubic to quadratic.
- In terms of real application, these developments can be applied for uncertainty quantification as described by Bekas et al in 2009 and 2012 for large covariance matrices and also to the OMEN application for Green functions computations. OMEN is a nanoelectronics application, developed at ETH Zuerich, which is widely used in the world of semi-conductors to design new chipsets and architectures.

Results and Methods

SEDI has been derived from an existing prototype developed at USI by F. Verbosio and O. Schenk. The objective of this work is to enhance the scalability of selective inversion algorithms.

The idea is to derive a high performance distributed memory implementation of this scalar Matlab prototype with various conjugate gradient implementations as the problem requires handling large symmetric positive definite matrices. We need to use a sparse operator which is a stencil that is applied on large 3D grids; thin dense bands are used as boundary conditions.

The problem might be non-symmetric: therefore we need to have a special implementation of the conjugate gradient, which is called BiCGstab. In numerical linear algebra, the biconjugate gradient stabilised method, often abbreviated as BiCGstab, is an iterative method developed by van der Vorst for the nu-

merical solution of non-symmetric linear systems. It is a variant of the biconjugate gradient method (BiCG) and has faster and smoother convergence than the original BiCG as well as other variants such as the conjugate gradient squared method (CGS). It is a Krylov subspace method. Note that the BiCGstab method that we have implemented here is unpreconditioned, which can lead to less efficient convergence of the iterative method. This might be improved in the future.

- We have a 3D Cartesian grid with n1xn2xn3 cells.
- This grid is decomposed into p1xp2xp3 MPI ranks and the computation is distributed among these MPI ranks.
- Communications are collective for the dot products computations (they are mainly reduce operations) and point-to-point communications (Send/Recv) are used for the stencil computations.

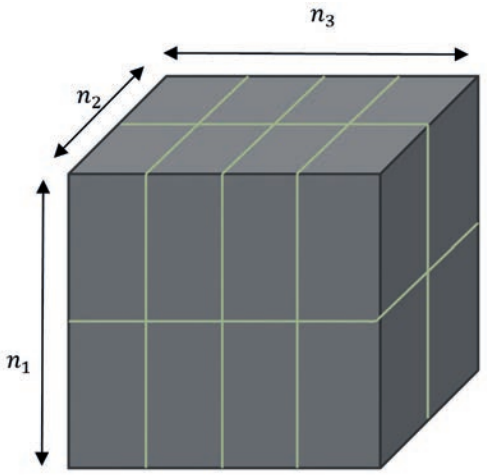


Figure 1.
Domain decomposition

The first results on a 2563 grid allow validation of the convergence of the method as compared to the Matlab reference prototype code. Figure 2 shows that, aside for some discrepancies, residuals converge at the same rate with the same level of accuracy up to 200 iterations, running up to 64 MPI ranks.
During the first iterations, the convergence behaviour is identical across all runs. The sharp spikes and irregular convergence behaviour demonstrate the sensitivity of the BiCGstab algorithm to numerical round off errors.
Scalability of the application has been assessed for various matrix sizes. Figure 3 shows the behaviour of SEDI up to 2,048 MPI ranks for 2563 and 1,0243 matrix sizes obtained on Salomon with the different linear solvers. As convergence is not the key criterion here, SEDI is operated in a slightly different mode: the number of iterations is fixed (here 50) together with an unreachable tolerance (typically 10-9) and a fixed number of samples for the stochastic estimation.
As expected, the three linear solvers exhibit the same behaviour, and the BiCGstab solver is slightly less efficient. This is due to its ability to handle non-symmetric problems and therefore, its workload is higher. The pipelined CG does not show any advantage compared to regular CG (it is even always a bit slower for a small number of MPI ranks) but the pipelined CG performs better for large numbers of MPI ranks.
In terms of memory consumption, the pipelined CG algorithm is the one which requires the biggest amount of memory: up to ~120 GB for the 1,0243 matrix on one core whereas the regular CG only requires 82 GB. Of course, as a domain decomposi-

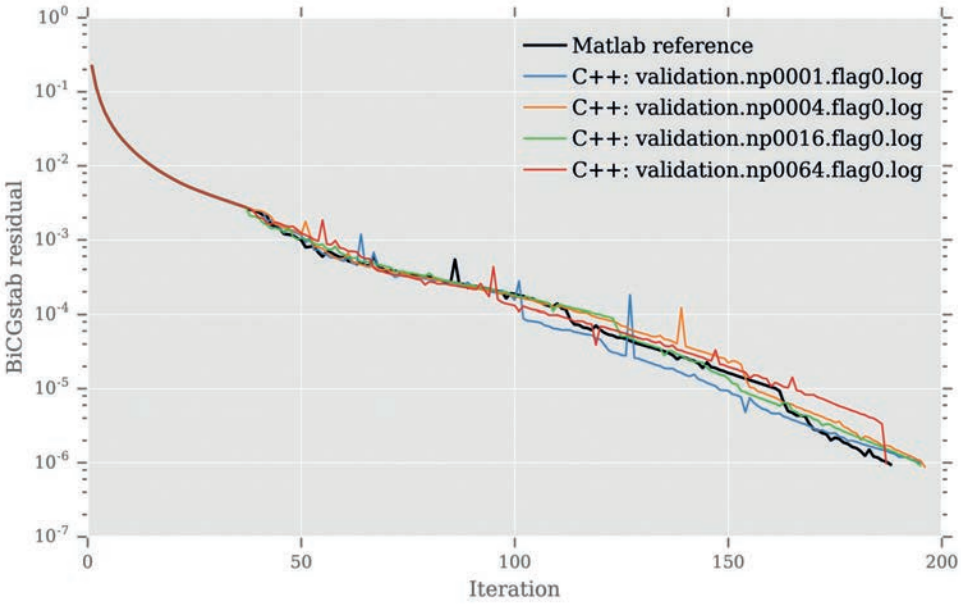


Figure 2.
Validation – BiCGstab residuals for a 2563 matrix size

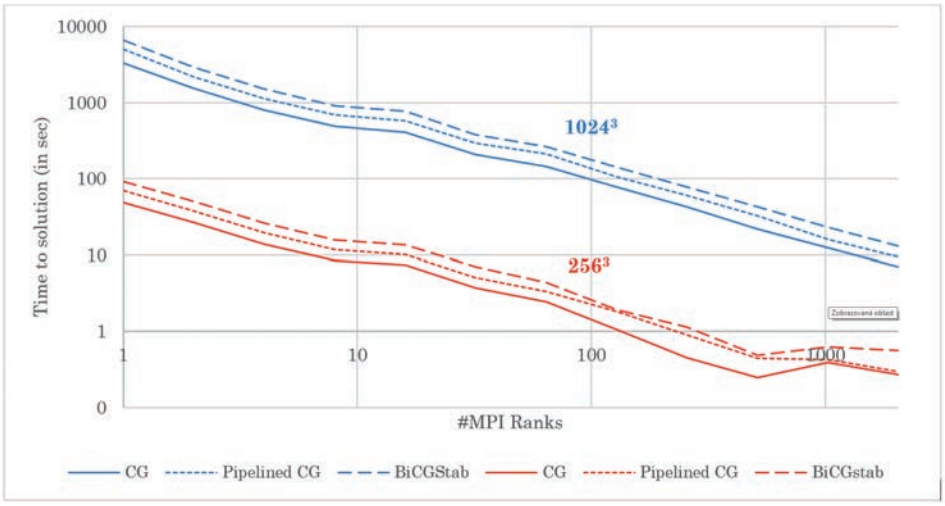


Figure 3.
Scalability of SEDI for various matrix sizes

tion is used within SEDI, the memory requirements are much lower when the number of MPI ranks used for the simulations is increased.

SEDI's scalability is almost linear for a large matrix size whereas we observe a scalability breakdown for smaller matrix size. Figure 4 shows, for the 2563 matrix case, the evolution of SEDI's different internal communication and computation timers described in the Table 1. below:

The communications in the dot products computations increase a lot while increasing the number of MPI ranks as shown by the DOT PRODUCTS timer. Their contribution becomes very important and drive the scalability of the whole application, leading to the scalability breakdown we observed in Figure 3. These communications are collective; basically, these are MPI_Allreduce calls. Both the axpys (MISC COMP timer) and stencil computations (spmv) scale quasi-perfectly for both algorithms (regular CG – not shown here – and pipelined CG). But the cost of the axpys operations is overall quite high compared to the stencil computation.

Evolution of the ratio between computation/communication times versus total time is shown in Figure 5 for a 2563 matrix size (black lines are for a 1,0243 matrix size).

For high number of MPI ranks (i.e. greater than 1,024), the application is fully dominated by MPI communications. Example given here for a 1,0243 matrix shows that this limit will be pushed far beyond. In this regime where the application is communication bound, collective communications account for the major part (as shown in Figure 4).

Communication timers		**Computation timers**	
Name	Description	Name	Description
UNPACK/WAIT	communication time to wait for incoming border cells from other ranks and unpack them into ghosts	DOT PRODUCTS	time spent in the computation of the dot products
DOT PRODUCTS	communications time for the computation of the dot products (these are non-blocking collective communications if using a MPI-3 implementation)	INNER STENCIL	time spent in the computation of the stencil on inner cells for which ghosts cells are not needed
HALO PACK	in the halo computation, this is the communication time to pack inner face cells	BORDER STENCIL	time spent to compute the stencil at inner border face cells
Isend/Irecv	in the halo computation, communication time to send the inner face cells and to post all the receives	MISC COMP	Other computations: combinations (z ← a*u + b*v), copies (z ← u), divisions (z ← u/v), etc...

Table 1.
Communication and computation timers

On-going Research / Outlook

Actual developments in SEDI are focused on the implementation of realistic boundary conditions, which are thin dense bands located at the border of the domain. These boundary conditions will introduce new collective communications in the operator.

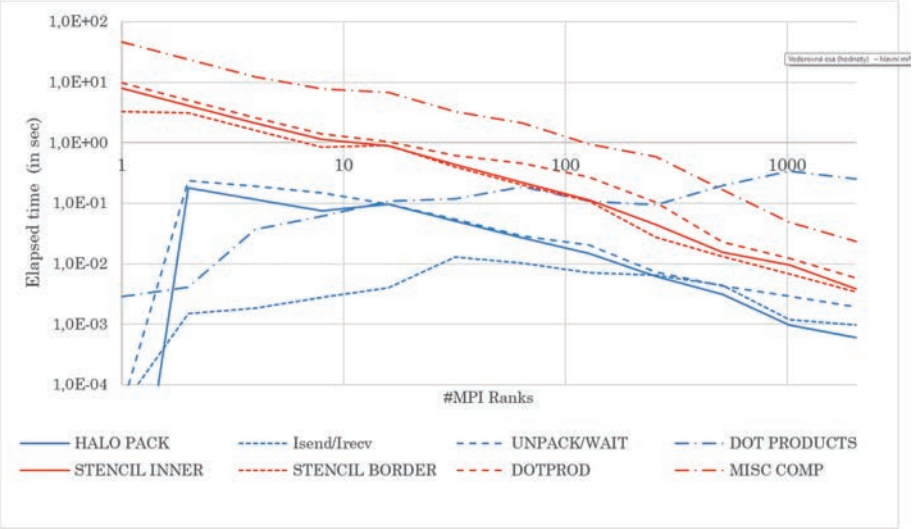


Figure 4.
2563 case – Communication and computation timers (pipelined CG solver)

Conclusion

SEDI is a stochastic approach to estimate the diagonal of the inverse of a huge sparse matrix that avoids the costly process of matrix inversion. It has been developed in the framework of the EXA2CT European project, within which IT4Innovations has been a major partner. SEDI has been validated on the Salomon supercomputer and has proven its parallel efficiency up to large numbers of CPU core counts.

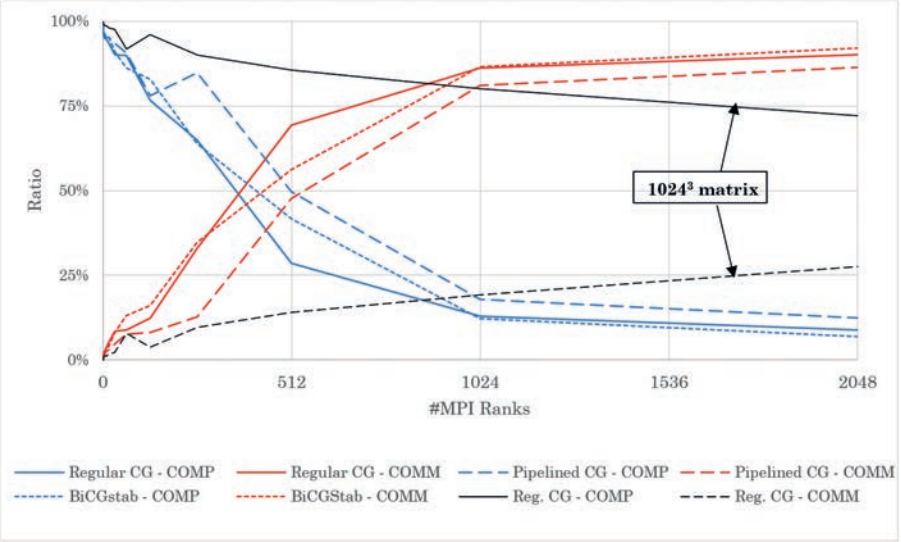


Figure 5.
Computation vs. communication time – 2563 matrix size

References

[1] Bekas et al. Low Cost High Performance Uncertainty Quantification, Proceedings of the 2nd Workshop on High Performance Computational Finance, SC09, Portland, OR, USA, Nov 2009. Doi: 10.1145/1645413.1645421

[2] Bekas et al. Low-cost data uncertainty quantification, Concurrency Computat.: Pract. Exper. 2012; 24:908–920. Doi: 10.1002/cpe.1770

[3] Ghysels P. and W. Vanroose. Hiding global synchronization latency in the preconditioned Conjugate Gradient algorithm, Parallel Computing. 2014, 40(7), 224-238. Doi: 10.1016/j.parco.2013.06.001

[4] Luisier M. et al. Atomistic nanoelectronic device engineering with sustained performances up to 1.44 PFlop/s, SC11, November 12-18, 2011, Seattle, Washington, USA. Doi: 10.1145/2063384.2063387

PERMON TOOLBOX DEVELOPMENT

Research institution:
IT4Innovations National Supercomputing Center

Researchers:
Martin Čermák, Václav Hapla, David Horák, Jakub Kružík, Marek Pecha, Lukáš Pospíšil, Radim Sojka, Jiří Tomčala, Alena Vašatová

Project ID:
IT4I-9-14

Introduction

PERMON (Parallel, Efficient, Robust, Modular, Object-oriented, Numerical) [1,2] is a newly emerging collection of software libraries, uniquely combining quadratic programming (QP) and Domain Decomposition Methods (DDMs). There are two core modules in PERMON: PermonQP and PermonFLLOP. They are built on top of PETSc [3], mainly its linear algebra part. They extend PETSc with new specific functionality, algorithms for large scale sparse QP problems and DDM of the FETI type. The same coding style is used so that users familiar with PETSc can utilise them with minimal effort. Among the main applications are contact problems of mechanics.

PermonQP provides a base for solution of quadratic programming (QP) problems. It includes data structures, transforms, algorithms, and supporting functions for QP. PermonQP is available for free under the FreeBSD open source license on GitHub, see the Download section at [1].

PermonFLLOP (FETI Light Layer on Top of PETSc) is an extension of PermonQP that adds support for DDM of the FETI type. PermonFLLOP is currently under preparation for publishing.

Results and Methods

PermonQP

PermonQP, a general purpose QP solver, allows solving QPs with an SPS Hessian and any combination of linear equality and inequality constraints, box constraints or no constraints. It provides a basic framework for QP solution (data structures, transformations, and supporting functions), a wrapper of PETSc KSP linear solvers for unconstrained and equality-constrained QP, a wrapper of PETSc TAO optimisation solvers offering several additional algorithms for unconstrained and box-constrained QP, a variant of the augmented Lagrangian method called SMALXE, and several concrete solvers for bound constrained minimisation. General linear inequality constraints can be converted to bound constraints using dualisation.

PermonFLLOP

PermonFLLOP (FETI Light Layer on Top of PETSc) is an extension of the PermonQP package, implementing the algebraic part of DDMs of the FETI type. Let us show how PermonFLLOP is implemented from the user's perspective. The domain has to be volume-meshed and decomposed using partitioning software such as METIS. Then virtually arbitrary Finite Element Method (FEM) implementation can be used to generate the subdomain stiffness matrices Ks and the subdomain load vectors fs as sequential data for each subdomain s independently. However, the local-to-global mapping l2g, mapping each subdomain's degrees of freedom to the global degrees of freedom, has to be produced in this phase.

Note that we assume each processor core owns only one subdomain here; PermonFLLOP has nevertheless a new experimental feature of allowing more than one subdomain per core, i.e. an array of Ks and fs is passed per subdomain.

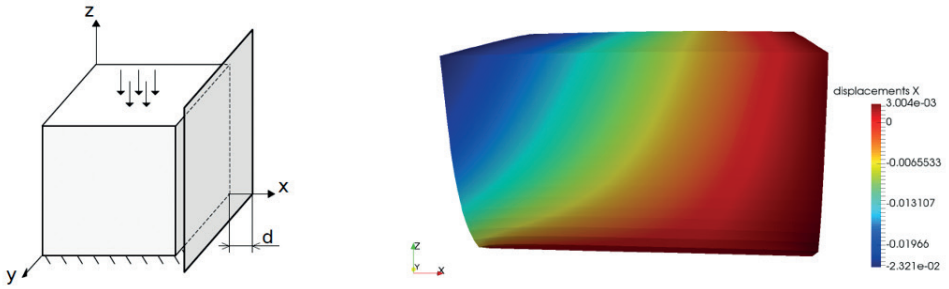


Figure 1.
Contact 3D elastic cube problem

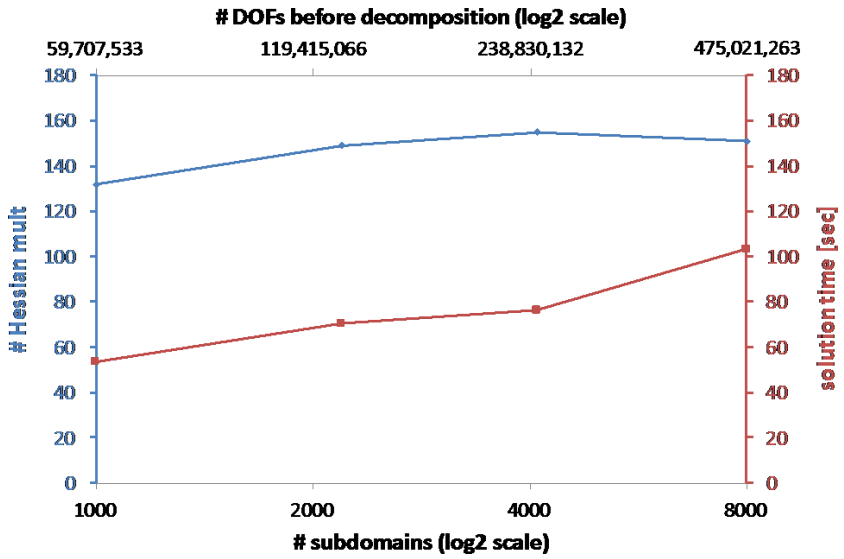


Figure 2.
Numerical and weak parallel scalability on the IT4Innovations Salomon supercomputer

The "gluing" signed Boolean matrix Bg is constructed based on l2g as described in [4]. The FEM software can skip the processing of the Dirichlet conditions and rather hand it over to PermonFLLOP, resulting in greater flexibility. PermonFLLOP allows to enforce Dirichlet boundary conditions either by the constraint matrix Bd (TFETI approach [5]), or by a classical technique of embedding them directly into Ks and fs (FETI-1 approach). It is also possible to mix these two approaches.

The inequality constraint matrix BI describes linearised non-penetration conditions on the contact zones. It is empty for linear (permanent contact only) problems. The global constraint right-hand side vector c possesses an analogous structure. Currently, PermonFLLOP requires BI and cI from the caller.The subdomain nullspace matrix Rs is assembled using one of the following options. The first option is to use a numerical approach, and the second one is to generate Rs as rigid body modes from the mesh nodal coordinates. The latter is typical for TFETI and is considered here.

In the PermonFLLOP's function FllopSolve, PermonFLLOP passes the global primal data K, f, BE = [BTg BTd]T, BI and R to PermonQP, calls a specific series of QP transforms provided by PermonQP, resulting in the bound and equality constrained QP, which is then solved with the QPSSolve function.

From the mathematical point of view, the called QP transforms (QPT) implement modifications transforming the original primal problem

$$\min \frac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u} - \mathbf{f}^T\mathbf{u} \text{ s.t. } \mathbf{B}_I\mathbf{u} \leq 0 \text{ and } \mathbf{B}_E\mathbf{u} = 0$$

into a dual one

$$\min \frac{1}{2}\widehat{\lambda}^T\mathbf{P}\mathbf{F}\mathbf{P}\widehat{\lambda} - \widehat{\lambda}^T\mathbf{P}\mathbf{d} \text{ s.t. } \widehat{\lambda}_I \geq -\widetilde{\lambda_I} \text{ and } \mathbf{G}\widehat{\lambda} = 0.$$

This dual problem formulated using Lagrange multipliers ensuring the gluing and non-penetration conditions is significantly smaller and better conditioned compared to the primal one.

Numerical experiments

The results were obtained at the Salomon cluster. As a model 3D linear elasticity contact problem, we consider an elastic cube with the bottom face fixed generated by our PermonCube benchmark generation with top face loaded with a vertical surface force fz = -465N/mm2 directed downwards, and the right one in contact with a rigid obstacle, Young modulus E = 2.105 MPa, Poisson ratio 0.33, see Figure 1.

The graph in Figure 2 demonstrates both numerical and weak parallel scalability up to 500 millions of unknowns and 8,000 subdomains with one subdomain per one computational core. The contact problem was solved using SMALBE and MPRGP with our new adaptive expansion steplength, which significantly improved this scalability and reduced not only the number of expansion steps but also the number of conjugate gradient steps. Parallel direct solver Super_LU was used for the coarse problem solution.

Conclusion

The PERMON team was successful to push the scalability limits for contact mechanics benchmarks on Salomon up to nearly 500 millions of unknowns and 8,000 subdomains (cores).

References

[1] PERMON (Parallel, Efficient, Robust, Modular, Object-oriented, Numerical). URL http://permon.it4i.cz.
[2] Hapla, V. et al.: Solving contact mechanics problems with PERMON. In: HPCSE 2015, Revised Selected Papers, Lecture Notes in Computer Science, vol. 9611, pp. 101-115. Springer International Publishing Switzerland, 2016. Doi 10.1007/978-3-319-40361-8_7.
[3] Vašatová, A., Čermák, M., and Hapla, V.: Parallel implementation of the FETI DDM constraint matrix on top of PETSc for the PermonFLLOP package. In: PPAM 2015, Revised Selected Papers, Part I, Lecture Notes in Computer Science, vol. 9573, pp. 150-159. Springer International Publishing Switzerland, 2016. Doi 10.1007/978-3-319-32149-3_15.
[4] Balay, S. et al.: PETSc (Portable, Extensible Toolkit for Scientific Computation). URL http://www.mcs.anl.gov/petsc.
[5] Dostál, Z., Horák, D., and Kučera, R.: Total FETI – an easier implementable variant of the FETI method for numerical solution of elliptic PDE. Communications in Numerical Methods in Engineering 22 (2006), 1155-1162. Doi 10.1002/cnm.881.

Publications

[1] Sojka R., Horák D., Hapla V.: The impact of enabling multiple subdomains per MPI process in TFETI domain decomposition method, Applied Mathematics and Computation, submitted 2016
[2] Kružík, Hapla, Horák, Pecha: Comparison of selected FETI coarse space projector implementation strategies, Applied Mathematics and Computation, submitted 2016
[3] Cermak, Hapla, Kruzik, Markopoulos, Vasatova: Comparison of different FETI preconditioners for elastoplasticity, Computers and Mathematics with Applications, accepted
[4] Merta, Zapletal, Brzobohaty, Markopoulos, Riha, Cermak, Hapla, Horak, Pospisil, Vasatova: Numerical libraries solving large-scale problems developed at IT4Innovations Research Programme Supercomputing for Industry. Perspectives in Science. 2016, 7, 140-150. Doi: 10.1016/j.pisc.2015.11.023. ISSN 22130209.
[5] Hapla, Horak, Cermak, Sojka, Pecha, Kruzik, Tomcala, Vasatova, Pospisil: PERMON toolbox development, in proceedings of PANM 2016, submitted
[6] Hapla, Horák, Pospíšil, Čermák, Vašatová, Sojka: Solving Contact Mechanics problems with PERMON, in proceedings of HPCSE 2015. Lecture Notes in Computer Science, 2016, vol. 9611. Doi: 10.1007/978-3-319-40361-8_7
[7] Cermak, Hapla, Horak, Pospisil, Vasatova, Markopoulos: PERMON software toolbox as solver of contact problems in mechanics. In proceedings of ICNAAM 2015, AIP Conf. Proc. 2016, 1738(1). Doi: 10.1063/1.4952135

TSUNAMI SIMULATION WITH PARALLEL ADAPTIVE MESH REFINEMENT ON THE XEON PHI NODES OF SALOMON

Research institution:
Technical University of Munich

Principal investigator:
Michael Bader

Researcher:
Chaulio Ferreira

Project partner:
IT4Innovations National Supercomputing Center

Project ID:
DD-16-23

Introduction

Sam(oa)² (Space-filling curves and Adaptive Meshes for Oceanic And Other Applications) is a parallel framework for solving 2D partial differential equations on dynamically adaptive triangular grids. With its kernel-based element-oriented interface, application programmers can imple-ment numerical methods efficiently without knowledge of the complex underlying mesh algorithms.

One of the scenarios currently implemented in sam(oa)² uses Shallow Water Equations and a Finite-Volume dis-cretization (based on [1]) to simulate Tsunami Wave propa-gation (Figure 1). This scenario has recently been opti-mised for Intel Xeon Phi architectures (via introduction of regularly refined patches and vectorisation), and we cur-rently have a situation where Xeon Phi coprocessors achieve performance comparable to that of regular "host" CPUs. This is strong motivation for using the so-called symmetric mode in heterogeneous platforms such as the Xeon Phi-accelerated nodes of the Salomon Supercomput-er. Thus, in this research we experimented with dynamic load balancing on heterogeneous compute nodes using a Haswell CPU and Xeon Phi coprocessor in the symmetric mode, and aimed to maximise time to solution on those nodes.

This extended abstract is a short version of a conference paper submitted to the PASC 2017 Conference [6].

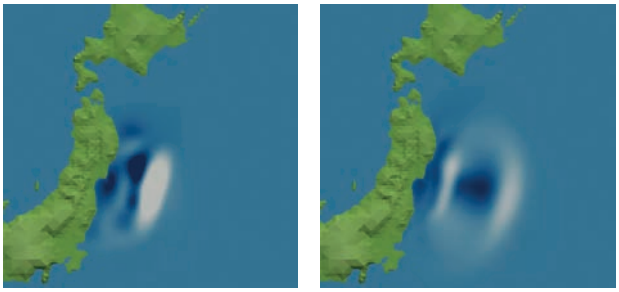


Figure 1.
Simulation of the Tohoku 2011 Tsunami

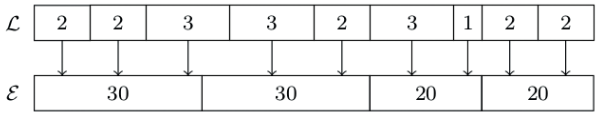


Figure 2.
Heuristic solution for the heterogeneous chains-on-chains problem. In this example, it distributes nine sections with loads L=(2,2,3,3,2,3,1,2,2) to four processors with execution speeds E=(30,30,20,20). The centre of each section defines which processor it is assigned to.

Results and Methods

Using heterogeneous platforms may lead to serious load imbalance issues. Sam(oa)² features a load balancing strat-egy that takes care of imbalances created by the mesh's dynamic adaptivity [4], but our original implementation assumed that all processors had uniform efficiency, dis-tributing the computation load homogeneously. Since this is not the case on the heterogeneous nodes in Salomon, we have extended the load balancing algorithms in sam(oa)² to consider heterogeneous systems as well [5].

The new implementation uses a heuristic based on ideas from [2], which maps a set of computational tasks with different loads to a set of processors with different execu-tion speeds, aiming to minimise the maximum execution time among all processors. That optimisation problem is formally known as heterogeneous chains-on-chains parti-tioning, for which an in-depth overview can be found in [3]. Figure 2 shows an example of how our partitioning algorithm works.

In order for the partitioning heuristic to achieve maximal performance, the model must contain appropriate values for the processors' execution speeds, which leads to the problem of defining them accurately. Two strategies for choosing these speeds were used: (1) defining them ex-plicitly after experimenting with different configurations and choosing the one with highest performance; and (2) using an in-situ auto-tuning algorithm that computes the actual execution for each processor speed in the previous time step, using on-the-fly runtime measurements.

Single-Node performance.
We started by considering a single Xeon Phi-accelerated Salomon node, and attempted to maximise the simulation performance using different combinations of the node's resources. In Figure 3, we compare the performance obtained when using only the dual-socket Haswell host, only one Xeon Phi, both Xeon Phis, and, lastly, using all four devices available in the node. This last case was actually split into three experi-ments, each applying a different load balancing strategy: "Homog." stands for our original implementation, which is unable to properly deal with processor heterogeneity. "Manual" and "Auto" stand for the heterogeneous load balancing implementation, with execution speeds defined either explicitly or implicitly, as discussed above. From these results we observe that using the symmetric mode does pay off, since even the original homogeneous load balancing performs better than using either only the Host or only the two Xeon Phis. Load balancing with ex-plicit values for processor speeds (in this case, giving 36% of the load to the Haswells and 64% to the Xeon Phis) was able to increase the symmetric mode performance by 16%. Our auto-tuning algorithm also delivered a decent speedup, performing very close to the best configuration found by the explicit strategy.
However, we also observe that the performance obtained by two Xeon Phis and by the symmetric mode are consid-erably smaller than what one could expect, when compar-ing with the performance of the individual devices. By breaking down the execution time into the different components of the simulation code (Figure 4), we found that the components that depend heavily on MPI commu-nication performed very badly in these experiments, while the ones with little or no communication did not have the same problems. We used the Intel® MPI Benchmarks to replicate the communication patterns used by sam(oa)², and noticed that intra-node communication involving Xeon Phis performed substantially slower than inter-node communication between Haswells. We concluded that this was the reason for the far-from-perfect scalability. The underlying problems with MPI communication on Salomon (only in symmetric mode) are still being investigated.

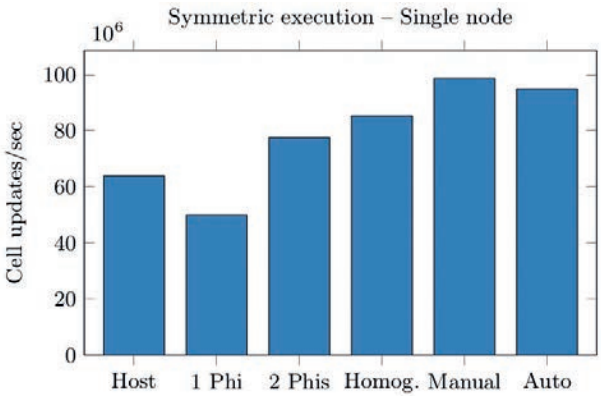


Figure 3.
Comparison of single-node performance with Haswell and Xeon Phi architectures and with symmetric mode using different load balancing strategies.

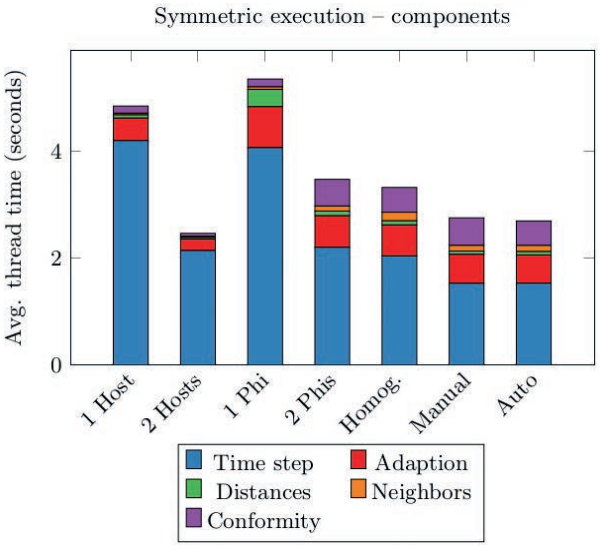


Figure 4.
Breakdown of the average thread time into components. Components that depend heavily on MPI communication (Adaption and Conformity) did not scale well on Xeon Phis.

Multi-Node performance.
We also performed similar experiments using multiple Salomon nodes, attempting to reproduce a weak scaling setting (which is not trivial for dynamically adaptive grids). The results are presented in Figure 5, where the performance is normalised to that achieved using only the Haswell processors.
We observe that, for all node counts, symmetric mode is faster than using either only the Haswells or only the Xeon Phis, and using heterogeneous load balancing with execu-tion speeds defined explicitly can substantially increase performance. However, these differences reduce as the number of nodes increase, due to the communication overhead. We also note that

our auto-tuning strategy based on runtime measurements was able to achieve good performance with only 1 and 2 nodes, but failed to do it with 4 and 8 nodes. Again, this is due to the growing communication costs, which also affect the runtime meas-urements used to compute the processor speeds.

Conclusion and Outlook

The combination of two Haswell host CPUs and two Xeon Phi coprocessors per node in Salomon is strong motivation for using the coprocessors in symmetric mode, instead of native or offload approaches. For a tsunami simulation in sam(oa)² we showed that symmetric mode can improve time-to-solution compared to using only Haswells or only Xeon Phis, especially when the load balancing algorithm takes the system heterogeneity into account. Particularly, we implemented an approach that uses on-the-fly runtime measurements to automatically determine the relative load distribution. This approach delivered reasonable performance on small node counts, but not on a greater number of nodes.

We also found that scalability is considerably limited due to the inefficiency of MPI communication involving Xeon Phi coprocessors in symmetric mode. As the next Xeon Phi generation (Knights Landing) can be run as a standalone CPU by default, we expect that these MPI communication issues will vanish. Hence, we may expect that sam(oa)² will be able to provide efficient parallel adaptive mesh re-finement on heterogeneous platforms that include KNL CPUs.

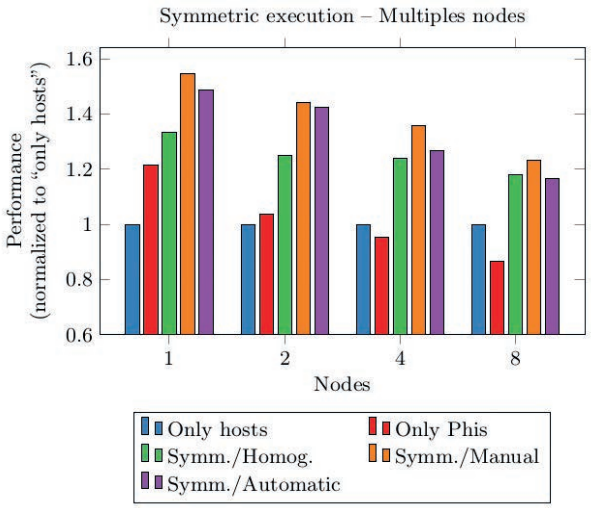


Figure 5.
Experiments with multiple nodes comparing performance of different load balancing strategies. All values were normalised to "Only hosts".

References

[1] LeVeque R. J.; D. L. George, and M. J. Berger. Tsunami modelling with adaptively refined finite volume methods. Acta Numerica, 20:211–289, 2011. Doi: 10.1017/S0962492911000043

[2] Pinar A. and C. Aykanat. Fast optimal load balancing algorithms for 1D partitioning. J. Parallel Distrib. Comput., 64(8):974–996, 2004. Doi: 10.1016/j.jpdc.2004.05.003

[3] Pınar A.; E. K. Tabak, and C. Aykanat. One-dimensional partitioning for heterogeneous systems: Theory and prac-tice. Journal of Parallel and Distributed Computing, 68(11):1473–1486, 2008. Doi: 10.1016/j.jpdc.2008.07.005

Publications

[4] Meister O.; K. Rahnema and M. Bader. Parallel, Memory Efficient Adaptive Mesh Refinement on Structured Triangu-lar Meshes with Billions of Grid Cells. ACM Transactions on Mathematical Software, 43(3): 19:1–19:27, 2016. Doi: 10.1145/2947668

[5] Meister O. Sierpinski Curves for Parallel Adaptive Mesh Refinement in Finite Element and Finite Volume Methods. PhD thesis, Department of Informatics, Technical University of Munich, 2016.

[6] Ferreira C. and M. Bader. Load Balancing and Patch-Based Parallel Adaptive Mesh Refinement for Tsunami Simulation on Heterogeneous Platforms Using Xeon Phi Coprocessors. Submitted to The PASC 2017 Conference.

PARALLELISATION OF THE PRAGTIC SOFTWARE FOR AN AUTOMATED FATIGUE DAMAGE CALCULATION

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
David Horák

Researchers:
Jiří Tomčala, Marek Pecha

Project partner:
Czech Technical University in Prague

Project ID:
IT4I-9-14

Introduction

Repeated service loading of machine parts leads to the reduction of their load-carrying capacity and often to the fatigue failure. Because the experimental verification of their long-term functionality is very expensive and time consuming, engineers are interested in virtual modelling and simulating fatigue failure. The quality of these simulations is limited by (i) reliability of phenomenological models – used prediction criteria, cyclic plasticity model; (ii) involved numerical methods, algorithms, and their implementations; (iii) computational resources being at disposal for the simulations.

Fatigue analysis requires very accurate results of the stress-strain calculation at critical locations of the finite element (FE) model. After results are obtained, the fatigue analysis takes place, usually being realised in another software. One of the available tools for computational fatigue analysis is PragTic.

PragTic software is developed at CTU in Prague. It serves as a fatigue analysis tool using the computation results of the FE-solution. It is provided as a freeware on the PragTic webpages [2]. Many new fatigue criteria and approaches have been implemented into this code during the last fifteen years. PragTic also has a direct connection to the large material and experimental databases FatLim (Fatigue Limits) and FinLiv (Finite Lives). The major focus of PragTic is multiaxial fatigue analysis. It is very important in cases where more load channels act on the component simultaneously. The analysis is therefore usually realised by evaluating stress or strain components on various planes in the analysed point.

The damage parameter built from these components is either maximised over all possible planes or integrated over them. Obtaining the single equivalent stress or strain by some criterion gets complicated, because many candidate planes have to be evaluated. In addition to it, the computation time is increased even more due to the need to correctly evaluate shear stress or strain components in cases of non-proportional loading, when the shear stress or strain vector rotates. The current normal solution is to compute the position and radius of the minimum circle circumscribed to the shear stress vector tip trajectory in each load cycle.

The major feature of the PragTic software is the multiaxial fatigue analysis. This analysis type is used if multiple load channels simultaneously act on a mechanical component. In these cases, the analysis process is usually proceeded by the stress-strain component evaluation on various planes, which go through the analysed point. The damage parameter is then set up from these components and further processed as follows: it is maximised or integrated over all evaluated planes [6].

The PragTic software predicts mechanical fatigue failure by generating large-scale computational simulations. This approach demands extensive computational resources and a large amount of time. By the parallelisation of this software at the various levels, we can save a lot of the run time. It enables finding a much more detailed solution. In this paper we present the parallelisation at the node level.

Former version and its inefficiency

A former version of the PragTic software is accessing the disc very often through the simulations. Therefore, it wastes a huge amount of the run time by writing, reading, writing, reading, etc., of partial results to/from the disc. The PragTic uses files on disc as the temporary buffers, which cause a significant slowdown of the program. These temporary buffer files are erased at the end of the run.

Development of the PragTic fatigue solver was started by Jan Papuga in 2000, and various institutions have gotten involved throughout the past years – above all the Czech Technical University, the Evektor Company, and the VŠB-Technical University of Ostrava. Since the last two decades, a whole range of new criteria and approaches on fatigue analysis have been implemented. It is currently distributed as a freeware application, and it can be downloaded from the PragTic webpages [2].

Results and Methods

Main idea

PragTic sequentially generates simulations node by node. The main idea of our first parallelisation is to assign the particular subset of nodes to every parallel process, then compute partial

solutions and finally merge the results into the one result file. If we use for example 2,000 processor cores, then the theoretical maximum speed up would be 2,000. Of course, there is always the run time overhead cost, so the real speed up is lower. It is caused by distributing the particular subsets of nodes at the beginning and then merging the results after the partial computations.

Further significant speed up can be reached by not saving the temporary buffers to the disc, but keeping them just in the memory. It is a current approach compared to the 20 year old one, when computers had significantly less operating memory than today.

Implementation

The first step was to compile the PragTic on the Salomon cluster. We then parallelised the PragTic using MPI and got the first results.

Results

We have chosen to test the small benchmark of 120 nodes on the Salomon cluster. The loading was sin(t)+ sin(1000*t) where t means time. See the illustrative shape of the loading in Figure 1.

The result run times are shown in Figure 2. The red colour represents the run time of the version saving temporary buffers to the real disc (former version of the PragTic), the green colour represents the run time of the version saving temporary buffers to the shared memory (/dev/shm), and the yellow colour represents the run time of the version saving temporary buffers to the ramdisk (/ramdisk). As the compute nodes do not see each other's shared memory and ramdisk, their run time graph line ends at 20 processor cores. The Salomon's compute nodes contain just 24 processor cores [1].

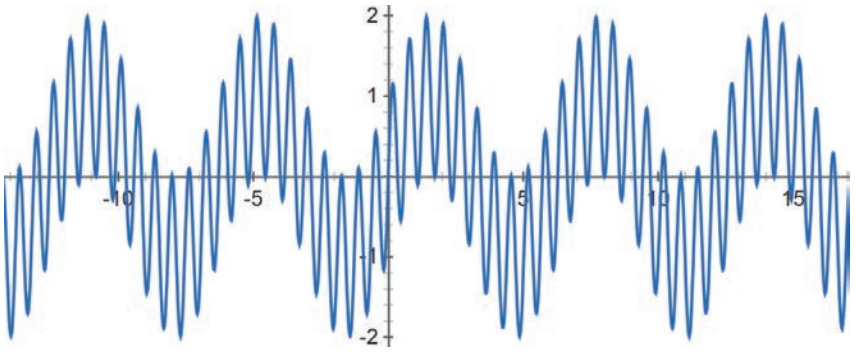


Figure 1.
The illustrative shape of the loading

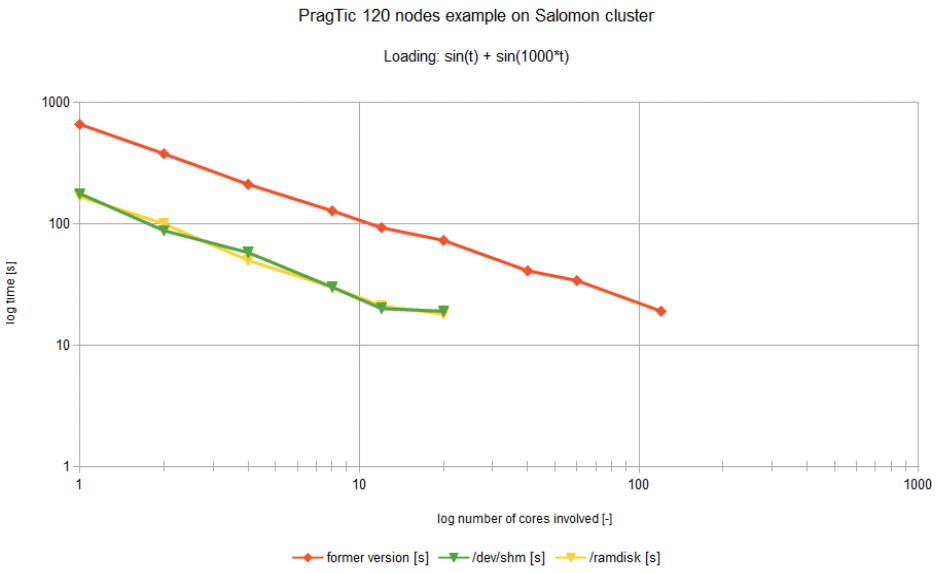


Figure 2.
Run time results of the parallelised PragTic

Conclusion

As can be seen from the results, by the parallelisation of the PragTic we reached speed up 35 on 120 processor cores, which is amazing. It is obvious that a much higher speed up could be reached on a bigger problem with more processor cores.

We also expect much higher speed up if we will be successful with the optimisation of the temporary buffers and with the plane and method parallelisation.

On-going Research / Outlook

The next step is to adjust whole code to keep the temporary buffers in the memory. That will be the hardest part, because the PragTic's source code consists of circa 80,000 lines.

An additional step will be to use this parallelised and optimised version of the PragTic to solve the real problem, including about million nodes.

There is also a possibility to not only parallelise the PragTic software at the node level, but at the plane and/or at the method level as well. This method of parallelisation will probably be even more complicated to implement than the optimisation of the temporary buffers mentioned above.

References

[1] IT4Innovations National Supercomputing Center, VSB-Technical University of Ostrava, Salomon Cluster Documentation – Hardware Overview: https://docs.it4i.cz/salomon-cluster-documentation/hardware-overview, 2016.

[2] PragTic Project. http://www.pragtic.com.

[3] Papuga J.; M. Lutovinov M. Help for FinLiv.VBA Excel Database. [FAD/12/001, ver. B]. FME CTU in Prague and Evektor, spol. s r.o., Prague 2014.

[4] Halama R.; M. Fusek, M. Šofer, M. Poruba, Z. Matušek, P. Fajkoš. R. Ratcheting Behavior of Class C Wheel Steel and Its Prediction by Modified AbdelKarim-Ohno Model. In Proceedings of the 10th International Conference on Contact Mechanics CM2015, Colorado Springs, Colorado, USA, August 30 – September 3, 2015.

[5] PERMON webpages. http://permon.it4i.cz/.

Publication

[6] Papuga J.; R. Halama, M. Fusek, J. Rojíček, F. Fojtík, D. Horák, M. Pecha, J. Tomčala, M. Čermák, V. Hapla, R. Sojka, J. Kružík: Efficient Lifetime Estimation Techniques for General Multiaxial Loading (proceeding ICNAAM 2016 conference, Rhodos 19.-26. 7. 2016)

CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE RESTORATION

Research institution:
Brno University of Technology

Principal investigator:
Michal Hradiš

Researchers:
Pavel Svoboda, Pavel Zemčík, Jan Kotera, Filip Šroubek, David Bařina

Project partner:
Czech Academy of Sciences

Project IDs:
OPEN-7-45, IT4I-9-16

Introduction

In this project, we are investigating if convolutional neural networks are able to learn to directly map degraded images to their high quality versions. That is, if feed-forward convolutional networks are able to take a degraded image as input and produce the corresponding restored image as output. This approach differs significantly from the traditional methods which usually require iterative optimization, are hand-designed for each specific type of degradation, and which require detailed understanding of the degradation process. In contrast, the investigated approach relies on a general and widely used computational model which is optimized in a supervised manner on pairs of high-quality and degraded images.

We have focused on two primary types of image degradations: JPEG compression and motion blur. JPEG is still the most widely used lossy image compression method. It encodes small non-overlapping image patches by quantizing frequency coefficients. Some information is lost in the quantization which results in discontinuities at the patch borders, ringing artifact on edges, and slight blurring of edges (see Figure 1). Motion blur is caused by a non-stationary camera or by objects moving relative to the camera. Motion blur can be modeled by linear convolution with a kernel representing the path of the object during the camera exposure. Blur is usually accompanied by noise created when the photons from the scene are captured and digitized. We consider the general blind deblurring problem, where the blur kernel is not known (blind deconvolution).

Figure 1.
JPEG artifacts: ringing, edge blur, and blocking

Traditional restoration and deblurring methods usually try to find a maximum a posteriori estimate $\hat{\mathbf{x}}$ of the reconstruction [1] by considering image priors $\mathrm{p}(\mathrm{x})$, priors over possible blur kernels $\mathrm{p}(\mathrm{k})$, and a model of possible degradations $\mathrm{p}(\mathrm{y}|\mathrm{x})$:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x},\mathbf{k})\, p(\mathbf{x}) p(\mathbf{k}),$$

where y is the observed image. This problem can be solved as an optimization problem, e.g.:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x} * \mathbf{k} - \mathbf{y}\| + \|\Delta^h \mathbf{x}\|^{\alpha} + \|\Delta^h \mathbf{x}\|^{b} .$$

Our approach estimates the expected value of the reconstruction given the degraded image by directly modelling this relation by convolutional network (CNN):

$$\hat{\mathbf{x}} = \mathrm{E}[p(\mathbf{x}|\mathbf{y})], \qquad \hat{\mathbf{x}} = \mathrm{CNN}(\mathbf{y})$$

Results and Methods

Convolutional networks

The network architectures we use are inspired by the very successful networks that recently redefined state-of-the-art in many computer vision tasks starting with image classification on ImageNet by Krizhevsky et al. [2], and which have since been used for other image processing tasks. The networks are composed of multiple layers of convolutions and element-wise non-linearities, e.g. Rectified Linear Units (ReLU).

The input and output are both 3-channel RGB images with values mapped to the interval [-0.5; 0.5]. Each layer applies convolutions with filters spanning all channels of the previous layer. The last layer is linear (without ReLU). As in previous works [5], we train the networks by minimizing mean squared error on a dataset of corresponding clean and corrupted image patches.

We explore various network architectures and variations including different loss functions, additional network connections, and different types of non-linearities. For example, the residual architecture in Figure 2 which predicts how the input image should be changed is very suitable for tasks where the input and output are highly correlated (e.g. JPEG compression and superresolution) as it does not have to propagate the image through the

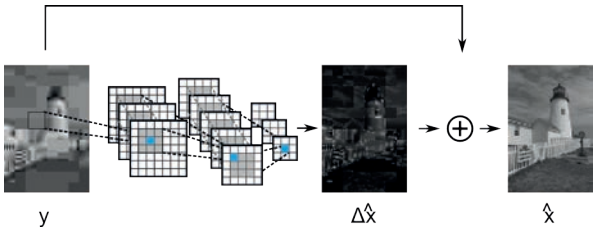


Figure 2.
Residual Convolutional Neural Network for imagerestoration

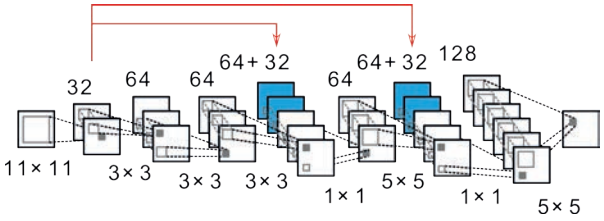


Figure 3.
Network with skip connections (in red)

whole network. Similarly, connections from early network layers to later layers may improve network training and the overall performance (skip architecture in Figure 3).

Text deblurring

Deblurring of text, such as when trying to digitize a printed document with a mobile phone, is a difficult task for the traditional methods as they may not be equipped to handle some of the degradations (e.g. compression, non-uniform blur, high noise, saturation). These methods don't fully exploit the knowledge of the image content in image priors, and they have problems recovering the high amount of small detail needed to retain legibility.

We achieved high quality deblurring results with a large 15-layer network (see Figure 4). This network produces state-of-the-art results with better quality even than non-blind methods which know what blur kernel caused the degradation [1] (measured as signal-to-noise ratio and optical character recognition accuracy). Although the network was trained on rendered PDF documents and artificial blur, it is able to reliably reconstruct real photographs (see Figure 4).

License plate deblurring

Reading of car license plates is an important task in traffic surveillance. We demonstrated that deblurring CNNs provides superior accuracy of a consequent OCR compared to state of the art blind deconvolution methods [1] on images from a real traffic surveillance system (see Figure 5 for examples). Thereby, we, for the first time, quantitatively verified that CNNs provide state-of-the-art results in blind deblurring tasks on real data.

The network from Figure 4 was trained on a large number of high-quality license plate images which were artificially blurred. The performance was improved by restricting the blurs to directions and lengths corresponding to the possible positions of the traffic flow surveillance cameras. The network was trained for 3 days on GTX 980.

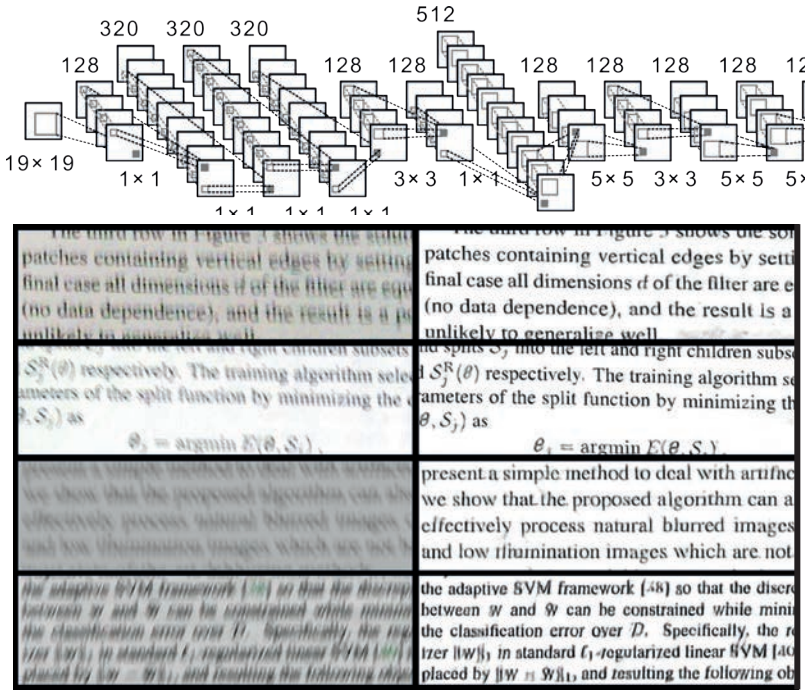


Figure 4.
Text deblurring network consisting of 15 layers and corresponding deblurring results on real photographs

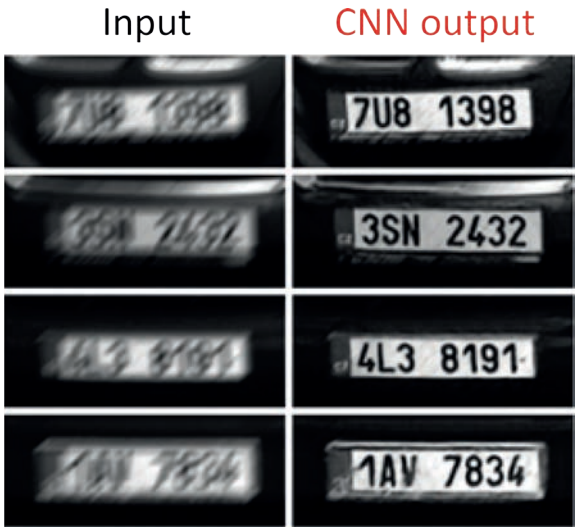


Figure 5.
Result of license plate deblurring on images from real-life traffic surveillance system

Figure 6.
JPEG artifact restoration. Top – original image; Bottom – restored image

JPEG restoration

Deep convolutional networks were unsuccessful in JPEG restoration for long time. For example, Dong et al. [3] reported that they were not able to train networks deeper than 4 layers. We managed to train even 8 layer networks by using proper weight initialization, high learning rates, skip connections (Figure 3) and residual architecture (Figure 2). An example of restoration is shown in Figure 6.

On-going Research / Outlook

We are currently investigating how to combine explicit motion kernel estimation [4] with the feed forward restoration networks into a single end-to-end trainable computational model. We have achieved encouraging preliminary results with motion kernel embeddings and multiplicative connections. In addition we intend to explore the possibility of adapting network weights based on motion estimates.

Conclusion

We have illustrated that direct mapping by convolutional networks is a competitive and practical approach to image restoration. It can be applied with minor changes to a wide range of restoration tasks, often outperforming traditional hand-designed methods.

References

[1] Krizhevsky A., Sutskever I., and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in neural information processing systems 25(2), 2012.
[2] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Doi: 10.1109/CVPR.2014.371
[3] Dong C., Deng Y., Loy C. C., and Tang X. Compression artifacts reduction by a deep convolutional network. In: 2015 IEEE International Conference on Computer Vision (ICCV). Doi: 10.1109/ICCV.2015.73
[4] Sun Jian, et al. Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Doi: 10.1109/CVPR.2015.7298677

Publications

[1] Hradiš Michal, Kotera Jan, Zemčík Pavel and Šroubek Filip. Convolutional Neural Networks for Direct Text Deblurring. In: Proceedings of BMVC, 2015. Doi: 10.5244/C.29.6
[2] Svoboda Pavel, Hradiš Michal, Maršík Lukáš a Zemčík Pavel. CNN for license plate motion deblurring. In: 2016 IEEE International Conference on Image Processing (ICIP). Doi: 10.1109/ICIP.2016.7533077
[3] Svoboda Pavel, Hradiš Michal, Bařina David a Zemčík Pavel. Compression Artifacts Removal Using Convolutional Neural Networks. Journal of WSCG, 2016. Vol. 24, no. 2, p. 63-72. ISSN 1213-6972

Trained networks, data, scripts, and other material can be downloaded from project website: http://www.fit.vutbr.cz/~ihradis/CNN-Deblur/

RENDERING IN BLENDER CYCLES USING MPI AND INTEL® XEON PHI™

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Milan Jaroš

Researchers:
Lubomír Říha, Daniel Krpelík

Project ID:
IT4I-9-9

Introduction

The purpose of rendering is to create visually realistic 2D images of 3D scenes created in 3D modelling environments. Due to the complex nature of light, propagation rendering is a computationally expensive task. Rendering is used to generate images, develop animated movies or computer games.

There are multiple software solutions for both scene modelling and rendering. We have focused in our development on the open source 3D creation suite Blender. Blender has a wide user base and gives user a vast possibility of extensions via C++ and Python programming. Blender includes a 3D modelling environment and two original renderers – Blender internal and Blender cycle, which can provide rendering in photorealistic quality, see Figure 1. These renderers are implemented to utilise multi-core CPU systems and GPU acceleration.

Figure 1.
Rendered worm scene using Cycles renderer

There is other rendering software available, but not many of them offer distributed rendering suitable for the HPC environment. Furthermore, almost none of the HPC rendering software is open source. An example of commercial product supporting utilisation of a graphic accelerator is Octane Render from OTOY or Iray directly from NVIDIA. There are photorealistic ray tracing kernels from Intel too, called Embree, which support Intel Xeon processors and also Intel Xeon Phi coprocessors. The example of a commercial product based on Embree is the Corona render, which only runs on CPUs. There is also an open source engine based on Embree. It is called OSPRay and it is meant for high-performance, high-fidelity visualisations on Intel CPUs and runs on anything from laptops or workstations to compute nodes in HPC systems.

Within the project we have developed an extension to the Blender code which can utilise Intel Xeon Phi coprocessors with many integrated core (MIC) architecture for stand-alone and also cluster computer setups. MIC architecture can be programmed using both shared memory models such as OpenMP or OpenCL (OpenCL provides compatibility with codes developed for GPU) and distributed memory models, such as MPI.

The implementation we provide has been developed and tested on the Salomon supercomputer [1].

Results and Methods

Rendering consist of creation of images from virtual 3D scenes. In 1986 Kajiya first introduced the rendering equation in computer graphics [2]. The rendering equation provides the basis for ray tracing methods like Path-tracing. Solving the rendering equation is computationally extensive. The most common method of solving the equation is to estimate it by means of the Monte Carlo (MC) simulation. In the case of Path-tracing algorithm, it integrates possible paths of light from the light source into selected pixels in the rendered image. Using MC, we select random subpixels and then track a light ray bouncing randomly from object surfaces.

The cycles rendering engine in Blender is based on the Path-tracing algorithm and uses a Quasi-Monte Carlo method [3,4] to increase convergence speed. It supports both interactive and offline rendering. The interactive rendering provides online changes of a scene, which offers users possibilities to change materials, the position of a light source, object geometry, etc. during the rendering process. There is no need to manually restart the renderer as opposed to the offline renderer, which is therefore mainly used for rendering of the final scenes.

The rendered scene in Blender is stored in KernelData and KernelTextures structures. At first the elementary information about

the scene (for instance the camera position or background information) is sent to the selected computing devices using the const_copy_to method. After that the objects and textures are saved to the memory of computing devices using the tex_alloc method. In addition, we have to allocate a buffer for rendered pixels and initialise the buffer for the random number generator state, the rng_state. The next step is to decompose the rendered image, or task. Each thread calls on the path_tracing method to calculate the render equation on the assigned part of the job using the thread_run method. The results are then stored to the buffer vector. The main data distribution functions are depicted in Figure 2.

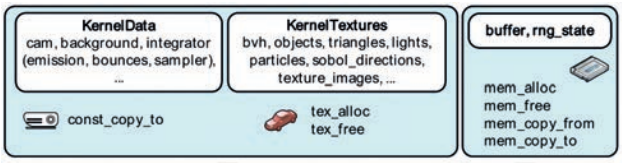


Figure 2.
Blender internal data manipulation functions for rendering

CPU only parallelisation

Possible alternation of the original Blender render code may be acquired by simply altering the thread driver technology. Instead of using original POSIX threads and job decomposition into tiles, we have implemented OpenMP based parallel code. In this case the job is not decomposed into smaller tiles, but rather parallel threads split the job by pixels – one at the time. This alternation leads to a slight increase of efficiency as shown in Figure 5.

Intel Xeon Phi utilisation

For utilisation of the symmetric mode of Intel Xeon Phi a "Blender Client" has been created to run on slave machines and to wait for the jobs given by the master machine running the modified version of Blender. The "Blender Client" requires being compiled and run on selected slaves.

The offload mode of Intel Xeon Phi can be accessed from both altered Blender and "Blender Client". Due to differences in CPU and MIC, special directives must be implemented to ensure proper communication between the devices. CPU code is divided into two main tasks. One task handles CPU/MIC communication and the other task deals with CPU rendering itself.

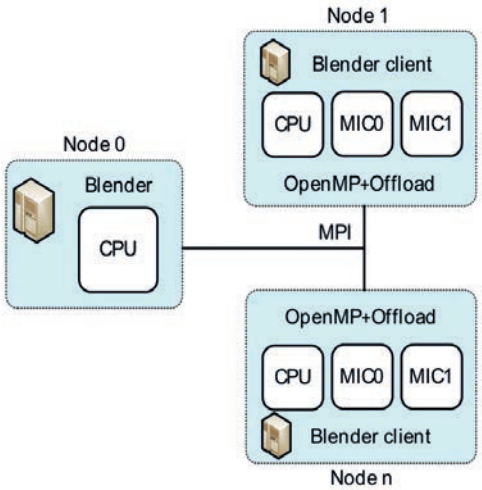


Figure 3.
Rendering on HPC cluster, Offload mode

Cluster parallelisation

Cluster parallelisation is achieved via MPI technology, where part of the node or whole node acts as the master running the modified version of Blender and the rest of the nodes act as slaves running "Blender Client". Our implementation provides the possibility to utilise Intel Xeon Phi accelerators in both symmetric and offload mode as shown in Figure 3 and Figure 4.

A series of tests were performed on two different 3D scenes (Tatra, worm) to measure rendering time speed-up. The first test compared the rendering times of various parallelisation schemes on a single node. Namely, this was original POSIX parallelisation (CPU24) on 24 CPU cores, our OpenMP (OMP24) version on 24 CPU cores, original GPU implementation on 2x NVIDIA GeForce GTX 970 (GPU2) and also our implementation on 2x Intel Xeon Phi running in the offload mode (MIC2). Results are shown in Figure 5.

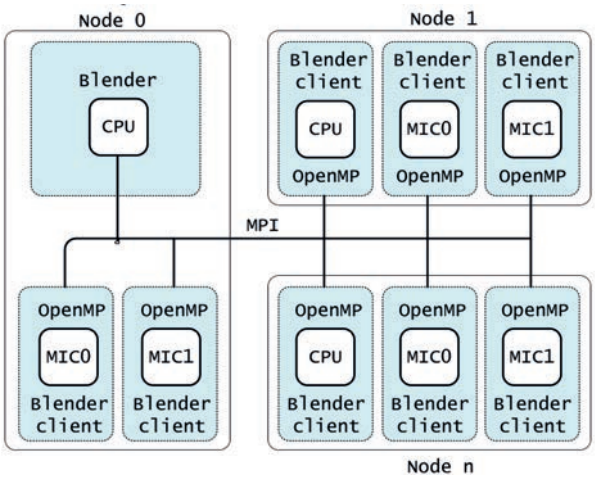


Figure 4.
Rendering on HPC cluster, Symmetric mode

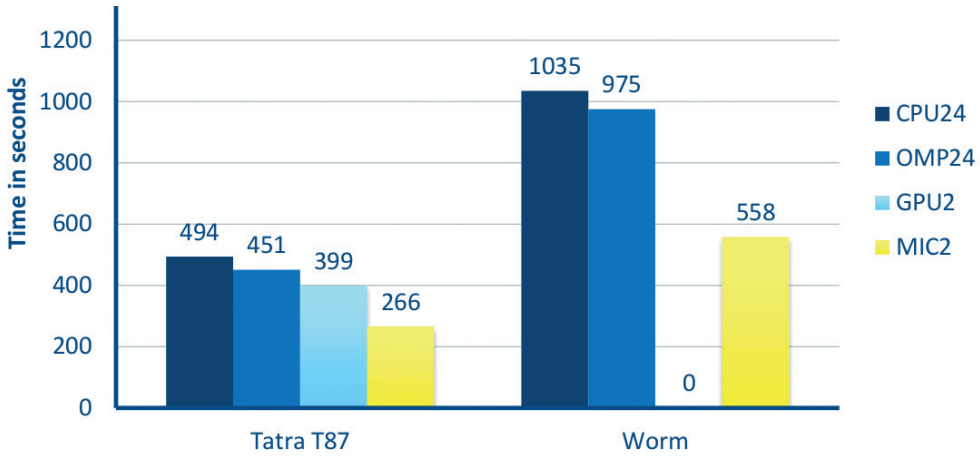


Figure 5. Rendering time for different scenes and parallelisation techniques on one node

Another test measured the strong scalability of the offline rendering on multiple computing nodes. The time necessary to render the final scene was measured using different parallelisation techniques and an increasing number of nodes. Results for the worm scene are depicted in Figure 6.

Finally, the interactive rendering capabilities were tested. The time necessary to render the scene preview for an increasing number of nodes (strong scaling performance) was measured up to the point where interactive rendering (24pfs) is reached. Further nodes were then utilised to reach higher preview quality while retaining interactive rendering (weak scaling performance). These results can be seen in Figure 7.

On-going Research / Outlook

We plan to further increase the efficiency of the modified Blender's renderer by improving its vectorisation. Other improvements in terms of reduction of rendering time lay in application of image filtering methods, which are already being implemented.

Conclusion

We have developed a modified version of Blender's rendering system that is capable of utilising Intel Xeon Phi accelerators. The resulting software is able to share the workload between CPU and additional MIC architectures very efficiently for both type of rendering tasks (interactive, offline). It is superior to GPU acceleration by allowing us to exploit the same functionality as on CPU-only systems (advanced shaders, etc.). Our software [5] can fully utilise HPC resources, as was tested on infrastructure at IT4Innovations.

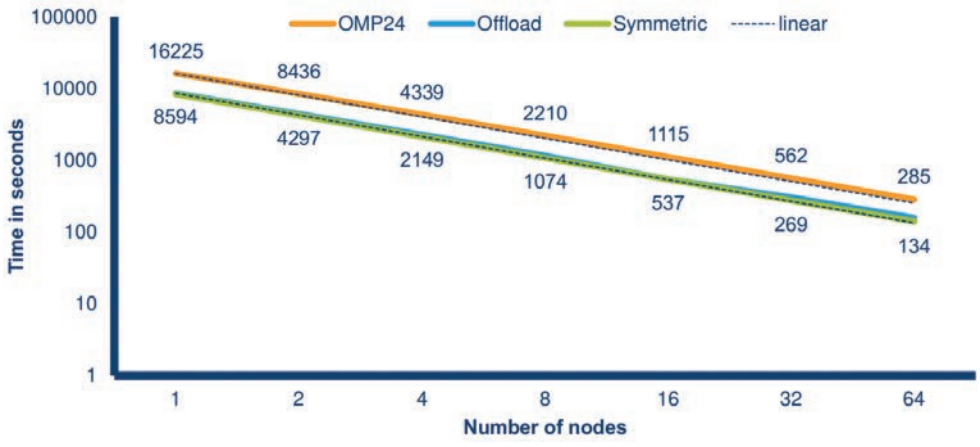


Figure 6. Strong scaling of the renderer using different parallelisation techniques

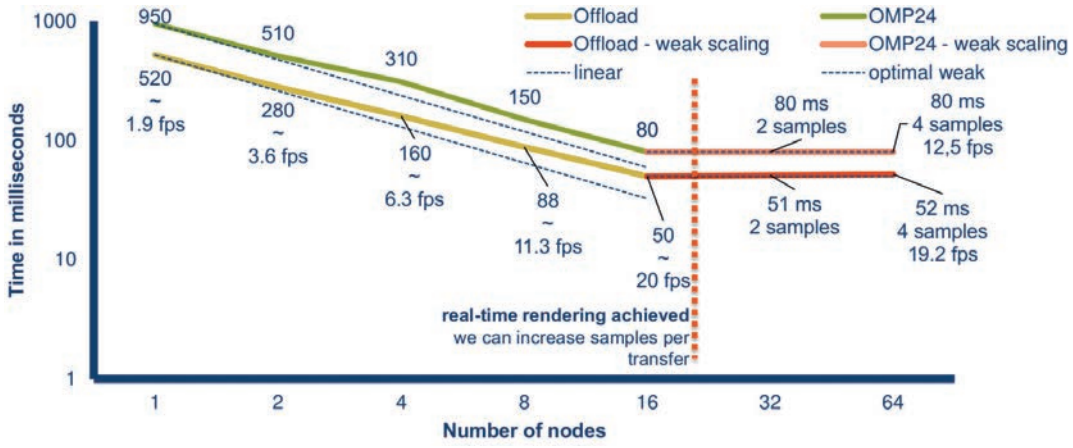


Figure 7.
Strong/weak scaling of the interactive rendering

References

[1] Salomon hardware overview. Available at: https://docs.it4i.cz/salomon/hardware-overview/
[2] Kajiya, J. The rendering equation. Computer Graphics. 1986, vol. 20, pp. 143-150, August 1986. Doi: 10.1145/15886.15902
[3] Niederreiter, H. Random number Generation and quasi-Monte Carlo Methods. SIAM, Philadelphia, ISBN 0-89871-295-5, 1992
[4] Morokoff, W.J. Generating quasi-Random Paths for Stochastic Processes, Working Paper, Mathematics Dept. of UCLA, 1997

Publications

[1] Jaroš M. The Fundamentals: How to accelerate Blender with the Intel Xeon Phi coprocessors, IXPUG/EMEA 2016
[2] Jaroš M.; P. Strakoš, and T. Karásek. Implementation of bm3d filter on intel® xeon phi™ for rendering in blender cycles. AMC2016. Jimp – Submitted.
[3] Jaroš M. and L. Říha. Interactive and offline rendering in Blender Cycles using MPI and Intel Xeon Phi Offload, SC2016
[4] Jaroš, M. and Říha, L. and Strakoš, P. and Karásek, T. and Vašatová, A. and Jarošová, M. and T. Kozubek. Acceleration of the blender cycles path-tracing engine using Intels' many integrated core architecture. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015, 9339, 86-97. Doi: 10.1007/978-3-319-24369-6_7

EVALUATION OF PARALLEL I/O STRATEGIES FOR DISTRIBUTED SPARSE MATRICES

Research institution:
Czech Technical University in Prague

Principal investigator:
Daniel Langr

Researcher:
Ivan Šimeček

Project ID:
OPEN-1-12

Introduction

Sparse matrices are commonly stored in a computer memory in storage formats that provide high performance/efficiency of related computations, e.g., the matrix-vector multiplication operation. However, such formats are generally not space-optimal. If we want to store a sparse matrix to a file system, we then have two options: either to store the matrix in its in-memory storage format, or to store it in some space-efficient storage format, which additionally requires executing a corresponding conversion algorithm.

The goal of this project was to compare these two options on the Anselm supercomputer. As an in-memory storage format, we chose the compressed sparse row (CSR) format, which is likely the most commonly used format for storing sparse matrices in memory in practice. As a space-efficient storage format, we took our adaptive-blocking hierarchical storage format (ABHSF) developed especially for the purpose of storing sparse matrices to file systems [1]. Accordingly, we developed an efficient conversion algorithm from CSR to the ABHSF [2].

In a distributed-memory runtime environment, such as on the Anselm supercomputer, parallel I/O must be exploited to provide fast file operations performed by multiple application processes. Such parallel I/O is provided by the MPI-IO parts of MPI library implementations. We accessed MPI-IO indirectly through the HDF5 library, namely its parallel I/O capability (pHDF5). HDF5 considerably simplifies parallel writing/reading of information into/from files stored in parallel file systems, such as Lustre in our case of the Anselm supercomputer.

Minimisation of storage time was not the only objective of the presented project. Possibly the most typical application of the problem of storing large sparse matrices to file systems is the implementation of application resiliency to system failures via checkpointing/restart techniques. In such cases, we want to accomplish the matrix storage process in parallel with running computations. The main goal is then not only a fast storage process, but also its low influence on the concurrently running computation. Moreover, parallel file systems of large supercomputers are shared resources used by many applications at once. Therefore, the I/O bandwidth stress to a file system was also a metric of our concern in the experiments.

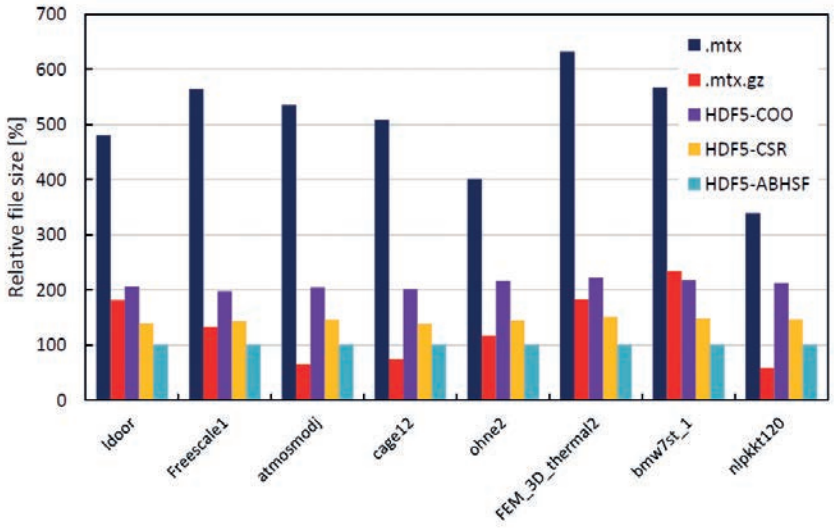


Figure 1.
File sizes relative to HDF5-ABHSF for benchmark matrices and different file/storage formats

Results and Methods

Figure 1 shows the comparison of file sizes with several stored matrices when using different file/storage formats. The benchmark matrices were taken from the University of Florida sparse matrix collection (UFSMC) and were chosen to emerge from different application domains, and thus having different structural and numerical characteristics. We may observe that matrices stored in the CSR format took considerably more file system space than matrices stored in the ABHSF, which supports our assumption of reducing storage time and bandwidth by reduction of matrix memory / file system footprints.
To measure scalability, we used our method of scalable parallel generation of benchmark sparse matrices [3]. Overlapping of computations and storage processes was implemented by using multiple threads via OpenMP. To evaluate the influence of the matrix storage process to running computations and vice versa, each experiment was performed with the following options:

- I/O only: matrix storage only and turned-off computation (inactive working thread);
- computation only: computation only and turned-off matrix storage (inactive I/O thread);
- I/O+computation: both computation and matrix storage (active both threads).

As for results, we set up the following measured metrics:

- Storage time – the time needed for the I/O thread to store the tested matrix into files. In case of the ABHSF, the storage times includes the on-the-fly conversion of the matrix from the CSR format.
- Storage size – the effective number of bytes written to files. This is the number of bytes containing useful information, thus without the overhead of the HDF5 library (which is negligible for larger files).
- Computational performance – the floating-point performance of the working thread. In case of computation only, it was measured for 100 iterations of the iterative process that involved sparse matrix-vector multiplication and normalisation of resulting vectors by their maximum elements. In case of I/O+computation, it was measured for the number of iterations that overlapped the matrix storage process.
- Effective bandwidth – the measured bandwidth of the utilised file system, calculated as the storage size divided by the storage time.

We show here the measured results for the Anselm supercomputer and the matrix generator seeded with the heart1 sparse matrix [3]. All the results were measured as a function of the number of processors (MPI processes), which always corre-

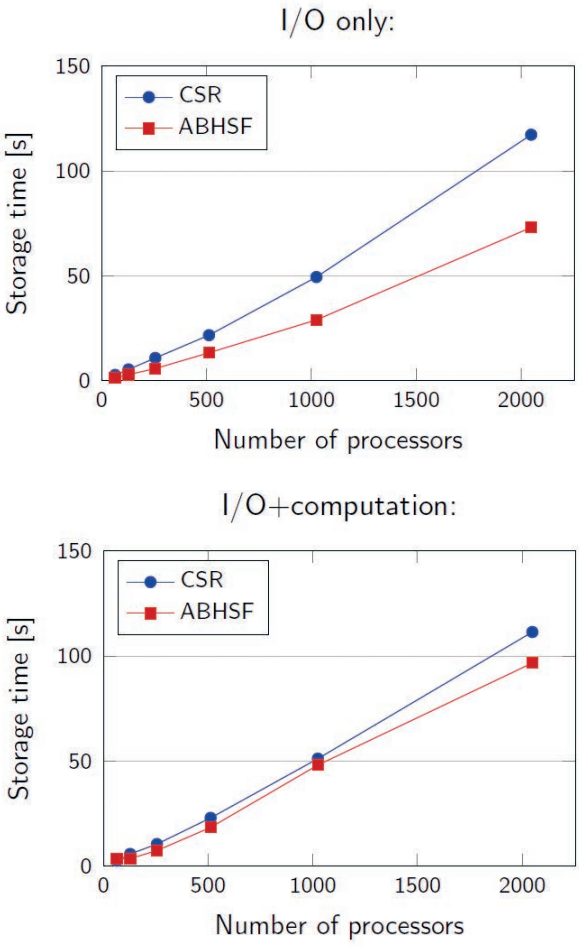


Figure 2.
Measured storage times in the I/O only (above) and I/O+computation (below) cases

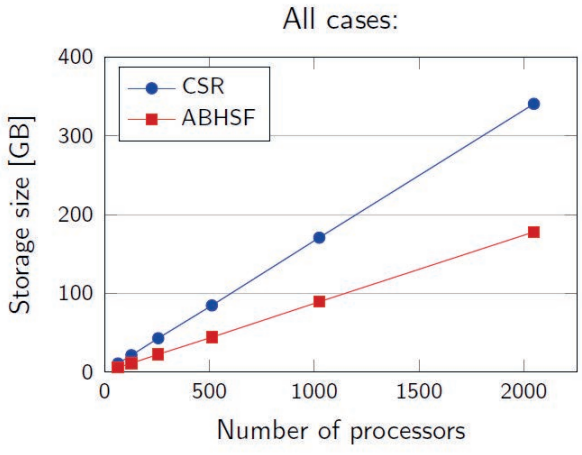


Figure 3.
Size of data written to a file system

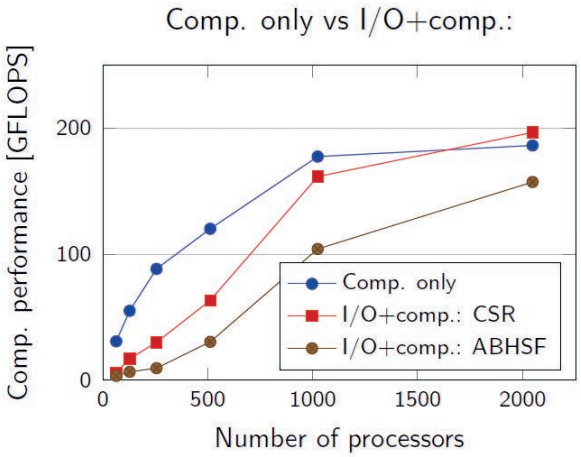


Figure 4.
Measured computational performance

sponded with the number of utilised CPU cores. The upper part of Figure 2 shows the storage time in case of pure I/O (no concurrent computation). The storage times measured in the case of concurrent I/O and computation are shown in the lower part of Figure 2. Figure 3 shows the total amount of stored data into a file system. Figure 4 evaluates the effect of the storage process to the concurrently running computation with the same matrices (iterative process described above). Finally, Figure 5 shows the measured effective bandwidth for both I/O only and I/O+computation cases.

On-going Research / Outlook

Reducing memory footprints of sparse matrices might be attractive not only for their storage to a file system, but for generic matrix computations as well. The performance of iterative computations based on sparse-matrix vector multiplication is known to be limited by memory bandwidth. In the outgoing and future research, we are therefore focusing on the utilisation of developed space-efficient formats also for generic sparse matrix computations, such as sparse iterative linear solvers or eigensolvers.

Conclusion

The experiments conducted within this project (for more results see [4]) imply that when large sparse matrices stored in memory

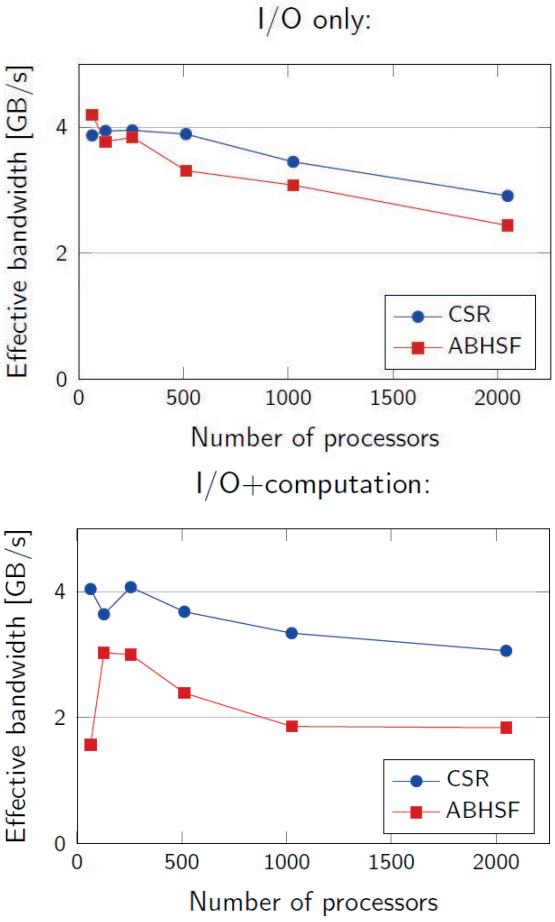


Figure 5.
Measured effective file system bandwidth in the I/O only (above) and I/O+computation (below) cases

in CSR need to be saved to a file system, it is more beneficial to choose the ABHSF storage format for this task; it provided shorter storage times in almost all experiments. The impact on simultaneously running computations were lower for CSR in some cases and for ABHSF in others. However, differences were mostly only small. Moreover, ABHSF either resulted in higher effective file system bandwidth together with considerably lower storage times, or in lower bandwidth, which is generally a wanted feature because of sharing a file system with other system users.

References

[1] Langr, D.; I. Šimeček, P. Tvrdík, T. Dytrych, J.P. Draayer. Adaptive-Blocking Hierarchical Storage Format for Sparse Matrices. In Proceedings of the Federated Conference on Computer Science and Information Systems. 2012, pp 545-551. ISBN 978-83-60810-48-4

[2] Langr, D.; I. Šimeček, P. Tvrdík. Storing sparse matrices to files in the adaptive-blocking hierarchical storage format. In proceedings of the Federated Conference on Computer Science and Information Systems. 2013, 479-486. ISBN 978-1-4673-4471-5

[3] Langr, D.; I. Šimeček, P. Tvrdík, T. Dytrych. Scalable Parallel Generation of Very Large Sparse Benchmark Matrices. In proceedings of the 10th International Conference on Parallel Processing and Applied Mathematics. Lecture Notes in Comp. Science. 2014, 8384, 178-187. ISBN 978-3-642-55224-3

Publication

[4] Langr, D. Algorithms and Data Structures for Very Large Sparse Matrices. Ph.D. thesis, Czech Technical University in Prague, 2015.

PARALLEL BOUNDARY ELEMENT METHOD FOR SOUND-HARD SCATTERING

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Michal Merta

Researchers:
Jan Zapletal, Lukáš Malý, Michal Kravčenko

Project IDs:
IT4I-9-12, Open-7-13

Introduction

The boundary element method (BEM) is a suitable tool for simulation of processes on unbounded domains, such as sound or electromagnetic wave propagation. Solution of these problems may find an application in various areas of engineering, science, and medicine, starting from non-destructive testing, to seismic wave propagation modelling, to ultrasonic imaging. While the modelling of time-harmonic wave propagation (described by the Helmholtz equation) by BEM is now well established, efficient solution of transient problems (described by the time-dependent wave equation) still needs to be investigated.

BEM4I is a library of parallel BEM-based solvers aiming at high performance environments. It has been under development at the IT4Innovations National Supercomputing Center since 2013 and currently includes modules for the solution of the Laplace, Lamé, Helmholtz, and time-dependent wave equations in 3D. To leverage the state-of-the-art technology available at the infrastructure of IT4Innovations, the library provides SIMD-vectorised functions [4], OpenMP parallelisation in shared memory, and MPI parallelisation in distributed memory [8]. The library structure is depicted in Figure 2. Computationally intensive parts of the code can be offloaded to the Intel Xeon Phi coprocessors [5,7]. Moreover, because of the interface to the Espreso domain decomposition library (espreso.it4i.cz), the BETI (boundary element tearing and interconnecting) domain decomposition method can be used for the solution of large scale problems or problems with material heterogeneities.

The library includes a module for massively parallel solution of the time dependent wave scattering problems described by the wave equation [6]. To overcome problems with numerical integration arising when using classical polynomial temporal basis functions, we use a novel approach with smooth temporal basis functions introduced by Sauter and Veit [2]. Since the systems of linear equations produced by this method are global in space and time and therefore of a large dimension, an efficient parallelisation of the method is necessary in order to enable the solution of large scale problems.

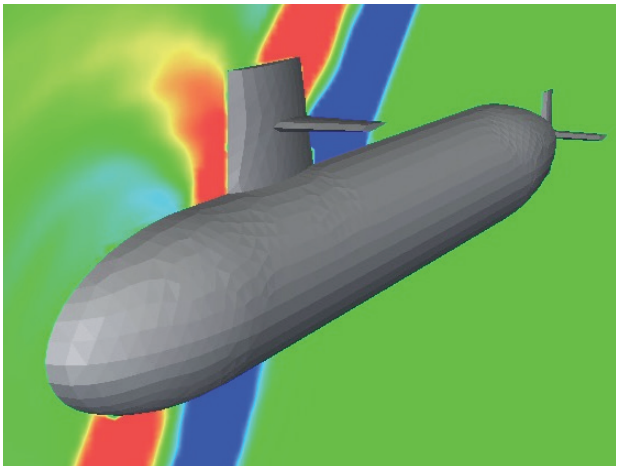


Figure 1.
Scattering off a submarine

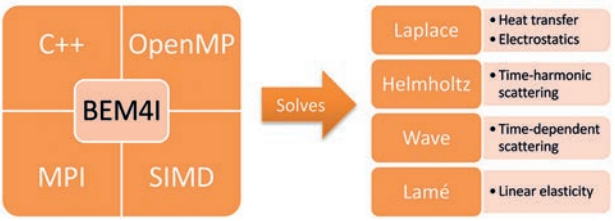


Figure 2.
Structure of the BEM4I library

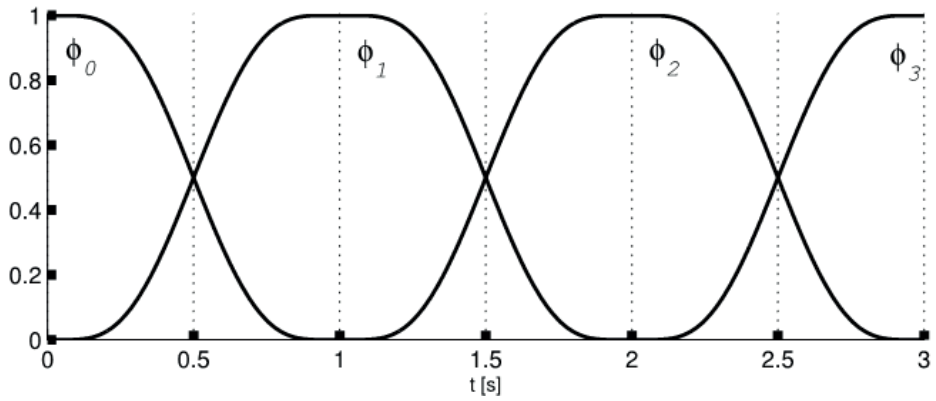


Figure 3.
Partition of unity functions

Results and Methods

The main advantage of the boundary element method in comparison to other methods used in mathematical modelling (such as the finite element or finite volume method) is the absence of the necessity to discretise the interior of a computational domain. This not only leads to a reduction of the number of variables, but also to easier mesh generation. However, the main drawback in the case of stationary problems is the quadratic complexity with respect to the computational and memory requirements, since the resulting system matrices are dense. Moreover, special quadrature routines are necessary due to the singularities occurring in integrals over surface elements.

When implementing the boundary element method for the solution of a transient wave equation in 3D, one has to deal with several additional challenges. Although the resulting system matrices are sparse, they are global in space and time and of the order MN(p+1), where M is the number of surface nodes, N is the number of time-steps, and p is the order of temporal basis functions. Moreover, the block-Hessenberg structure of the matrix does not allow for the solution using marching-on-time algorithms. Therefore, an iterative solver together with a suitable preconditioner has to be employed.

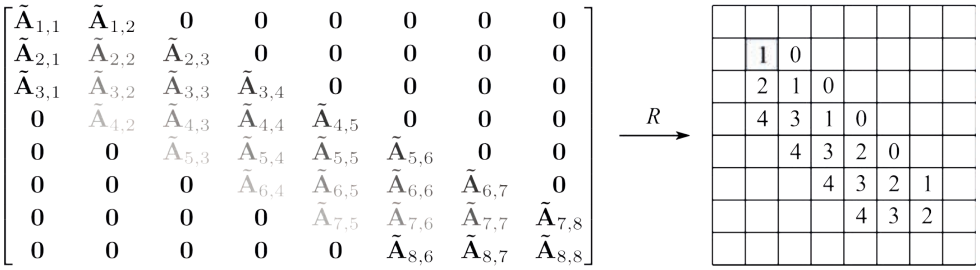


Figure 4.
Structure of the system matrix and its distribution. The blocks printed in the same shade of grey are identical.

The main problem during the system matrix assembly is an efficient accurate quadrature over the pairs of surface elements. When using classical piece-wise polynomial basis functions, one has to integrate over an element and the intersection of another element with a discrete light cone to obtain optimal convergence of the numerical quadrature. To avoid this integration over complicated shapes we use a new approach utilising smooth temporal basis functions based on the partition of unity method [2], see Figure 3. However, the evaluation of these functions is costly, which even further increases the computational demands of the method and makes its parallelisation necessary. The parallelisation of the method consists of the distribution of the non-zero matrix blocks among available MPI processes taking into account a special duplicating structure of the matrix blocks (see Figure 4). Within each MPI process the assembly is parallelised by OpenMP in shared memory.

To solve the resulting system of linear equations an iterative solver with a suitable preconditioner has to be used. In Table 1 we compare our implementations of GMRES, DGMRES, and FGMRES [3] in combination with a new algebraic recursive preconditioner with respect to the number of time-steps. See [6] for a detailed analysis.

	GMRES		DGMRES		FGMRES	
N	# it	t [s]	# it	t [s]	# it	t [s]
15	4021	44.5	580	8.5	51	7.3
20	5448	99.0	510	14.9	48	9.7

Table 1.
Comparison of iterative solvers

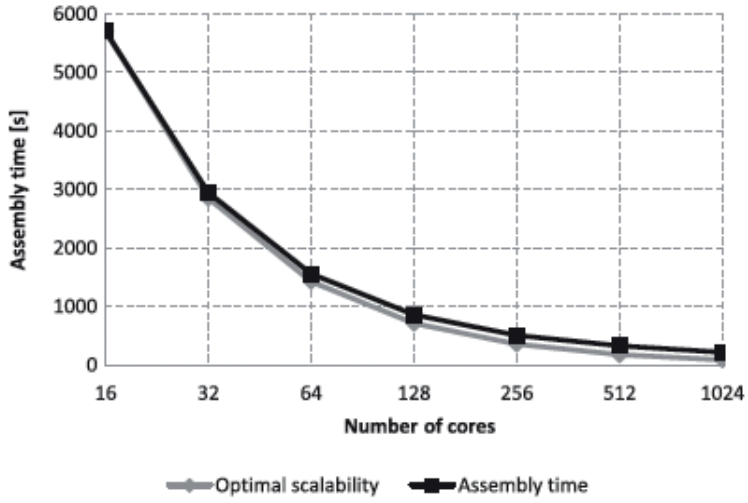


Figure 5.
Scalability of the system atrix assembly

The scalability of the system matrix assembly was tested on a submarine-shaped domain (see Figure 1) and is depicted in Figure 5. The test was carried out using up to 64 nodes of the Anselm cluster, each equipped with two eight-core Intel Xeon E5-2665 Sandy Bridge processors and 64 GB of RAM.

On-going Research / Outlook

We plan to further develop and optimise the solver for time-dependent problems, and to establish cooperation with the Graz University of Technology in this field. Development of a suitable parallel preconditioner is also necessary in order to be able to solve large scale real world problems.

Conclusion

Solution of time-dependent problems using the space-time boundary element method has recently been gaining in popularity. The method produces very large matrices which are global in space and time. While this may be from one point of view considered a disadvantage, from another point of view it enables the introduction of another level of parallelism in the temporal dimension, which would not be possible when using classical marching-on-time schemes. The additional parallelism leads to a more efficient utilisation of modern peta- and upcoming exa-scale supercomputers. Current development and testing of the solver at the infrastructure of the IT4Innovations National Supercomputing Center show the significant potential of the solver.

References

[1] Rjasanow Sergej and Steinbach Olaf. The fast solution of boundary integral equations. Springer 2007. ISBN: 978-0-387-34041-8.
[2] Sauter Stefan and Veit Alexander. A Galerkin method for retarded boundary integral equations with smooth and compactly supported temporal basis functions. Numerische Mathematik, 2013, 123, pp. 145-176. ISSN: 0029599X.
[3] Saad Yousif. A flexible inner-outer preconditioned GMRES algorithm. SIAM Journal on Scientific Computin, 1993, 14, pp. 461-469. ISSN: 1064-8275

Publications

[4] Merta Michal and Zapletal Jan. Acceleration of boundary element method by explicit vectorisation. Advances in Engineering Software, 2015, 86, pp. 70-79. ISSN: 0965-9978.
[5] Merta Michal, Zapletal Jan and Jaros Jiri. Many core acceleration of the boundary element method. High Performance Computing in Science and Engineering: Second International Conference, HPCSE 2015, Soláň, Czech Republic, pp. 116-125. Springer International Publishing, ISBN: 978-3-319-40361-8.
[6] Veit Alexander, Merta Michal, Zapletal Jan and Lukáš Dalibor. Efficient solution of time-domain boundary integral equations arising in sound-hard scattering. International Journal for Numerical Methods in Engineering, 2016, 107, pp. 430-449. ISSN: 1097-0207.
[7] Zapletal Jan, Merta Michal and Malý Lukáš. Boundary element quadrature schemes for multi- and many-core architectures. Submitted.
[8] Lukáš Dalibor, Kovář Petr, Kovářová Tereza and Merta Michal. A parallel fast boundary element method using cyclic graph decompositions. Numerical Algorithms, 2015, 70, 807-824. Doi: 10.1007/s11075-015-9974-9

Project website: bem4i.it4i.cz

DEVELOPMENT OF A PARALLEL ALGORITHM FOR FUZZY ASSOCIATION RULES MINING

Research institution:
Institute for Research and Applications of Fuzzy Modelling, University of Ostrava

Principal investigator:
Viktor Pavliska

Researchers:
Radek Valášek, Michal Burda, Pavel Rusnok

Project ID:
IT4I-6-6

Introduction

Searching for association rules [1-3] is a broadly accepted data mining technique. An association rule is commonly understood as an expression X -> Y, where antecedent X and consequent Y are conditions in the form of elementary conjunctions, and the rule is usually interpreted as "if X is satisfied then Y is also very often true".

There exist many algorithms to obtain such rules automatically from data. The usefulness of association rules is multitudinous: for human-readable description of relationships in data, for creating classification or regression models, for prediction, etc.

Searching for association rules is a very computational power demanding task. Although various algorithms exist, all of them are of exponential time complexity. The task of searching for association rules fits particularly well on binary or categorical data and many has been written on that topic [2-5].

For association analysis on numeric data, a prior discretisation is proposed by Srikant et al. [6]. Unfortunately, that may lead to danger of undiscovering important knowledge due to information loss caused by discretisation. The problem of information loss due to discretisation is tightly connected to the fact that a quantitative attribute is transformed to several categories given by crisp boundaries. A rational request to "soften" these boundaries leads us to the fuzzy sets theory. The use of fuzzy sets in connection with association rules has been motivated by many authors (see [7] for recent overview). By allowing for soft boundaries of discretisation intervals, fuzzy sets can avoid certain undesirable threshold effects [8]. Fuzzy association rules are also appealing because of the use of vague linguistic terms such as "small", "very big", etc. [9],[10]. Also see [11].

Association rules mining, especially of fuzzy rules where the amount of attributes multiplies by using different linguistic terms, may be a very computational time demanding task. Therefore, substantial research effort is devoted to the design of parallel algorithms.

Parallel algorithms for association rules primarily focus on processing large, but sparse categorical data sets. That data often does not fit into RAM, therefore the reading of data from the disk is the most time consuming task. The approaches for parallelisation are, e.g., count distribution [12], where each processor computes counts of appearance of candidates in a locally stored data set, which are then combined to obtain global results. The data distribution algorithm [12] partitions candidate item sets among the processors. Also combinations of these approaches exist [13]. All these approaches are based on the candidate-generation technique developed originally in the Apriori algorithm [3]. Other approaches are based on the FP-growth mining algorithm [14], which avoids generating candidate sets. Hu and Yang-Li [15] propose FP-Forests.

Results and Methods

This project was focused on a novel implementation of fuzzy association rules mining algorithm that is based on OPUS search [1,16] that is suitable for dense data sets. We develop a parallel version of OPUS, implement it as a program suitable to run on a high-performance computing cluster with the Open MPI parallelisation environment [17], and evaluate its scalability on real data. The parallelised version of the OPUS search algorithm is based on the master/slave architecture with the master node handling the queue of search tasks and distributing the workload to the slave nodes. The performance of the algorithm was tested on several datasets from the UCI Machine Learning Repository [18]. Each of the numeric columns was transformed to 7 fuzzy sets by using linguistic expressions [19].

Within this project, a large experiment with practical results was also performed. We have used the implemented algorithm to search for rules on a data set that contains information about several statistical properties of a set of time series together with the quality of their forecasts computed by various forecasting methods. A large experiment optimising various parameters of the model was performed in order to find an optimal setting of the parameters. As a result, we have obtained a rule base that helps for selection of an appropriate prediction method in

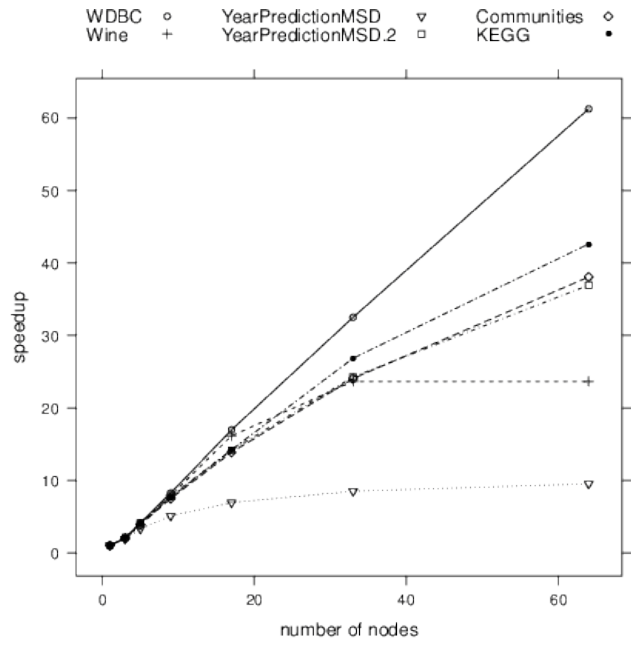


Figure 1.
Scalability of the parallel algorithm. Axes of the graph are logarithmic. Note that the running time nearly halves if doubling the number of nodes being used, which indicates excellent scalability.

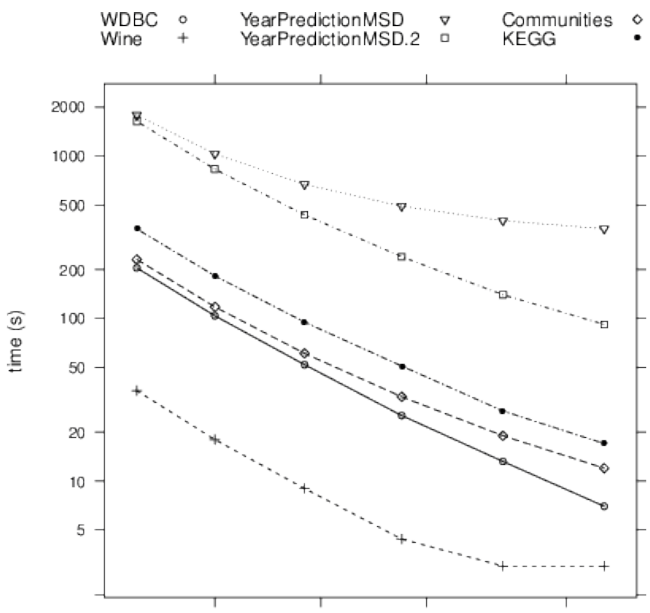


Figure 2.
Speedup of a parallel algorithm, i.e. a ratio of the run time of a sequential algorithm to the run time of the parallel version running on a given number of nodes. Linear speedup (i.e. speedup equal to the number of nodes) is "ideal".

time series forecasting. The rule base is now a part of the Fuzzy Rule-based Ensemble (FRBE) of forecasts, a general tool for time series forecasting that may be used in many different areas such as economy, health, computer science and other.

On-going Research / Outlook

Performance analysis shows very good scalability of the algorithm. The algorithm also provides the means for adjusting the load-balancing by optimally setting the recursion threshold, whose value was only roughly estimated in the presented performance analysis. Future work will therefore address elaboration of better load balancing by dynamically changing the recursion threshold based, e.g., on the task queue size, amount of available slaves, average time to compute the task, etc.

Conclusion

A novel parallel algorithm for mining of fuzzy association rules was developed and performance-tested. It was also used to create a rule-base that can be used to ensemble time series forecasts. As a result, we have obtained a rule base that helps the selection of an appropriate prediction method in time series forecasting. The rule base is now a part of the Fuzzy Rule-based Ensemble (FRBE) of forecasts, a general tool for time series forecasting that may be used in many different areas such as economy, health, computer science and others. The FRBE is implemented as a package for the R statistical environment, its source code is open, and freely available for a wider audience. The superiority of the method was reported in several papers [20-23].

References

[1] Webb, G. I. Opus: An efficient admissible algorithm for unordered search. Journal of Artificial intelligence Research. 1995, 3, pp. 431-465.
[2] Hájek, P.; Havel, I., and Chytil M. "The GUHA method of automatic hypotheses determination," in Computing 1, 1966, pp. 293-308.
[3] R. Agrawal, T. Imielinski, and A. Swami, "Mining associations between sets of items in massive databases," in ACM SIGMOD 1993 Int. Conference on Management of Data, Washington D.C., 1993, pp. 207-216.
[4] A. Berrado and G. C. Runger, "Using metarules to organise and group discovered association rules," Data Min. Knowl. Discov. 2007, vol. 14, no. 3, 409–431. Doi: 10.1007/s10618-006-0062-6
[5] G. I. Webb, "Discovering significant patterns," Mach. Learn., vol. 68, no. 1, pp. 1-33, 2007. Doi: 10.1007/s10994-007-5006-x
[6] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," SIGMOD Rec., vol. 25, no. 2, pp. 1-12, 1996. Doi: 10.1145/233269.233311
[7] H. Kalia, S. Dehuri, and A. Ghosh, "A survey on fuzzy association rule mining," International Journal of Data Warehousing and Mining (IJDWM), vol. 9, no. 1, pp. 1-27, 2013. Doi: 10.4018/jdwm.2013010101
[8] T. Sudkamp, "Examples, counterexamples, and measuring fuzzy associations," Fuzzy Sets and Systems, vol. 149, no. 1, pp. 57-71, 2005. Doi:10.1016/j.fss.2004.07.017
[9] K. C. Chan and W.-H. Au, "Mining fuzzy association rules," 1997.
[10] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, and P. Yan, "Mining pure linguistic associations from numerical data," Int. J. Approx. Reasoning, vol. 48, no. 1, pp. 4-22, Apr. 2008. Doi:10.1016/j.ijar.2007.06.005
[11] M. Burda, "Fast evaluation of t-norms for fuzzy association rules mining," in the 14th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2013). Budapest: IEEE, 2013, pp. 465-470. Doi: 10.1109/CINTI.2013.6705242
[12] R. Agrawal and J. C. Shafer, "Parallel mining of association rules," IEEE Transactions on Knowledge and Data Engineering, vol. 8, pp. 962-969, 1996. Doi: 10.1109/69.553164
[13] E.-H. Han, G. Karypis, and V. Kumar, "Scalable parallel data mining for association rules," IEEE Trans. Knowl. Data Eng., vol. 12, no. 3, pp. 377-352, 2000. Doi: 10.1109/69.846289
[14] O. R. Zaiane, M. El-hajj, and P. Lu, "Fast parallel association rule mining without candidacy generation," in In ICDM, 2001, pp. 665-668. Doi: 10.1109/ICDM.2001.989600
[15] J. Hu and X. Yang-Li, "A fast parallel association rules mining algorithm based on fp-forest." in ISNN (2), ser. Lecture Notes in Computer Science, F. Sun, J. Zhang, Y. Tan, J. Cao, and W. Y. 0001, Eds., vol. 5264. Springer, 2008, pp. 40-49. Doi: 10.1007/978-3-540-87734-9_5
[16] G. I. Webb, "Discovering associations with numeric variables," in Knowledge Discovery and Data Mining, 2001, pp. 383-388. Doi: 10.1145/502512.502569
[17] E. Gabriel et al., "Open MPI: Goals, concept, and design of a next generation MPI implementation," in Proceedings, 11th European PVM/MPI Users' Group Meeting, Budapest, Hungary, September 2004, pp. 97-104. Doi: 10.1007/978-3-540-30218-6_19
[18] K. Bache and M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml
[19] V. Novák, "A comprehensive theory of trichotomous evaluative linguistic expressions." Fuzzy Sets and Systems, vol. 159, no. 22, pp. 2939-2969, 2008. Doi:10.1016/j.fss.2008.02.023

Publications

[1] Štěpnička, M., Burda, M., Štěpničková, L. Fuzzy Rule Base Ensemble Generated from Data by Linguistic Associations Mining. FUZZY SET SYST. 2015 (submitted)
[2] Burda, M., Štěpnička, M., Štěpničková, L. Fuzzy Rule-Based Ensemble for Time Series Prediction: Progresses with Associations Mining. Strengthening Links between Data Analysis and Soft Computing. Heidelberg: Springer. 2015, 261-271. ISBN 978-3-319-10764-6
[3] Štěpnička, M., Štěpničková, L., Burda, M. Fuzzy Rule-Based Ensemble for Time Series Prediction: The Application of Linguistic Associations Mining. IEEE International Conference on Fuzzy Systems. Beijing, China: IEEE. 2014, 505-512. ISBN 978-147992072-3
[4] Burda, M., Pavliska, V., Valášek, R. Parallel Mining of Fuzzy Association Rules on Dense Data Sets. IEEE International Conference on Fuzzy Systems. Beijing, China: IEEE. 2014, 2156-2162. ISBN 978-147992072-3

Project website: bem4i.it4i.cz

TASK PROGRAMMING AND ASYNCHRONOUS COMMUNICATION FOR IRREGULAR COMPUTATIONS

Research institution:
University of Versailles (UVSQ)

Principal investigator:
Eric Petit (Intel)[1]

Researchers:
Loïc Thébault (UVSQ)
Quang V. Dinh (Dassault Aviation)

Project partner:
Dassault Aviation

Project IDs:
IT4I-3-4,
DD-16-20

Introduction

With the increasing requirement for energy efficiency and performance of new HPC system architectures, recent advances in hardware design result in an increasing number of nodes and cores, and an increasing usage of heterogeneous architectures. In future post-exascale systems, one can reasonably foresee thousands of nodes with more than a thousand cores per node. However, some resources like memory per core or memory bandwidth per core do not scale at the same rate. Most of the applications and runtimes currently in use struggle to scale with the present trend. Evolutions are requested for applications, runtimes, and programming models. In our view, to take benefit of the new systems, two major aspects prevail when designing an application: concurrency and locality.

During the EXA2CT project, we propose two contributions. The first one is DC_lib, a library for efficient parallel computation on unstructured meshes. The second is a proto-application that can be reused to demonstrate other approaches to unstructured meshes computation optimisations.

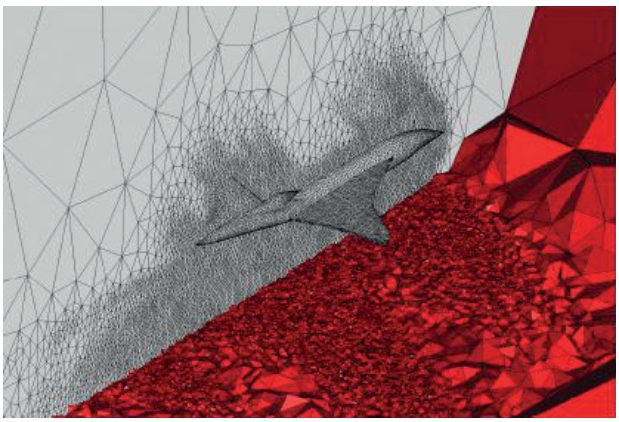


Figure 1.
3D unstructured meshes from Dassault Aviation

Results and Methods

In the DC_lib work, we propose a new approach for efficient hybrid parallelisation of unstructured mesh applications. It is based on domain decomposition at distributed memory level (process), recursive divide-and-conquer, D&C, at shared memory level (thread), and colouring for vectorisation at core level (SIMD).

The rationale is to adopt the best-suited strategy at each level of the architecture. We replace the rigid loop-based approach at shared memory level by a versatile and efficient recursive approach with an architecture oblivious design. This recursive approach naturally exposes data locality both in the mesh computation and in the associated sparse matrix system. This leads to a better locality than the original ordering using the Cuthill-McKee approach. We also observe that current colouring strategies are not efficient on the very small data partition size of the fine grain task-based parallelism. We propose a new colouring heuristic to reveal data-parallelism in small partitions. We

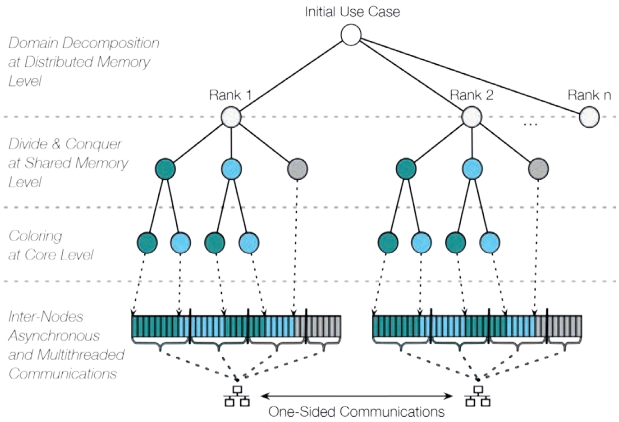


Figure 2.
Global view on DC_lib parallelisation scheme

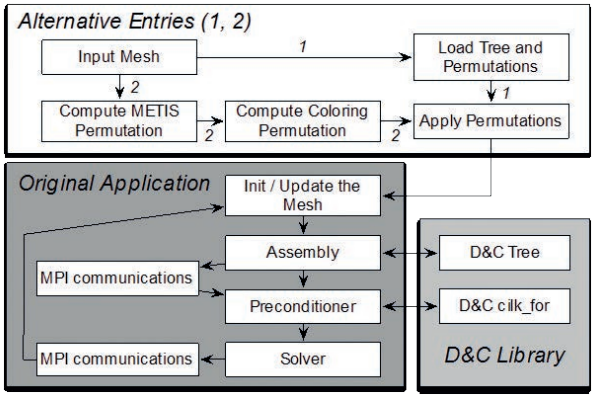


Figure 3.
DC_lib integration in the original application

identify and expose the new arising trade-off on many-cores between data structuredness, memory locality, and vectorisation. We propose a vectorisation ratio prediction model as a function of the vector length and the data partition size. The model is applied to evaluate the state-of-the-art with our original colouring strategy. We show that a higher speed-up can be achieved by limiting the vector length. The overall objective is to produce an application with no compromise on current performing solutions and able to scale with the current trends in system design.

Our target application, DEFMESH, is an industrial fluid dynamics Fortran code using the Finite Element Method, FEM, developed by Dassault Aviation. We implement our hybrid D&C approach using GASPI for the communication layer and the Cilk Plus task-based runtime as a C++ library that can be interfaced with minimum intrusion to the original code. To exploit the distributed memory parallelism, we propose a strategy to efficiently enforce communication between worker threads while computing. The proposed solution is to trigger communication as soon as enough data are ready to efficiently use the networks bandwidth.

Results

We evaluate D&C on two different architectures: a standard cluster of NUMA nodes and on 4 Intel KNC Xeon Phi nodes. D&C achieves a high parallel efficiency, a good data locality, and improved bandwidth usage. Furthermore, it competes favourably on current NUMA nodes with an optimised pure MPI version with a 10% speed-up at least and a maximum 3.47x at scale. It reaches an impressive 65% parallel efficiency in a strong scaling experiment on 32 nodes with only 2,000 vertices per core and less than an arithmetic intensity of one flop per byte.

Finally, the Intel Xeon Phi version shows a 97% parallel efficiency on the 60 physical cores. While running on 4 Xeon Phi, miniFEM has 92% parallel efficiency on the 240 physical cores compared to the 42% of the MPI version, and a performance equivalent to 96 Intel Xeon E5-2665 SNB cores.

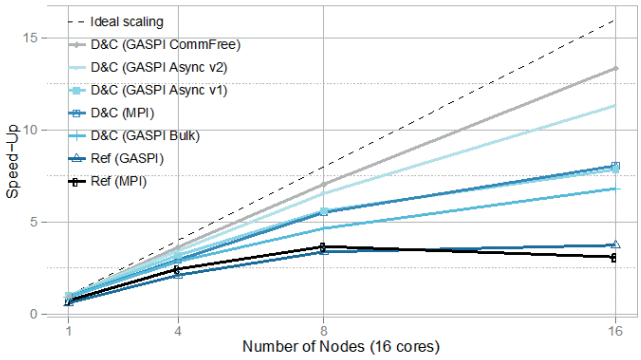


Figure 4.
Strong-scaling performance of DC_lib in miniFEM on 32 nodes

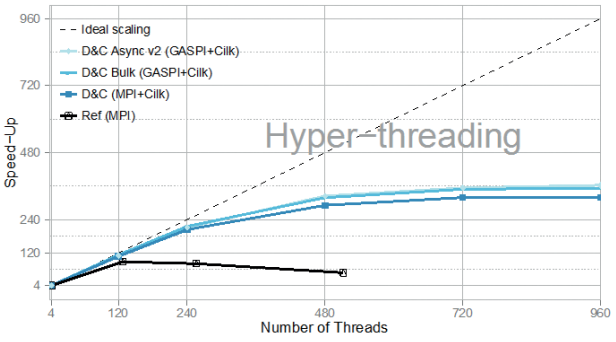


Figure 5.
Strong scaling experiment on 4 Intel Xeon Phi (KNC)

On-going Research / Outlook

Dassault Aviation is working on integrating the DC_lib in their main production code, AETHER. Initial results confirm the speedup obtained on the proto-application.

Furthermore, DC_lib is being assessed on the solver part (deflatedCG) of Yales2 in collaboration with Intel. We expect that the higher arithmetic intensity of the Yales2 solver will allow exploring strong scaling with a 100's mesh point per core, the figure of merit for Exascale.

Conclusion

The results we obtained in the context of the EXA2CT project and experimented on the Salomon platform in IT4Innovations are two-fold. First it has demonstrated how a proto-application extrapolated from an industrial use case can be effectively used as a vehicle for code modernisation. Second, it has demonstrated the potential of task based parallel programming and asynchronous communication for efficient computation on unstructured meshes. This work will be extended and applied to a wider range of applications. Our next target will be MLFMM for CEM computation.

References

[1] Frigo, M.; C.E. Leisersonand K.H. Randall. The implementation of the cilk-5 multithreaded language, in: ACM Sigplan Notices. 1998, volume 33, pp. 212-223. Doi: 10.1145/277652.277725

[2] Cuthill, E. and McKee, E. C. Reducing the Bandwidth of Sparse Symmetric Matrices. The 1969 24th National Conference. New York, NY, USA: ACM. 1969, pp. 157-172. Doi: 10.1145/800195.805928

Publication

[1] Thebault, L.; Petit, E., Dinh, Q., Jalby, W. Scalable and efficient implementation of 3d unstructured meshes computation: A case study on matrix assembly. In ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPoPP '15. USA, 2015. ISBN 978-1-4503-3205-7

Links

www.gaspi.de
https://github.com/EXAPARS

[1] at UVSQ until 31/08/2016

SPARSE BLOCK MATRIX LINEAR ALGEBRA (SBMLA)

Research institution:
Brno University of Technology

Principal investigator:
Lukáš Polok

Researchers:
Viorela Ila,
Marek Šolony,
Pavel Svoboda,
Pavel Smrž,
Pavel Zemčík

Project ID:
IT4I-6-9

Introduction

The aim of this project was to investigate more efficient numerical computation methods for problems in robotics, related to 3D reconstruction and robot navigation. These problems are sparse and have a particular structure. Because the variables involved in these problems exist in higher dimensions, such as R^2, R^3 or even SE(3), the associated matrices have a block pattern. There are a few implementations that take advantage of this pattern, typically with underwhelming results. In this project, we explored algorithms that can take advantage of explicit sparse block representation and form highly efficient algorithms for performing arithmetic operations.

Results and Methods

The state of the art graph optimizers are working with element-wise sparse matrices, and are typically not accelerated on GPUs. We have developed and implemented efficient sparse block matrix BLAS and applied it to nonlinear least squares solving. We have substantially extended existing block matrix methods by using C++ template meta-programming, allowing loop unrolling and automatic vectorization of BLAS routines. We have also explored suitable functionality for GPU acceleration.

We have also devised highly efficient methods for marginal covariance recovery, which outperform the state of the art by two orders of magnitude, while providing the same or better precision. These methods are used for quality estimation in problems in robotics and 3D reconstruction, as well as reduced pose SLAM and information theoretic distance calculation on graphs. We have also been able to apply the methods in digital cinema production, where they are needed for special effects creation.

On-going Research / Outlook

The research related to sparse block matrices is ongoing. We are interested in investigating the possibilities of GPU acceleration of blocked sparse primitives. Furthermore, pivoting matrix factorizations are of interest as pivoting at the granularity of blocks could reduce the fill-in while yielding factorizations of reasonable precision. The results look promising so far.

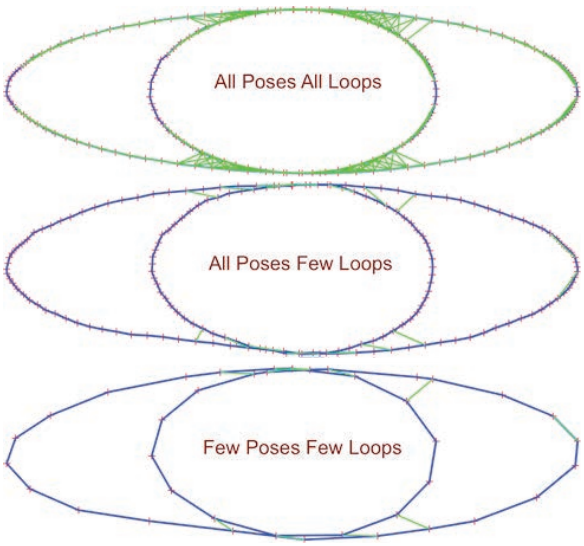


Figure 1.
Compact pose SLAM on a synthetic Ellipse dataset. This algorithm takes advantage of efficient covariance recovery in graphical inference problems to calculate the information gain of loop closures and of poses. That way, it is possible to produce a sparse map of the trajectory while not throwing away too much information. This technology is crucial for long-term missions e.g. in underwater surveying or robotic navigation.

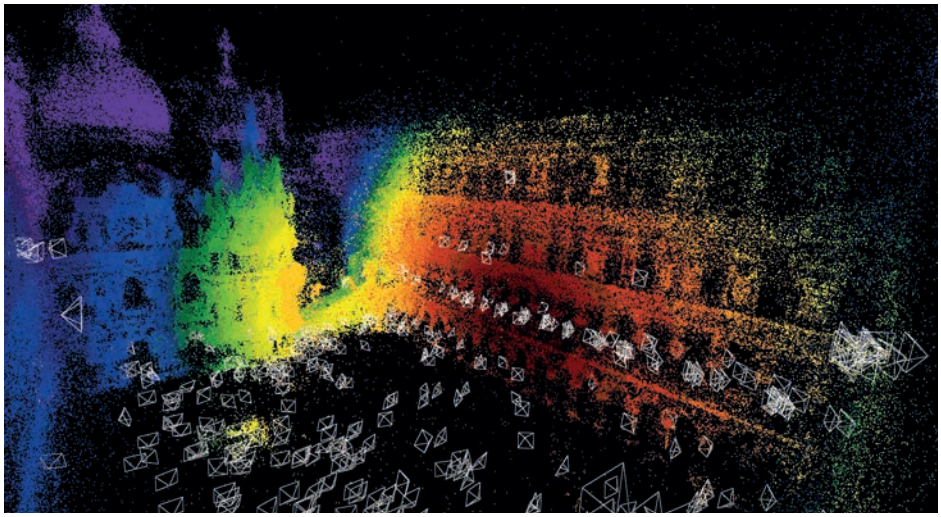


Figure 2.
3D reconstruction of San Marco square in Venice, Italy, computed entirely from Flickr pictures. The image shows false color rendering of 3D reconstruction quality.

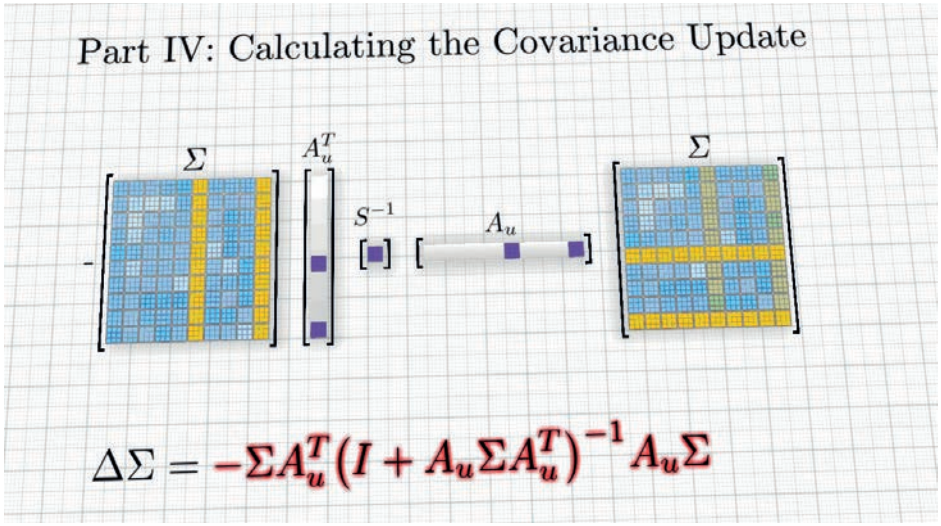


Figure 3.
An artistic rendering of the incremental covariance recovery equation and the related sparsity patterns

Conclusion

The IT4I infrastructure was indispensable for the achieved results. We particularly made use of fat nodes for calculating inverses of large matrices to support the covariance recovery research. Additionally, we made several large scale computations related to 3D reconstructions and loop closing for robotic navigation, requiring hundreds of compute nodes to be able to finish in a reasonable time.

Publications

[1] Ila V.; L. Polok, M. Solony and P. Svoboda. SLAM++. A Highly Efficient and Temporally Scalable Incremental SLAM Framework. The International Journal of Robotics Research. 2016, no. 123, pp. 1-22. ISSN 1741-3176

[2] Solony M.; E. Imre, V. Ila, L. Polok, H. Kim and P. Zemcik. Fast and Accurate Refinement Method for 3D Reconstruction from Stereo Spherical Images. In: Proceedings of the 10th International Conference on Computer Vision Theory and Applications. Berlin: Institute of Electrical and Electronics Engineers, Germany, 2015. ISBN 978-989-8425-47-8

[3] Ila V.; L. Polok, M. Solony, P. Zemcik and P. Smrz. Fast covariance recovery in incremental nonlinear least square solvers. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), Seatle, USA. Doi: 10.1109/ICRA.2015.7139841

Link

http://sf.net/p/slam-plus-plus/

ESPRESO – ACCELERATED SOLVER LIBRARY FOR INTEL XEON PHI SYSTEMS

Research institution:
IT4Innovations National Supercomputing Center

Principal investigators:
Lubomír Říha, Tomáš Brzobohatý

Researchers:
Alexandros Markopoulos, Michal Merta, Ondřej Meca

Project partner:
Universita della Svizzera italiana (Pardiso SC)

Project ID:
DD-16-20

Introduction

ExaScale PaRallel FETI SOlver (ESPRESO) is a sparse linear solver library that the IT4Innovations National Supercomputing Center in Ostrava, the Czech Republic has been developing since 2014. In 2016, the alpha version was released to the public and can be downloaded from the project website (espreso.it4i.cz).

ESPRESO contains not only the linear solver, but also several Finite Element (FEM) and Boundary Element (BEM) preprocessing tools designed particularly for FETI solvers, see Figure 2. The BEM support was produced in collaboration with developers of the BEM4I library (bem4i.it4i.cz). The preprocessor supports FEM and BEM discretisation for Advection-diffusion equation, Stokes flow, and Structural mechanics. Real engineering problems can be imported from Ansys Workbench or OpenFOAM. In addition, a C API allows ESPRESO to be used as a solver library for a third-party application. This has been used for integration with CSC ELMER. For large scale tests, the preprocessor also contains a multi-block benchmark generator. The post-processing and visualisation is based on the VTK library and Paraview, including Paraview Catalyst for inSitu visualisation.

The funding for the development of the library is provided by several sources, each focused on the development of particular features. For instance, the EU FP7 EXA2CT project funding has been used for implementation of the Hybrid FETI algorithm. This method provides excellent numerical and parallel scalability that allowed scientists to create this massively parallel library. A significant part of this work was done during the research internship at the Department of Aeronautics and Astronautics at Stanford University. The Intel Parallel Computing Centre funding is used to develop a new approach for acceleration of FETI methods in general, not just the Hybrid FETI, by Intel Xeon Phi accelerators.

This particular research was designed to take full advantage of the IT4Innovations Salomon supercomputer which, when installed, was the largest Intel Xeon Phi installation in Europe.

The project resources have been used to carry out two main tasks: (1) scalability testing and evaluation of the Hybrid Total FETI method, and (2) development of the Intel Xeon Phi and GPU acceleration of the FETI methods.

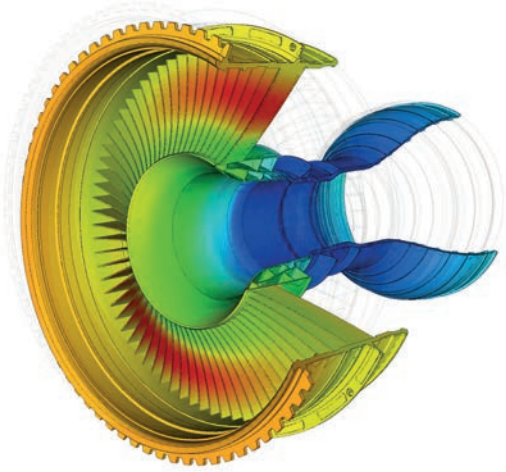


Figure 1.
The real world engine benchmark

Results and Methods

For many years, the Finite Element Tearing and Interconnecting method (FETI) [1] has been successfully used in the engineering community for solving very large problems arising from the discretisation of partial differential equations.

In such an approach, the original structure is decomposed into several non-overlapping subdomains. Mutual continuity of primal variables between neighbouring subdomains is enforced afterwards by dual variables, i.e., Lagrange multipliers (LM). They are usually obtained iteratively by one of the Krylov subspace methods, and then the primal solution is evaluated locally for each subdomain.

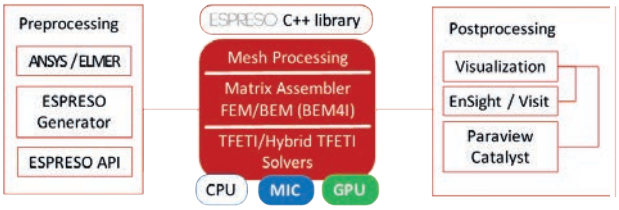


Figure 2.
The block diagram of the ESPRESO solver

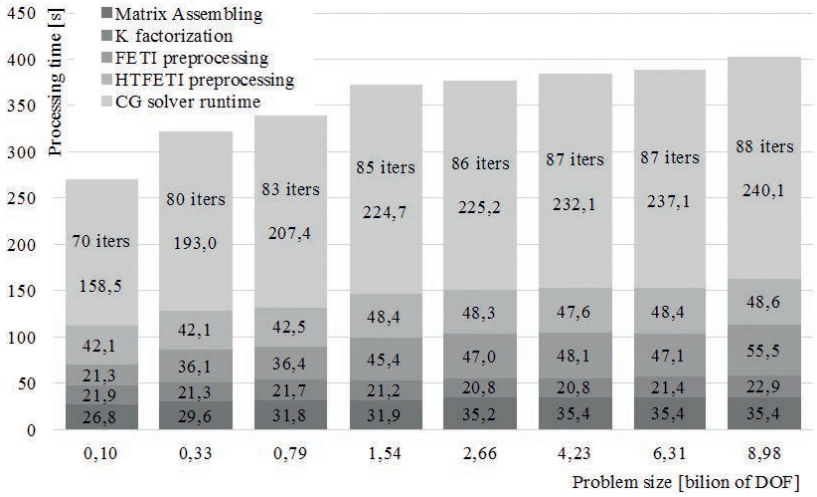


Figure 3.
The weak scalability of the HTFETI method on the Salomon supercomputer executed on up to 729 compute nodes

In 2006 Dostal et al. [2] introduced a new variant of an algorithm called Total FETI (or TFETI) in which Dirichlet boundary condition is also enforced by LM. The HTFETI method is a variant of hybrid FETI methods introduced by Klawonn and Rheinbach [3] for FETI and FETIDP. In the original approach, a number of subdomains are gathered into clusters. This can be seen as a three-level domain decomposition approach. Each cluster consists of a number of subdomains and for these, a FETI-DP system is set up. The clusters are then solved by a traditional FETI approach using projections to treat the non-trivial kernels. In contrast, in HTFETI, a TFETI approach is used for the subdomains in each cluster and the FETI approach with projections is used for clusters.

The main advantage of HTFETI is its ability to solve problems decomposed into a very large number of subdomains.

On the Salomon supercomputer, we have performed both weak and strong scalability tests. The weak scalability gives users an idea of how large problems can be solved by a given machine. These results are shown in Figure 3. It can be seen that the HTFETI method has been able to solve up to 8.9 billion unknowns on 729 compute nodes when solving a structural mechanics problem. For the strong scalability we have used the real world engine benchmark depicted in the Figure 1. These tests showed that HTFETI can achieve superlinear scalability, see Figure 4, from 43 (1,024 MPI processes) to 343 (8,192 MPI processes) compute nodes solving a structural mechanics problem of approximately 300 million unknowns.

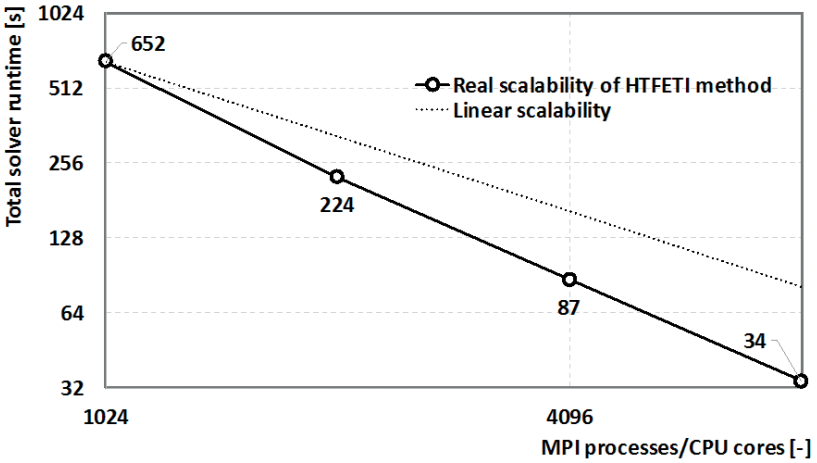


Figure 4.
The superlinear strong scalability of the HTFETI method on the real world engine benchmark

Acceleration of the Hybrid Total FETI domain decomposition method using the Intel Xeon Phi coprocessors provided key research for taking advantage of the Salomon machine. The HTFETI method is a memory bounded algorithm which uses sparse linear BLAS operations with irregular memory access pattern. We have developed a local Schur complement (LSC) method, which has a regular memory access pattern that allows the solver to fully utilise the Intel Xeon Phi fast memory bandwidth.

This translates to a speedup of as much as 7.8 on the HTFETI iterative solver when solving a 3 billion unknown heat transfer problem (Laplace equation) on almost 400 compute nodes

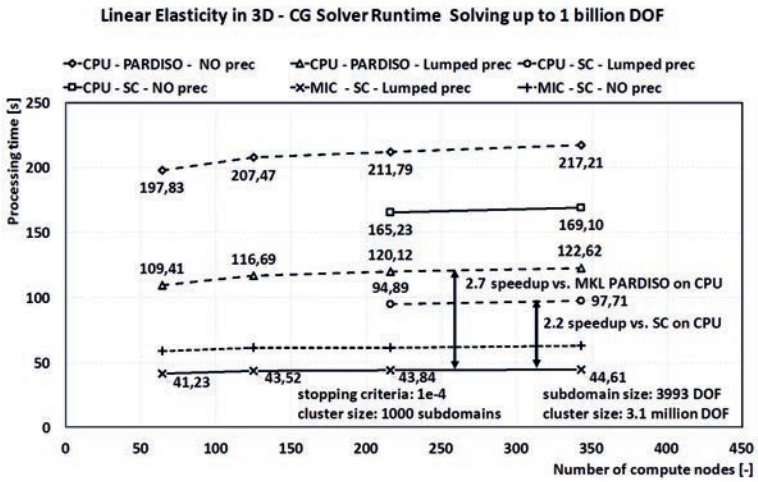


Figure 5.
The speed-up achieved by the Intel Xeon Phi acceleration in combination with the local Schur complement method can be as high as 7.8

and 800 Xeon Phi accelerators. The comparison is carried out between the CPU computation using sparse data structures (PARDISO solver) and the local Schur complement computation on Xeon Phi. In the case of the structural mechanics problem (linear elasticity) of size 1 billion DOFs, the respective speedup is 3.4.

The presented speedups are asymptotic and they are reached for problems requiring a high number of iterations (e.g. ill-conditioned problems, transient problems, contact problems). For problems which can be solved with less than one hundred iterations, the local Schur complement method is not suitable. For these cases, we have implemented sparse matrix processing using PARDISO also for the Xeon Phi accelerators. The weak scalability of the Xeon Phi accelerated version is shown in Figure 5.

On-going Research / Outlook

In the coming months we would like to further develop the ESPRESO FEM package, which will allow us to more efficiently solve real world problems. From the solver point of view, we will work on improvements of the numerical scalability of the Hybrid Total FETI method and its acceleration using next generation Intel Xeon Phi Knights Landing processors.

Conclusion

Being able to run scalability tests on a near full scale of the Salomon machine gives us initial insight on the scalability issues that have to be addressed in order to run the library on the world's largest petascale systems – which is our next goal.

Intel Xeon Phi coprocessor is a modern, many-core architecture which shares several key features (fast memory, large number of SIMD cores) with GPU accelerators. The proposed Local Schur complement method efficiency has been successfully evaluated on the Xeon Phi (the Knights Corner generation) in this report, but one can expect similar behaviour on GPUs.

Based on our survey of upcoming pre-exascale machines, one cannot ignore many-core architectures and heterogeneous systems with some type of accelerator. Therefore, we will put more effort in this type of research in the near future.

References

[1] Farhat, C.; Mandel, J., and Roux., F.-X. Optimal convergence properties of the FETI domain decomposition method. Computer Methods in Applied Mechanics and Engineering. 1994, 115, 365-385. Doi: 10.1016/0045-7825(94)90068-X

[2] Dostal, Z., Horak, D., and Kucera, R. Total FETI – an easier implementable variant of the FETI method for numerical solution of elliptic PDE. Communications in Numerical Methods in Engineering. 2006, 22(12), 1155-1162. Doi: 10.1002/cnm.881

[3] Klawonn, A. and Rheinbach, O. Highly scalable parallel domain decomposition methods with an application to biomechanics. ZAMM. 2010, 90(1), 5-32. Doi: 10.1002/zamm.200900329

Publications

[1] Riha Lubomir; Brzobohaty, Tomas; Markopoulos Alexandros; Meca Ondrej; Kozubek Tomas. Massively Parallel Hybrid Total FETI (HTFETI) Solver. Platform for Advanced Scientific Computing Conference, PASC, ACM, 2016. ISBN: 978-1-4503-4126-4/16/06

[2] Riha Lubomir; Brzobohaty, Tomas; Markopoulos Alexandros; Meca Ondrej; Kozubek Tomas; Schenk Olaf; Vanroose Wim. Efficient Implementation of Total FETI Solver for Graphic Processing Units Using Schur Complement. Conference, HPCSE 2015, LNCS 9611, 2016.

[3] Riha, Lubomir; Brzobohaty, Tomas; Markopoulos, Alexandros. Hybrid parallelisation of the Total FETI solver. Advances in Engineering Software. 2016, 103, 29-37. ISSN: 0965-9978.

[4] Riha, Lubomir; Brzobohaty, Tomas; Markopoulos, Alexandros; Jarosova, Marta; Kozubek, Tomas; Horak, David; Hapla, Vaclav. Implementation of the Efficient Communication Layer for the Highly Parallel Total FETI and Hybrid Total FETI Solvers. Parallel Computing. 2016, 57, 154-166. ISSN: 0167-8191.

Project website: espreso.it4i.cz

COMPUTATIONAL METHODS FOR SOLVING STOCHASTIC EQUILIBRIUM MODELS

Research institution:
Universita della Svizzera italiana

Principal investigators:
Olaf Schenk, Simon Scheidegger

Researcher:
Dmitry Mikushin

Project partner:
University of Zurich

Project ID:
DD-16-7

Introduction

Dynamic stochastic general equilibrium models with heterogeneous agents are of paramount interest in quantitative economics. One particular subclass of them which is particularly well suited for policy analysis is the so-called overlapping generation (OLG) model. Computable models with OLG are routinely used in public finance for counterfactual policy analysis as they allow for a careful modelling of individuals' decisions over the life cycle and their interactions with capital accumulation and economic growth. Solving for the global solution of a model with substantial heterogeneity is indeed very costly: the computation time and storage requirements increase dramatically with the amount of heterogeneity, i.e. with the dimensionality of the problem. It is therefore often far beyond the scope of current methods to include as much heterogeneity as a natural modelling choice would suggest. Building on [1],[2], the work carried out in the Anselm and Salomon clusters aim at porting the massively parallel solver for high-dimensional dynamic stochastic models to Intel Xeon Phi accelerators hardware in order to solve OLG models of unprecedented and, previously unreached performance by making use of AVX-512 as well as many threads available on the node level.

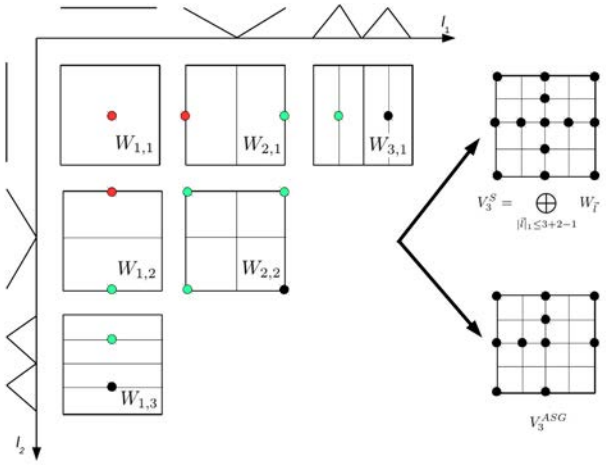


Figure 1.
Construction of a classical sparse grid (top right panel) as well as an adaptive sparse grid (lower right panel)

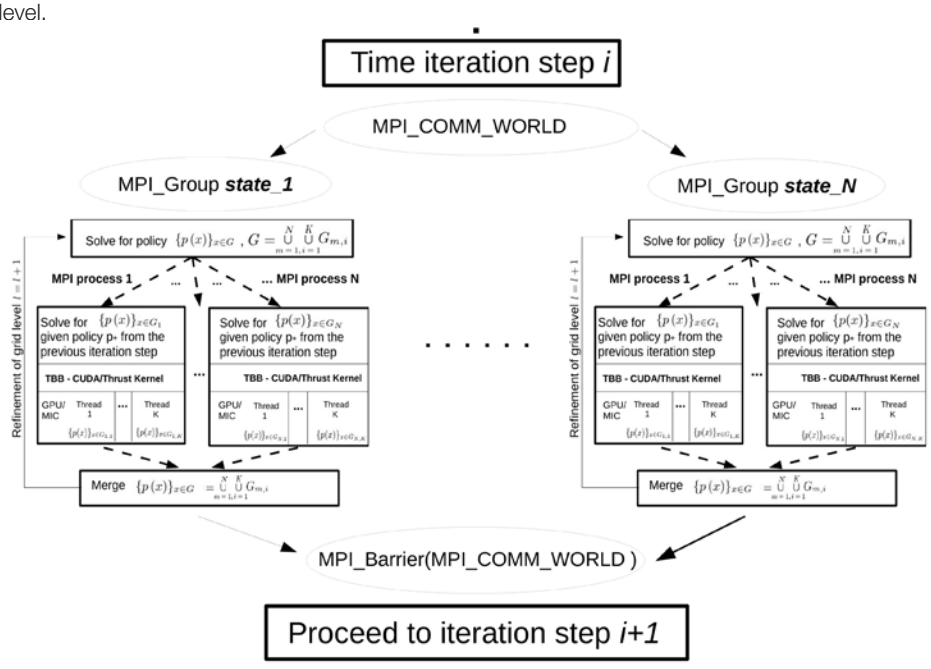


Figure 2.
Comparison of wall times for different stochastic OLG code variants on a single node of Salomon (speedup against scalar version)

Methods

Stochastic dynamic models are often solved by iterating on the functions that represent economic choices (by so-called time iteration algorithms; see, e.g. [1]. There are two major bottlenecks when solving large-scale dynamic stochastic OLG models with time iteration, namely, (i) in each iteration step, an economic function needs to be approximated. For this purpose, the function value has to be determined at many points in the high-dimensional state space, and (ii) each point involves solving a system of nonlinear equations (around 60 equations in 60 unknowns). These two items of the considered problems create difficulties in achieving a fast time-to-solution process. Thus, an extremely efficient and most of all highly scalable implementation of a time iteration scheme becomes crucial. We overcome these difficulties by massively reducing both the number of points to be evaluated and the time needed for each evaluation. Issue (i) is resolved by the use of adaptive sparse grids (see Figure 1 and e.g., [3],[4]), while task (ii) is resolved using a hybrid parallelisation scheme that uses message passing interface (MPI) among compute nodes and minimises interprocess communication by using Intel threading building blocks (TBB). Furthermore, it relies on AVX or AVX-512 vectorisation and offloads the function evaluations partially to Xeon Phi coprocessors to further speed up computations. This scheme, displayed in Figure 2, enables us to make efficient use of the emerging hybrid high-performance (HPC) computing facilities that consist of a variety of special purpose as well as general purpose hardware, and whose performance can today reach multiple petaflop/s.

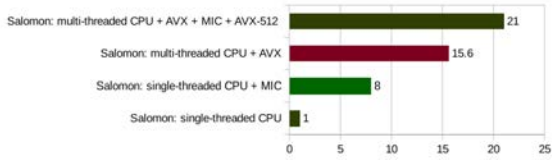


Figure 3.
Comparison of normalised wall times for the OLG code variants on a single node of Salomon (speedup against scalar version)

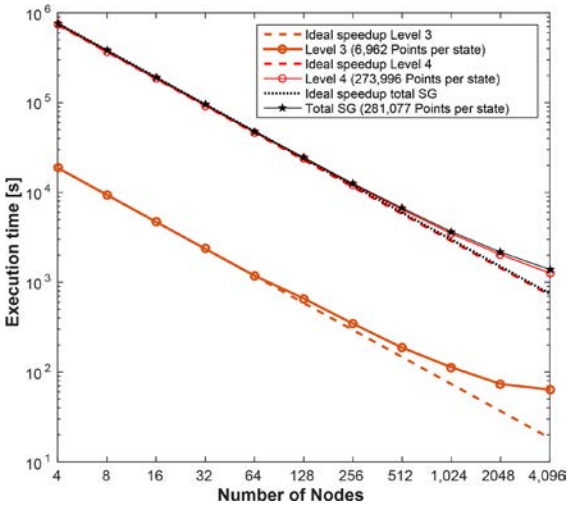


Figure 4.
Strong scaling plot on CSCS-s Piz Daint for an OLG model. Dashed and dotted lines show the ideal speedup.

Their architecture usually comes in two flavours, namely, CPU compute nodes with either attached GPUs or Xeon Phi co-processors. In greater detail, our massively parallel framework is parallelised as follows: In each step of the above time iteration procedure the updated policy function is determined using a hybrid-parallel algorithm, as displayed in Figure 2. Conceptually, the highest levels of parallelism are the discrete states of the OLG model, which are completely independent from each other within one time step. Hence, the MPI_COMM_WORLD communicator is split into N groups (one per state), each of them consisting of an individual discrete state. Inside one MPI Group, we construct an adaptive sparse grid by distributing the newly generated grid points via MPI within a refinement step among multiple, multithreaded processes. The points that are sent to one particular compute node are further distributed among different threads. Each thread then solves a set of nonlinear equations for every single grid point assigned to it. On top of this, we add an additional level of parallelism. When searching for the solution to the equation system at a given point, the algorithm has to frequently interpolate the function computed in the previous iteration step. These interpolations take up 99% of the computation time needed to solve the equation system. As they have a high arithmetic intensity – that is to say, many arithmetic operations are performed for each byte of memory transfer and access – they are perfectly suited for SIMD devices such as MICs. We therefore offload parts of the interpolation from the compute nodes to their attached accelerators. Hence, CPU cores and the MIC device of a single node are utilised through multiple threads, and MPI is used for internode communication only. Multithreading on compute nodes is implemented with TBB. Moreover, one of the TBB-managed threads is exclusively used for the MIC dispatch. Moreover, note that the CPU and

MIC threads leverage TBB's automatic workload balancing based on stealing tasks from slower workers.
In the next step, we consider the Xeon Phi offload mode as the primary mode of operation for the coprocessor. The device receives the sparse grid assembly from the host processor and reports back upon completion of the assignment without the host being involved in between these two events. It is well known that Intel does not offer an explicit programming model for the Xeon Phi. Instead of data transfer and kernel launches, the user must orchestrate the Xeon Phi with offloading directives. However, an explicit interface exists only at low-level and is called COI (Intel Coprocessor Offload Infrastructure). On top of COI and liboffloadmic (GCC open source offloading library), we have developed an MIC runtime API (micrt), which mimics CUDA Runtime API. Using micrt, we perform data transfers and kernel launches on Xeon Phi. When performing computations on the Xeon Phi, we assigned 4 threads to a single MIC core. On top of that, each thread handles a vector of 8 doubles at once (512 bits) thanks to AVX-512.

Performance and Results

To give a measure of how the various hybrid single node parallelisation schemes discussed before impact the performance of the time iteration code on a single CPU thread, MIC, and the entire node, we display a representative test case in Figure 3. We believe that the combined speedup observed on the individual Salomon/Anselm nodes as well as among nodes (cf. Figure 4) that our code framework will achieve on the emerging large-scale Xeon Phi clusters is at least 4 orders of magnitude, thus enabling us to solve dynamic stochastic economic OLG models of unprecedented complexity. As a real application of our framework, we solved an annually calibrated OLG model with 4 discrete shocks (REF). This model consists of roughly 11,800,000 nonlinear equations and unknowns per time step.
Figure 5 displays the improving error quality with the number of grid points.
Figure 6 displays the error of the time iteration algorithm with respect to node hours for a complete simulation of a 59-dimensional model. It is apparent from Figure 6 that convergence of the time iteration algorithm is rather slow. This is to be expected, as time iteration has, at best, a linear convergence rate [5].

On-going Research / Outlook

In the coming months we would like to further improve our node level parallelism on the Xeon Phi devices. This will allow us further speedup that will be required to solve OLG models of greater detail and thus greater realism. The steps we are planning will ensure that our code framework will be ready for Intel Xeon Phi Knights Landing processors.

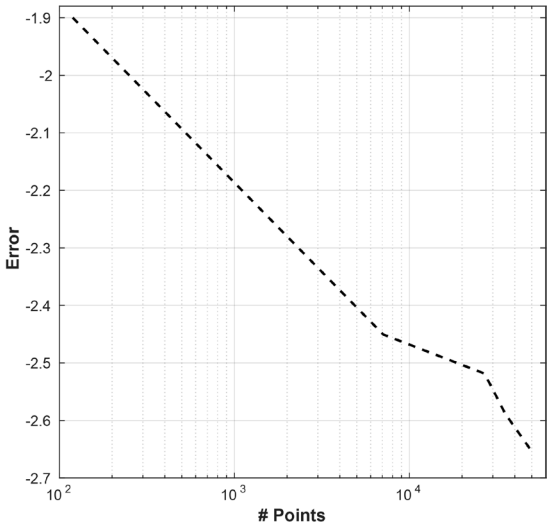


Figure 5.
Sparse grid solutions of the 59-dimensional OLG model as a function of the number of average grid points per state, resulting from decreasing the refinement threshold

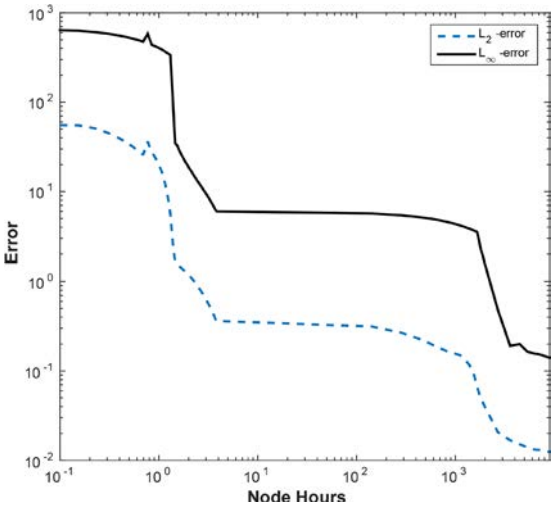


Figure 6.
Comparison of the average L2 and L∞-error of state 1 for adaptive sparse grid solutions of the 59-dimensional OLG model as a function of compute time

Conclusion

We demonstrate in our work that was partially carried out on Salomon and Anselm that by combining adaptive sparse grids (that ameliorate the curse of dimensionality imposed by the large heterogeneity of the economic model) with a time iteration algorithm (that deals with the recursive nature of the problem formulation), and hybrid HPC compute paradigms (which drastically reduces the time-to-solution process), we are able to handle the difficulties imparted by this particular model class up to a level of complexity not seen before. By exploiting the generic structure of the economic models under consideration, we implemented a hybrid parallelisation scheme that uses state-of-the-art parallel computing paradigms. It minimises MPI interprocess communication by using TBB, AVX-512 vectorisation and partially offloads the function evaluations to Xeon Phi coprocessors Numerical experiments on IT4Innovations' Salomon, Anselm (and Piz Daint at the Swiss National Supercomputer Centre) show that our code is highly scalable. In the case of a stochastic OLG model with 60 generations and four discrete states, we found very good strong scaling properties.

References

[1] Brumm, J., Mikushin, D., Scheidegger, S. and Olaf Schenk. Scalable high-dimensional dynamic stochastic economic modelling. Journal of Computational Science. 2015, 11, 12-25. Doi: 10.1016/j.jocs.2015.07.004.

[2] Brumm, J. and Simon Scheidegger. Using adaptive sparse grids to solve high-dimensional dynamic models: Working paper. SSRN Electronic Journal. 2014. Doi: 10.2139/ssrn.2349281.

[3] Bungartz, Hans-Joachim and Michael Griebel. Sparse grids. Acta Numerica. 2004, 13, 1-123. Doi: 10.1017/S0962492904000182.

[4] Ma, Xiang and Nicholas Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. Journal of Computational Physics. 2009, 228(8), 3084-3113. Doi: 10.1016/j.jcp.2009.01.006.

[5] Maldonado, W.L. and B.F. Svaiter. Holder continuity of the policy function approximation in the value function approximation. Journal of Mathematical Economics. 2007, 43(5), 629-639. Doi: 10.1016/j.jmateco.2007.01.004.

Publication

[1] Simon Scheidegger; Dmitry Mikushin, Felix Kubler; Olaf Schenk, "Tackling large-scale stochastic overlapping generation models on heterogeneous HPC systems": to be submitted.

MATSOL – DEVELOPMENT, TESTING AND SUPPORT LIBRARIES

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Oldřich Vlach

Researcher:
Václav Ryška

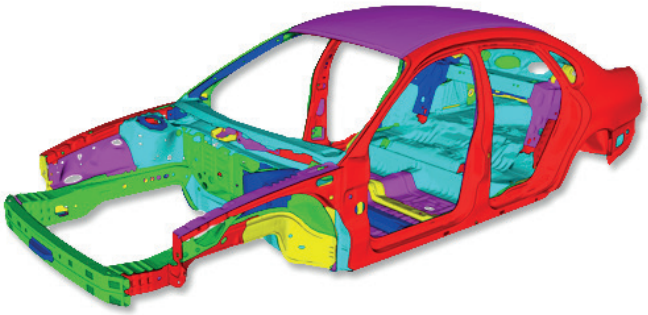
Project ID:
IT4I-2-3

Introduction

MatSol is a Matlab based library for engineering computation with the spacial focus on contact problems. The algorithms implemented in the MatSol library are based on our long-term research focused on the development of scalable algorithms for multibody contact and contact shape optimisation problems, 2D or 3D. Most of them were developed by Zdeněk Dostál and are described in detail in [1]. On the development we cooperate with the Department of Applied Mathematics at VSB.

The Matsol library serves as a referential implementation of those algorithms. Variants of FETI and BETI known for their parallel and numerical scalability are used for domain decomposition. The resulting class of QP problems are solved using our special QP algorithms. The unique feature of them is their capability to solve these QP problems with a bounded spectrum in O(1) iterations. Theory yields the error bounds that are independent of conditioning of constraints and the results are valid even for linearly dependent equality constraints.

To parallelise the algorithms, the MATLAB® Distributed Computing Server™ and Parallel Computing Toolbox™ are used. Hence, MatSol has the full functionality to solve efficiently large problems of mechanics. Users can solve static or transient analysis, optimisation problems, problems in linear and nonlinear elasticity, and contact problems.



For discretisation, the finite or boundary element methods are used. As the domain decomposition techniques, the FETI or BETI methods are implemented. The decomposition into subdomains is done using Metis and spectral methods. The solution process could be run either in sequential or parallel mode. The solution algorithms are implemented in such a way that the code is the same for both sequential and parallel mode.

The MatSol library also includes tools for postprocessing and postplotting. The results of the problem are then converted into the modelling tools for further postprocessing through the model database.

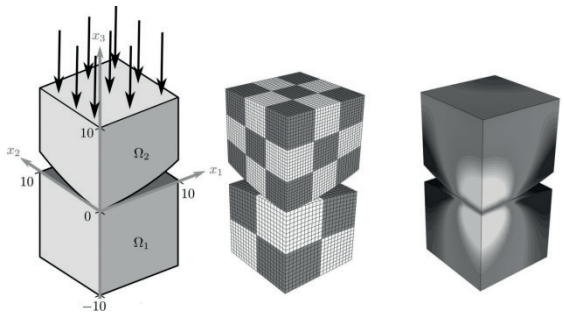


Figure 1.
3D Hertz example – setting (left), decomposition case (middle), von Mises stress (right)

h_{slave}/h_{master}	1/6	1/3	2/3	1	3/2	3	6
$\lambda_{min}(B_I B_I^T)$	0.49	0.40	0.43	0.73	0.86	0.83	0.88
$\lambda_{max}(B_I B_I^T)$	19.2	7.00	2.30	1.32	1.16	1.31	1.20
$\kappa(B_I B_I^T)$	42.8	17.5	5.36	1.82	1.35	1.57	1.37
$\lambda_{min}(F\|KerG)$	1.6e-6	2.4e-6	2.8e-6	3.8e-6	2.7e-6	2.2e-6	1.4e-6
$\lambda_{max}(F\|KerG)$	1.1e-4	1.1e-4	1.1e-4	1.1e-4	9.9e-5	9.9e-5	9.9e-5
$\kappa(F\|KerG)$	72.8	46.7	40.7	29.9	37.2	44.0	72.7

Table 1.
Conditioning of the mortar matrices B for a 3D Hertz problem with non-matching grids and domain decomposition

Results and Methods

During the year 2016 our main interests were focused on two topics. The first topic was the computation of scalability tests of quadratic programming algorithms with mortar contact matrices and different domain decompositions. The second topic was the investigation of the possible extension of the FETI approach to non-symmetric problems.

Recent developments in constrained quadratic programming problems included theoretically supported scalable algorithms for the solution of multibody contact problems discretised by the finite and boundary elements. These results were also adapted to the solution of problems with friction and to transient contact problems. Let us recall that an algorithm is numerically scalable for a given class of problems if the cost of the solution is asymptotically proportional to the number of unknowns, and it enjoys parallel scalability if the time of the solution can be reduced nearly proportionally to the number of available processors. The theoretical results on the numerical scalability of the contact algorithms assume strong linear independence of the rows of the "gluing" constraint matrices and the non-penetration conditions which are constructed using the biorthogonal mortar approach. We show that the constraint matrices arising from the discretisation by some biorthogonal bases are well conditioned under natural restrictions and satisfy the assumptions that guarantee efficiency of the QP algorithms [1,5]. The results in [5] were extended to the case when the domains are decomposed into subdomains and the FETI method was chosen to glue them together. The conditioning numbers of mortar matrices for the Hertz 3D benchmark from Figure 1 is depicted in Table 1.

So far, the scalability results for our QP algorithms (see [1]) assumed that the bodies and subdomains involved in a problem are made of the same material and that the stiffness matrices of the subdomains are reasonably conditioned. The reorthogonalisation-based preconditioning [4] was implemented and tested in MatSol. It improves the Hessian condition number but also preserves the structure of the inequality constraints, so that they affect both the linear and nonlinear steps.

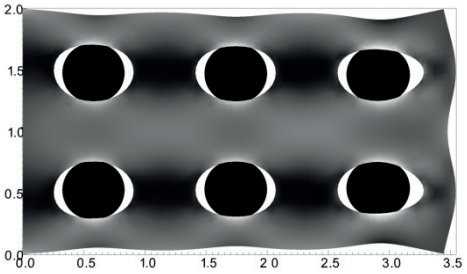


Figure 2.
Material with circular insets – deformation and the von Mises stress for E2 /E1 = 10,000

Now to the second topic. Although FETI is originally suited for symmetric problems, we can successfully use its benefits for non-symmetric matrix systems as well. Such system arises for example from convection-diffusion governing equations when convection dominates over diffusion. In this case the FEM solution contains spurious oscillations and, therefore, we need to stabilise it. We solved these equations for scalar unknown quantity without reaction term using stabilisations described in [3]. We used BiCGSTAB and GMRES iteration solvers for non-symmetric systems without a preconditioner and with a standard Lumped and Dirichlet preconditioner. The last preconditioner we applied was a simplified version of the Dirichlet preconditioner marked as DIAG D. Further we implemented "non-scaled" and enhanced "scaled" preconditioners containing the ones mentioned above. The performance is depicted in Table 2.

K = 1e^{-5}, CAU, GMRES

preconditioner		NONE		LUMPED		DIAG D		DIRICHLET	
GMRES		non-scaled	scaled	non-scaled	scaled	non-scaled	scaled	non-scaled	scaled
H/h	1/H	number of iterations							
8	2	40	40	17	16	14	12	12	10
8	8	98	98	39	36	30	23	34	30
8	16	148	148	45	44	35	28	47	45
8	32	218	218	53	48	39	34	71	71
16	2	54	54	25	23	19	16	16	13
16	8	143	143	73	67	57	40	65	54
16	16	238	238	106	96	73	51	128	104
16	32	381	381	133	124	86	62	232	223
32	2	75	75	36	34	26	23	22	18
32	8	225	225	120	109	89	67	115	96
32	16	412	190	211	190	143	99	261	222
32	32	745	745	346	289	224	143	559	660

K = 1e^{-5}, CAU, BiCGSTAB

preconditioner		NONE		LUMPED		DIAG D		DIRICHLET	
BiCGSTAB		non-scaled	scaled	non-scaled	scaled	non-scaled	scaled	non-scaled	scaled
H/h	1/H	number of iterations							
8	2	47	47	14	14	10	8	8	7
8	8	129	129	24	22	19	15	23	20
8	16	212	212	29	27	22	19	32	35
8	32	319	319	34	31	25	22	52	97
16	2	80	80	23	24	14	13	13	10
16	8	261	261	55	51	38	25	51	52
16	16	471	471	67	71	44	34	119	115
16	32	762	762	88	86	60	45	19	365
32	2	146	146	39	37	25	21	18	16
32	8	565	565	130	124	74	49	115	127
32	16	1114	1114	202	185	116	72	-	-
32	32	1822	1822	277	277	386	209	-	-

Table 2.
The performance of GMRES and BiCGSTAB using different preconditioners

On-going Research / Outlook

In the coming months we would like to further extend the capability of the MatSol library to compute the problems of shallow water simulations, incorporate the parallel contact search algorithm as well as the new variants of constrained QP algorithms.

Conclusion

MatSol serves us as the platform to fast implementation of new problems and algorithms. It allows us to perform primary tests and benchmarks for the parallel performance of new algorithms. The performance results obtained by MatSol serve as the foundation for the decision of which algorithm is worthy to implement in C/C++ to another in-house code.
Fast implementation in Matlab reduces the time of benchmarking new algorithms and allows us to publish our actual research with minimal time delay.

References

[1] Dostál, Zdeněk, “Optimal Quadratic Programming Algorithms: With Applications to Variational Inequalities”, Springer, (2009). ISBN 978-0-387-84806-8
[2] Pechstein, Clemens, “Finite and boundary element tearing and interconnecting solvers for multiscale problems”. Berlin: Springer, 2013. ISBN 978-3-642-23588-7
[3] Gomes Dutra do Carmo, Eduardo; Benitez Alvarez, Gustavo, “A new upwind function in stabilised finite element formulations, using linear and quadratic elements for scalar convection-diffusion problems”, Computer Methods in Applied Mechanics and Engineering, Volume 193, Pages 2383-2402, (2004). Doi: 10.1016/j.cma.2004.01.015

Publications

[1] Dostál, Zdeněk; Kozubek, Tomáš; Vlach, Oldřich; Brzobohatý, Tomáš, “Reorthogonalisation-based stiffness preconditioning in FETI algorithms with applications to variational inequalities”, Numerical Linear Algebra with Applications, Volume 22, Issue 6, Pages 987-998, (2015). Doi: 10.1002/nla.1994
[2] Vlach, Oldřich; Dostál, Zdeněk; Kozubek, Tomáš, “On conditioning of constraints arising from variationally consistent discretisation of contact problems and duality based solvers” Computational Methods in Applied Mathematics, Volume 15, Issue 2, Pages 221-231, (2015). ISSN 1609-4840

Project website: industry.it4i.cz/en/products/matsol/

BEM4I – VECTORISED BOUNDARY ELEMENT LIBRARY FOR MANY-CORE SYSTEMS

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Michal Merta

Researchers:
Jan Zapletal, Lukáš Malý, Michal Kravčenko

Project IDs:
IT4I-9-12, OPEN-7-13

Introduction

BEM4I is a boundary element library that has been under development at the IT4Innovations National Supercomputing Center in Ostrava since 2013. The library aims at the efficient assembly of the boundary element matrices both on modern PCs and within the high performance computing environments including the Anselm and Salomon clusters of IT4Innovations.

In its current state, the library provides assemblers and solvers for the Laplace (heat transfer, electrostatics), Helmholtz (time-harmonic wave scattering, see Figure 1 for a solution of a wave scattering transmission problem), and Lamé equations (linear elasticity) in 3D, see Figure 2 for the structure of the library. An efficient massively-parallel space-time approach [6] is also implemented for the solution of the time-dependent wave equation. Visualisation of the output results is available in the ParaView format.

To take full advantage of modern processors BEM4I leverages several layers of parallelism. While it has become customary in similar scientific codes to use shared-memory parallelism provided by the OpenMP standards, vectorisation of the code inside each available core is still often overlooked. With the rise of modern multi- and many-core platforms including the Intel Xeon Phi architecture providing wide Single Instruction Multiple Data (SIMD) registers, this leads to computational programs hardly able to reach their peak performance. In BEM4I the vectorisation is enabled for the evaluation of singular surface integrals and has been successfully tested on different architectures including Salomon's Haswell and Knights Corner (co) processors and the newest generation of the Intel Xeon Phi processor, codenamed Knights Landing, available at Intel's Endeavour cluster [4,7].

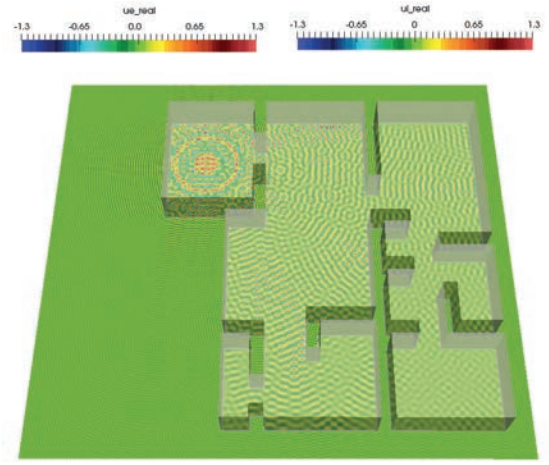


Figure 1.
Wave scattering transmission problem

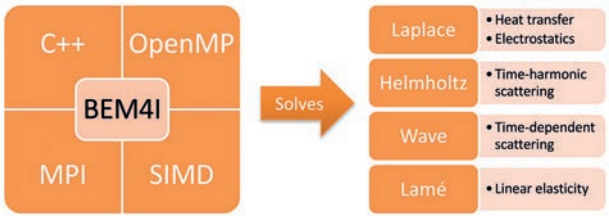


Figure 2.
Structure of the BEM4I library

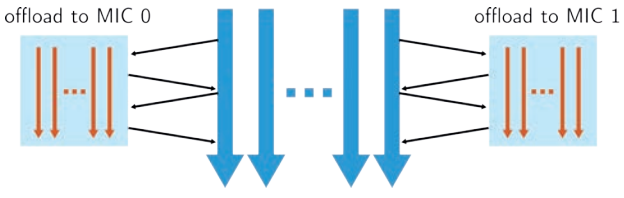


Figure 3.
Offload model for the matrix assembly

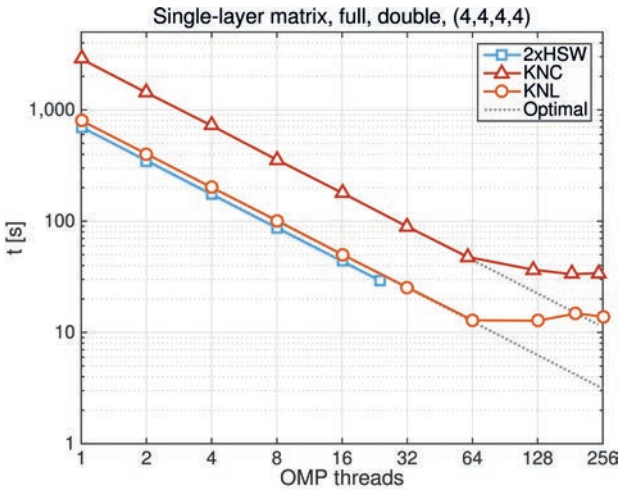


Figure 4.
Strong scalability with respect to the number of threads

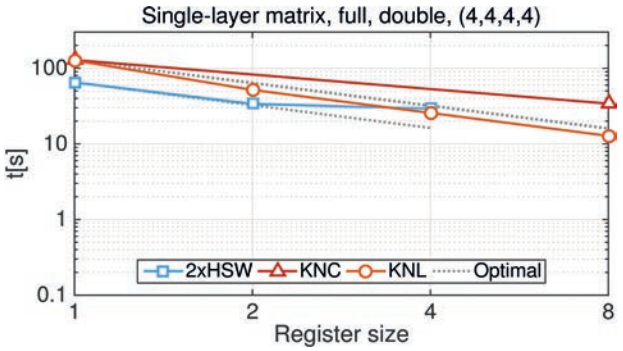


Figure 5.
Strong scalability with respect to the SIMD register width

The research within the project granting access to the nodes of the Salomon and Anselm clusters also aimed at the offloaded model with the Knights Corner coprocessors, where the main application runs on the host CPU and computationally intensive parts are offloaded to the coprocessor [5], see Figure 3 for an illustration.

To deploy BEM4I at massively parallel architectures, the library can be linked to the Espreso library (espresso.it4i.cz) providing an implementation of a hybrid domain decomposition method. This combination leads to the boundary element tearing and interconnecting (BETI, [1]) method utilising all available parallelisation layers including MPI over distributed memory, OpenMP threading within a single node, OpenMP SIMD vectorisation, and offload to the available coprocessors.

Results and Methods

The boundary element method (BEM) is a method for numerical solution of partial differential equations and represents a counterpart to the widely used finite element method (FEM). The major advantage over FEM is the dimension reduction since the problem under consideration is reformulated as a boundary integral equation. Unlike FEM, the boundary element method thus only requires discretisation of the boundary.

On the other hand, the setup of system matrices requires evaluation of singular kernels. In BEM4I we rely on two approaches, namely the semi-analytic assembly [3] and fully numerical regularised quadrature scheme [2] allowing for efficient vectorisation. Also, the system matrices produced by the standard BEM are fully populated and thus the complexity is quadratic with respect to both memory requirements and computational effort. To overcome this issue, the implementation of the adaptive cross approximation (ACA) is provided to lower the complexity to log-linear.

To test the efficiency of BEM4I on a single node of a computational cluster with respect to the number of OpenMP threads working over shared memory and, more importantly, to the width of the SIMD register, we performed scalability experiments on three different architectures. Two of them, namely the Haswell processors (HSW) and Knights Corner coprocessors (KNC), have been provided by the Salomon cluster at IT4Innovations. To prepare for the next generation of Intel Xeon Phi standalone processors, the library was also tested on Intel's Endeavour cluster providing access to the Knights Landing processors (KNL).

In Figure 4, the strong scalability of the assembly code for the full single-layer operator matrix is presented for all architectures mentioned above. The speedup with respect to scalar code reaches 24.07 on HSW with 2x12 cores, 85 on KNC with 61 physical cores, and 62 for KNL with 64 physical cores. Similar results hold for other matrices and also in the ACA-sparsified format.

The vectorisation results are presented in Figure 5. Utilising the most recent AVX-512 instructions available on KNL, the speedup with respect to the scalar version of the code reaches 9.94. The experiments performed show that vectorisation becomes a crucial part of the scientific code design process.

The offload mode presented in [5] also delivers further speedup of the assembly times. The graph in Figure 7 presents the computational times for the pure host setup and the assembly partly offloaded to one or two KNC coprocessor cards available at the accelerated nodes of the Salomon cluster. This approach leads to a further speedup of 2.66.

On-going Research / Outlook

A further aim of the BEM4I package is to add vectorisation of the semi-analytic assembly as presented in [3]. Compared to the fully numerical assembly, the difficulty lies in a low number of quadrature points for the outer integral and thus a rather short SIMD vectors. Also, the evaluation includes several branching points, which slow up the vectorised computations.

To fully incorporate the offload mode to the structure of BEM4I, an object-oriented approach will be implemented based on the target pragmas provided by the OpenMP 4.5 standard. In connection with the Espreso library. this will allow for large scale BETI simulations across all parallelism layers available.

In addition to the BETI method, a distributed ACA algorithm based on the graph theory [8] is under development. This approach is suitable for the treatment of large bodies made of homogeneous materials.

Conclusion

The numerical experiments performed on various multi- and many-core architectures show that the need for proper vectorisation of scientific codes becomes more important with the increasing size of SIMD registers. Both full and ACA-sparsified assembly of BEM matrices performs well on these architectures and our aim is to continue in incorporating more vectorised routines in the library.

With regard to the architecture of upcoming heterogeneous supercomputing systems involving some type of many-core accelerators, a more concentrated effort will be put into the development of routines leveraging such hardware.

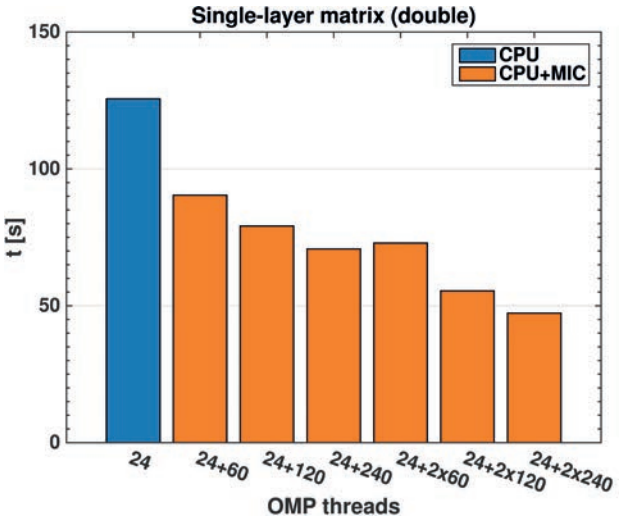


Figure 6.
Assembly times for offload mode with various number of threads on the target

References

[1] Langer Ulrich and Steinbach Olaf. Boundary element tearing and interconnecting methods. Computing, 2003, 71, pp. 205-228. ISSN: 1436-5057. Doi: 10.1007/s00607-003-0018-2

[2] Erichsen Stefan and Sauter Stefan. Efficient automatic quadrature in 3-d Galerkin BEM. Computer Methods in Applied Mechanics and Engineering, 1998, 157, pp. 215-224. ISSN: 0045-7825. Doi: 10.1016/S0045-7825(97)00236-3

[3] Steinbach Olaf. Galerkin- und Kollokations-Diskretisierung für Randintegralgleichungen in 3D. Technical report. Technische Universität Graz, 2004.

Publications

[1] Merta Michal and Zapletal Jan. Acceleration of boundary element method by explicit vectorisation. Advances in Engineering Software, 2015, 86, pp. 70-79. ISSN: 0965-9978.

[2] Merta Michal, Zapletal Jan and Jaros Jiri. Many core acceleration of the boundary element method. High Performance Computing in Science and Engineering: Second International Conference, HPCSE 2015, Soláň, Czech Republic, pp. 116-125. Springer International Publishing, ISBN 978-3-319-40360-1

[3] Veit Alexander, Merta Michal, Zapletal Jan and Lukáš Dalibor. Efficient solution of time-domain boundary integral equations arising in sound-hard scattering. International Journal for Numerical Methods in Engineering, 2016, 107, pp. 430-449. ISSN: 1097-0207.

[4] Zapletal Jan, Merta Michal and Malý Lukáš. Boundary element quadrature schemes for multi- and many-core architectures. Submitted.

[5] Lukáš Dalibor, Kovář Petr, Kovářová Tereza and Merta Michal. A parallel fast boundary element method using cyclic graph decompositions. Numerical Algorithms, 2015, 70, 807-824. Doi:10.1007/s11075-015-9974-9

Project website: bem4i.it4i.cz

03 | EARTH AND ENVIRONMENTAL SCIENCES

HPC FRAMEWORK FOR RAINFALL-RUNOFF UNCERTAINTY MODELLING

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Štěpán Kuchař

Researchers:
Michal Podhorányi, Radim Vavřík, Martin Golasowski

Project ID:
IT4I-10-9

Introduction

Software presented in this text is developed as a part of the automatized online flood monitoring and prediction process used in the Floreon+ disaster management system. This system has to provide accurate results for the current state of the modelled area, while taking into account its hydrological parameters and present meteorological situation. It is crucial to accurately estimate the possible uncertainty of its output and overcome it by using inverse modelling and calibration methods. Some of the dynamic inputs of these models can still contain inaccuracies based on the methods of their acquisition (e.g. weather forecast models [1], rain gauge density and distribution).

These errors are stochastic in nature and have to be analysed using statistical methods. The result of the analysis can be used as a parameter of a selected probabilistic uncertainty modelling method. These methods execute a high number of basic model simulations that can take a lot of computational time. This is in conflict with time constraints of the flood prediction process which has to provide its results as quickly as possible, in a time frame still relevant for early warning [9].

In an effort to overcome the computational complexity, we have created a software framework which allows us to create efficient implementations of the various modelling methods along with their deployment to the HPC infrastructure. As the Floreon+ system supports several models, to lower the impact of their individual weaknesses, an abstraction code layer is created to provide a single interface for all supported models, which is used by the individual methods.

This abstraction can be used to implement calibration and uncertainty modelling methods, what-if analyses or approaches based on input data ensembles. In this text, we present our implementation of an uncertainty simulation of the Rainfall-Runoff (RR) models based on the Monte Carlo method [2,8]. It is tested in experiments on the catchments located in the Moravian-Silesian region of the Czech Republic. Results of these experiments are presented later in this text.

Results and Methods

The uncertainty simulation runs a large number of RR simulations whose input parameters are sampled from their corresponding probability distributions. A more precise result from the simulation can be obtained by executing more simulations. Behaviour of the simulation error vs. number of model runs (samples) is presented in Figure 1. By picking selected p-percentiles from all results of the performed RR simulations, confidence intervals of the simulated water discharge are formed. Each confidence interval represents a boundary in which values of a model output can occur with a certain probability. Obtained confidence intervals are usually combined with the output of a given standard RR simulation to provide additional information about precision of the simulation output.

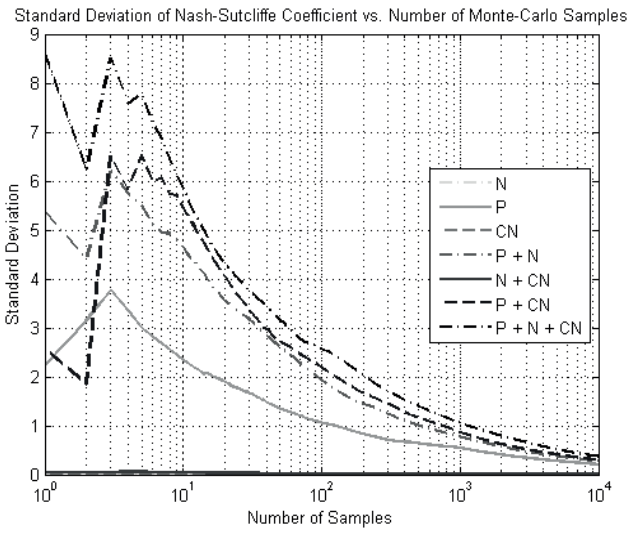


Figure 1.
Uncertainty simulation error convergence

Although the input space of the RR model has several parts, our method mainly focuses on the forecast precipitation data as this parameter has the biggest influence on the model output. The probabilistic properties of the forecast precipitation error can be estimated by comparing the historical forecast data with historical measurements. The probability distribution of the precipitation forecast error was estimated by using a non-parametric kernel density estimation method [3]. This method provides an estimate of the probability density, which is then used for sampling of the precipitation forecast error. Individual sets of the forecast input precipitation are generated by adding the sampled error values to the input precipitation values. An example of the estimated probability distribution is presented in Figure 2. These estimations were obtained by analysis of historical forecast data provided by the ALADIN model [1].

Our approach to uncertainty modelling is universal in the domain of RR modelling, and possibly in other similar domains, therefore it is possible to implement it in the same way for other models of a similar type. This led us to the idea of creating a framework that would allow us to generalise the execution of all supported models, and use these models in pluggable simulators implementing various modelling methods. The methods can be implemented and parallelised independently on the model, and deployed to the HPC cluster. The models themselves can also take advantage of more efficient parallelization, for example by using vectorization and optimization techniques for their specific challenges.

The main objective of the described HPC framework is to create a common interface for initialization, execution and gathering of simulation results for different implementations of the Rainfall-Runoff and hydrodynamic models. These models usually share common input parameters such as catchment schematisation and hydro-meteorological data but their application programming interfaces (API) and implemented modelling methods differ. The framework hides these differences behind a layer represented by the ModelAdapter class as shown in Figure 3.

The uncertainty simulation process can be roughly divided into three stages, which are executed in order. The first stage ensures sampling of the corresponding probability distributions of the uncertain parameters, and creation of the model parameters set. Probability distributions for other model parameters are selected arbitrarily by a hydrology expert. Generated parameter sets are then passed to the specific implementation of the model adapter where each set is converted to an internal representation suitable for the particular model. The second stage of the process is the execution of the model which is being performed for each imported parameter set. The third and final stage of the process is the statistical analysis and GIS post-processing of the simulation results, as shown in Figure 4.

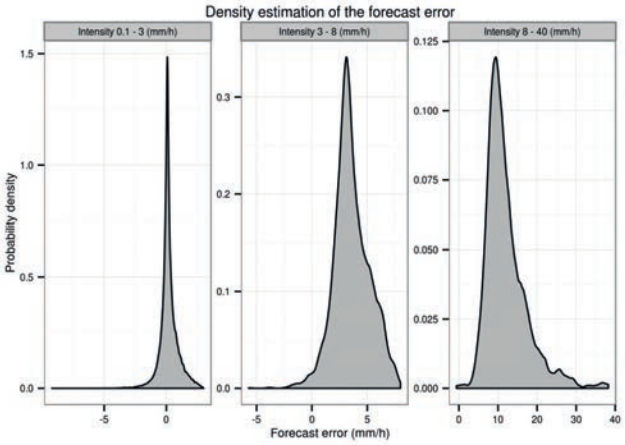


Figure 2.
Estimated probability density of forecast precipitation error

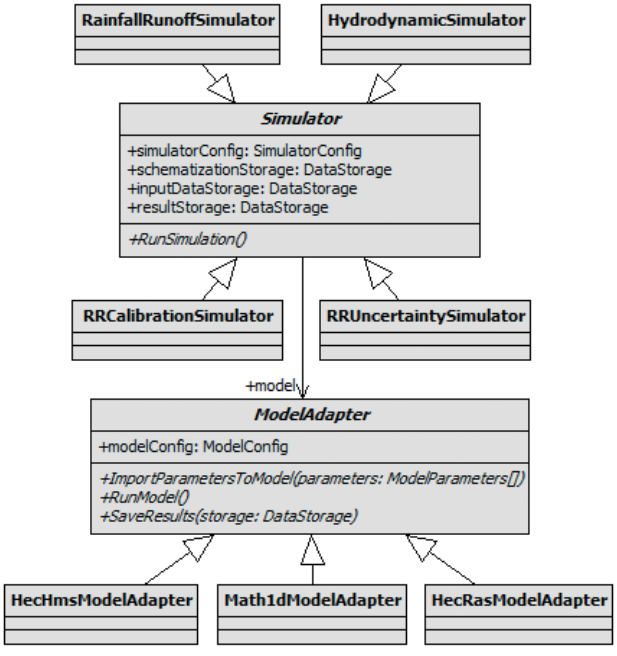


Figure 3.
Class diagram of the developed framework

The multi-node execution of the individual simulations is necessary to combat the computational demands of the simulation and to obtain the result as fast as possible. Each MPI rank generates its own chunk of the parameter sets and runs its respective number of simulations. The individual runs of the model on a single rank can be further parallelised by running multiple model instances in multiple threads, but this option is currently only available for the Math1D model that is executed concurrently 16 times on each node - one model execution for each available CPU core. Scalability of an experimental uncertainty simulation of the Math1D [4,8] model is shown in Figure 5. The separate runs of the HEC-HMS [5] model for a single MPI rank are sequential. After all simulations are finished, the results are converted from a model-dependent representation to an abstract representation recognized by our framework. Then, the results are passed via MPI to the master rank, which selects given percentiles. Gathering of results to the master rank is necessary to sort the results to select proper percentiles for each time step and each river channel. The results of the percentile selection are stored by a specific implementation of the DataStorage class.

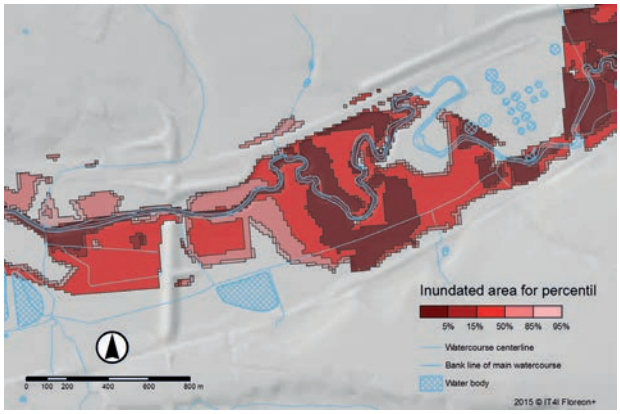


Figure 4.
Result of an uncertainty simulation – inundation map with percentiles

On-going Research / Outlook

Our current work is focused on the implementation of a what-if analysis simulator. This simulator allows the end users to execute a number of different scenarios represented by a set of variations in the model input parameters. The what-if analysis also offers a comparison of the scenarios with the original unchanged simulation result. Moreover, we plan to integrate models from the hydrodynamic simulation domain. This can bring additional challenges in terms of more complex input data sets and specific data formats.

Conclusion

We have developed our own approach to the uncertainty simulation of the RR models. Experimental results of the simulations are presented on a real-world scenario deployed on the HPC infrastructure. During the development, we created a software framework which allows seamless integration of various computationally intensive modelling methods with models coming from the hydrologic simulation domain. Its implementation targets the HPC infrastructure which allows integration in time sensitive applications such as disaster management systems.

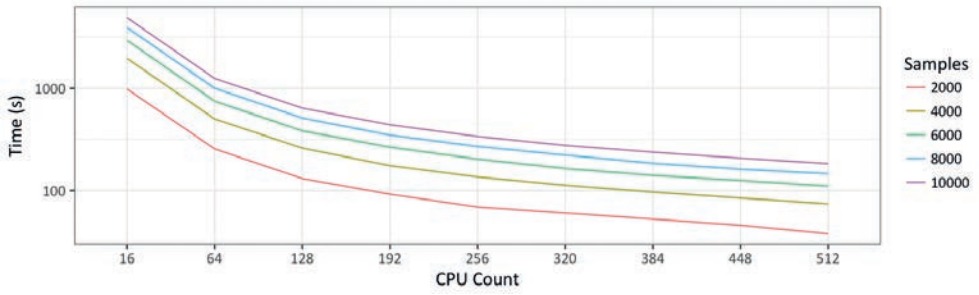


Figure 5.
Scalability of the Math1D rainfall-runoff model uncertainty simulation executed on the Anselm cluster

References

[1] Farda A, et al. Model Aladin as regional climate model for Central and Eastern Europe. Studia Geophysica et geodaetica. 2010, 54(2), 313-332. Doi 10.1007/s11200-010-0017-7
[2] Kroese D. P.; T. Taimre, Z.I. Botev. Handbook of Monte Carlo Methods. John Wiley & Sons, 2013. ISBN 1118014952.
[3] Zucchini, W., Berzel, A., & Nenadic, O. Applied smoothing techniques. Part I: Kernel Density Estimation. Philadephia, PA: Temple University, 2003, pp. 15–19.
[4] Golasowski M.; M. Podhorányi, Š. Kuchař, J. Martinovič, M. Litschmannová. Uncertainty modelling in Rainfall-Runoff simulations based on parallel Monte Carlo method. Neural network world. 2015, 25(3), 267-286. Doi 10.14311/NNW.2015.25.014
[5] Feldman A. D. Hydrologic modeling system HEC-HMS: technical reference manual. US Army Corps of Engineers, Hydrologic Engineering Center, 2000.
[6] Beven K. J. Rainfall-Runoff Modelling - The Primer. Wiley-Blackwell, 2nd edition, 2012. ISBN: 978-0-470-71459-1

Publications

[7] Vondrak V.; S. Kuchar, M. Golasowski, R. Vavrik and J. Martinovič (2015). Scalable Algorithms for Inverse and Uncertainty Modelling in Hydrology. In Computational Techniques for Civil and Structural Engineering. J. Kruis, Y. Tsompanakis and B.H.V. Topping (editors), Saxe-Coburg Publications, Stirlingshire, Scotland, 2015. Doi 10.4203/csets.38.20
[8] Golasowski M.; M. Podhorányi, Š. Kuchař, J. Martinovič, M. Litschmannová. Uncertainty modelling in Rainfall-Runoff simulations based on parallel Monte Carlo method. Neural network world. 2015, 25(3), 267-286. Doi 10.14311/NNW.2015.25.014
[9] Kuchar S.; M. Golasowski, R. Vavrik, M. Podhoranyi, B. Sir, J. Martinovic. Using High Performance Computing for Online Flood Monitoring and Prediction. International Journal of Environmental, Ecological, Geological and Geophysical Engineering. 2015, 9(5), 267-272

URBAN CANOPY CLIMATE FORCING ON A REGIONAL SCALE MODELLED BY THE REGCM AND WRF MODELS

Research institution:
Charles University in Prague

Principal investigator:
Peter Huszár

Researchers:
Michal Belda, Jan Karlický

Project ID:
OPEN-7-37

Introduction

There is an important environmental stress cities represent on its surroundings: the direct meteorological forcing of urban canopy on a local and regional scale. In certain regions like central Europe, this forcing is much stronger (often by two orders of magnitude) than the impact on meteorology and climate via the emissions of radiatively active gases and aerosols (Huszar et al. 2014, 2016a,b).

The urban canopy effects have their origin in characteristics of urban surfaces: they are largely covered by artificial objects and are often paved, therefore clearly distinguished from natural surfaces by mechanical, radiative, thermal, and hydraulic properties. These surfaces represent additional sinks and sources of momentum and heat, affecting the mechanical, thermodynamic and hydrological properties of the atmosphere (Lee et al., 2011).

In Huszar et al. (2014), we examined the urban heat island (UHI) and other urban canopy effects using the regional climate model RegCM4 and a single layer urban canopy model (SLUCM) for the 2001-2005 period. In order to examine these effects in a more robust way, including the analysis of the inter-model and inter-annual variability, we extended this study to cover the longer 2001-2010 period and, what is more important, we applied further parameterisations of urban canopy. Namely the CLM (Community Land Surface Model) urban scheme, also included in RegCM4, and finally the urban parameterisations implemented in the Weather Research Model (WRF) have been, along with the "carrying model," applied as well.

It is further expected that in a changing climate these urban canopy effects might be modified, however due to the complex character of urban forcing, it remains unclear as to how they will evolve in a warmer climate, thus using modelling tools is inevitable.

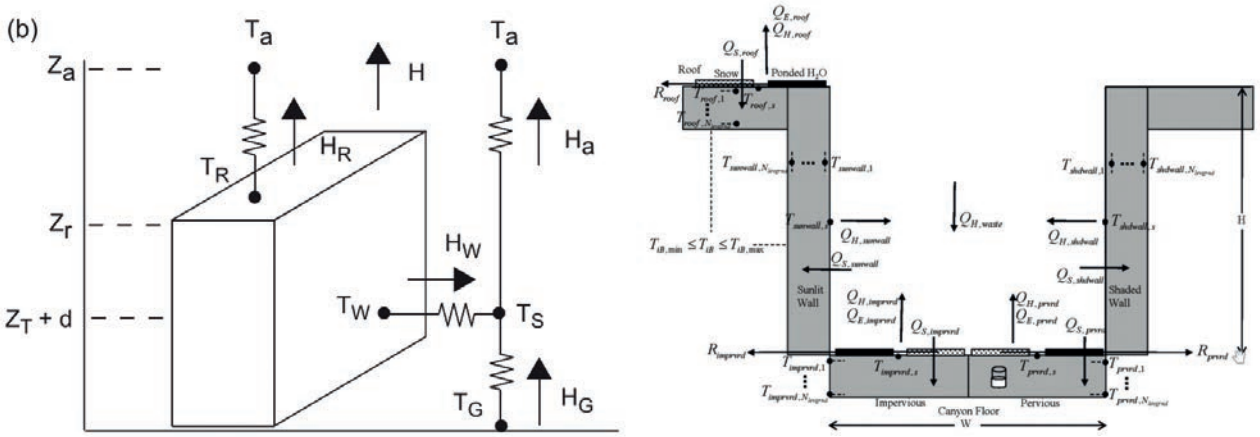


Figure 1.
The schematic representation of the BATS/SLUCM (left) and CLM4.5 urban module (right) with the fluxes of energy considered

Results and Methods

Two models were applied for studying the urban canopy impacts on climate: the Regional Climate Model RegCM4 and the Weather Research and Forecast Model WRF.

The regional climate model used in this study is The International Centre for Theoretical Physics' (ICTP) regional climate model RegCM version 4 (hereafter referred to as RegCM4). It is a three-dimensional mesoscale model, fully described in Giorgi et al. (2012). RegCM4 incorporates two land surface models: BATS (Biosphere-Atmosphere Transfer Scheme and CLM4.5 (Community Land Surface Model) by Oleson et al. (2013). The surface model CLM contains a single layer urban parameterisation scheme the CLM-urban (CLMU), while BATS did not originally contain any special urban canopy treatment. Therefore we included the Single-Layer urban Canopy Model (SLUCM), originally developed by Kusaka et al. (2001). We implemented SLUCM into RegCM4 as a sub-layer of the BATs surface scheme at 2 km x 2 km subgrid resolution (SUBBATS). Within BATS, SLUCM is called whenever urban land use categories are found in the land use data supplied for RegCM. RegCM4 by default does not consider urban type land use categories. We extracted the urban land use information from the EEA Corine 2006 database and those parts of the domain where this was not available, the GLC2000 (GLC, 2000) database was used. We considered two categories, urban and suburban.

The urban model implemented into the CLM4.5 surface module is the CLMU (Oleson et al. 2013). Urban areas in CLM are represented by up to three urban land units per gridcell according to density class. The urban land unit is based on the "urban canyon" concept of Oke (1987) in which the canyon geometry is described by building height and street width. The canyon system consists of roofs, walls, and canyon floor. Walls are further divided into shaded and sunlit components. The canyon floor is divided into pervious and impervious fractions. Vegetation is not explicitly modelled for the pervious fraction; instead evaporation is parameterised by a simplified bulk scheme. The schematic representation of the flux interactions between the urban canyon elements within CLM4.5-urban and BATS/SLUCM is shown in Figure 1.

In WRF, there are 4 choices of urban canopy treatment available: the bulk description, the SLUCM model (basically the same as in RegCM), and two multilayer approaches, the BEP and BEM modules (Chen et al., 2011).

The model domain chosen for the experiments cover the "larger" Central European area (Figure 2) with the horizontal resolution of 10 km x 10 km (160x120x23) gridboxes in horizontal/vertical. We modelled 2001-2010, where the meteorological boundary conditions were taken from the ERA-interim reanalysis data. For the selected period, simulations with and without considering urban canopy were carried out to separate the effect of urban surfacs

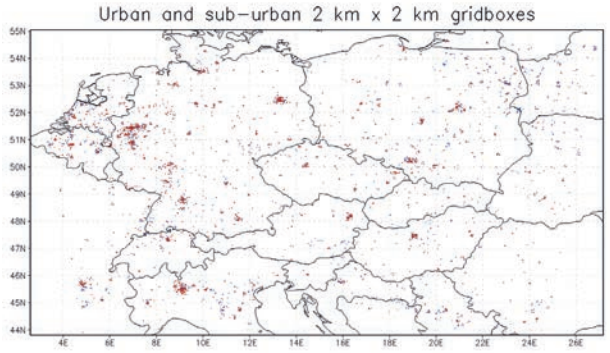


Figure 2.
The computational domain with urban and suburban gridboxes indicated

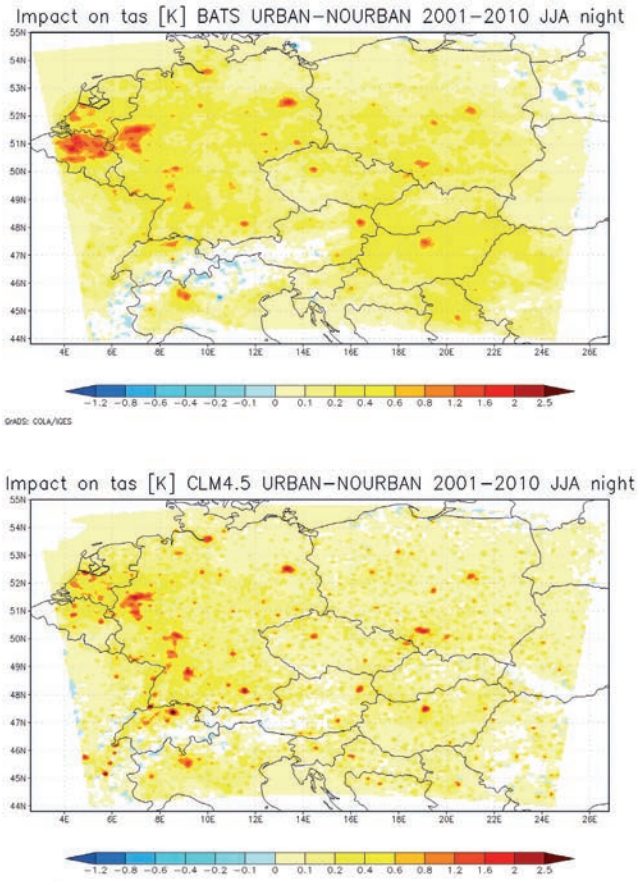


Figure 3.
Urban canopy impacts on the near surface temperature in Kelvins (upper figure – BATS/SLUCM module, lower figure – CLM4.5/CLMU module). Shaded areas represent statistically significant changes on the 95% level.

in the models. Calculations for the RegCM4 model were performed on the Salomon cluster, while WRF was run on Anselm. Results showed a statistically significant and strong impact of urban canopy on surface temperature in all model experiments. The warming pattern resembles the location of large cities and can reach 2.5K for night-time (see Figure 3 for RegCM and Figure 4 for WRF). Strong impacts are modelled consistently within all model set-ups also for the height of the planetary boundary layer and humidity (not presented here).

On-going Research / Outlook

Apart from the experiments shown above, we performed runs for future climate conditions, namely 2046-2050 and 2090-2099. These decades are currently analysed for the purpose of quantifying the possible evolution of the urban canopy meteorological effects in a changing climate. Furthermore as the listed urban parameterisation depend on a large number of parameters (e.g. those describing the urban geometry), we are currently performing sensitivity runs of these effects on the choice of parameters.

Conclusion

The computer capabilities of IT4Innovations clusters offered us the possibility to perform long term climate model experiments in order to investigate the stress that urban surfaces represent on the atmosphere. We showed (among others) that there is a large impact especially on temperature. If we consider that the climate change and the associated warming over the region in focus have a similar magnitude, we can conclude that urbanisation can significantly contribute to the observed temperature increase. In conclusion, urban canopy effects have to be described in future models (as parameterisations, sub-models, etc.) as neglecting the urban canopy impact on the climate can lead to large errors and incorrect conclusions.

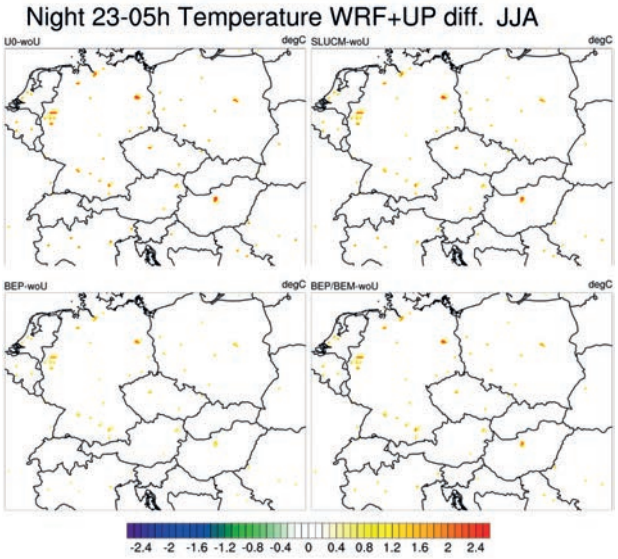


Figure 4.
Urban canopy impacts on the near surface temperature in Kelvins from the WRF model. Shaded areas represent statistically significant changes on the 95% level.

References

[1] Chen, F., et al.: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems, Int. J. Climatol. 31, 273-288, 2011. Doi: 10.1002/joc.2158

[2] Giorgi, F., et al., 2012: RegCM4: Model description and preliminary tests over multiple CORDEX domains, Clim. Rev. 52, 7-29. Doi: 10.3354/cr01018

[3] Huszár, P., Halenka, T., Belda, M., Zak, M., Sindelarova, K., and Miksovsky, J.: Regional climate model assessment of the urban land-surface forcing over central Europe, Atmos. Chem. Phys. 2014, 14, 12393-12413. Doi 10.5194/acp-14-12393-2014

[4] Huszár, P., Belda, M., and Halenka, T.: On the long-term impact of emissions from central European cities on regional air quality, Atmos. Chem. Phys. 2016a, 16, 1331-1352. Doi 10.5194/acp-16-1331-2016

[5] Huszár, P., Belda, M., Karlický, J., Pišoft, P., and Halenka, T.: The regional impact of urban emissions on climate over central Europe: present and future emission perspectives, Atmos. Chem. Phys. 2016b, 16, 12993-13013. Doi 10.5194/acp-16-12993-2016

[6] Kusaka, H. et al.: A simple single layer urban canopy model for atmospheric models: comparison with multilayer and slab models. Bound Layer Met, 2001, 101: 329-358.Doi: 10.1023/A:1019207923078

[7] Lee, S.-H. Et al.: Evaluation of urban surface parameterisations in the WRF model using measurements during the Texas Air Quality Study 2006 field campaign, Atmos. Chem. Phys. 2011, 11, 2127-2143. Doi 10.5194/acp-11-2127-2011

[8] Oke, T. R.: Boundary Layer Climates, 2nd edn., Routledge, 1987.

[9] Oleson, K.W., et al.: Technical Description of version 4.5 of the Community Land Model (CLM). Ncar Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, 2013, 422 pp. Doi 10.5065/D6RR1W7M

Publication

[1] Karlicky, J., Huszar, P. and M. Belda A multi-model assessment of the urban land-surface impact on regional climate over central Europe. In preparation, 2017.

PARALLELIZATION OF FLOOD MODEL CALIBRATION

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Štěpán Kuchař

Researchers:
Michal Podhorányi, Radim Vavřík, Martin Golasowski, Vít Vondrák

Project IDs:
IT4I-5-7, IT4I-6-3

Introduction

Successful decision making in disaster management needs to have very accurate information about disaster phenomena and its potential development in time. Rainfall-runoff simulations are an integral part of flood warning and decision making processes. To increase their accuracy, it is crucial to periodically recalibrate their parameters. It is necessary to use a High Performance Computing (HPC) facility as calibration is a very time-consuming process, requiring speeding-up. However, the required speed-up can only be achieved by avoiding any human-computer interaction in so-called automatic calibration. In order to develop the capability of the Floreon+ online hydrologic, environmental and traffic modelling system [1], appropriate methods were created, tested and parallel automatic calibration software was developed.

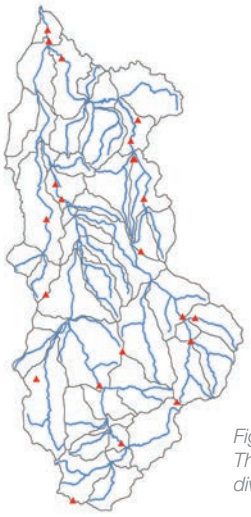


Figure 1.
The Ostravice basin (catchment) divided into sub-basins

Results and Methods

We used our in-house numerical semi-distributed rainfall-runoff model Math1D to divide the whole simulated area (basin) into smaller areas (sub-basins) for testing and development. An example of such schematization is shown in Figure 1, where blue lines represent river segments, grey lines represent boundaries of individual sub-basins and red points are hydrologic and/or meteorologic gauges that measure precipitation and/or river discharge.

The Math1D model is based on a modification of Unit Hydrograph and SCS-CN methods [2] to compute transformation of rainfall to effective runoff and its distribution in time. The contribution to a sub-basin outlet from river segments is computed using the kinematic wave approximation based on Manning's formula [3].

The simulation models depend on their input parameters. The correct setup and further adjustment of those parameters, so-called model calibration, is crucial for accurate simulation results. In the case of rainfall-runoff modelling, there are many hydrologic parameters that have a significant influence on the model. We identified two of the most significant - Runoff Curve Number (CN) for the SCS-CN runoff method and Manning's Roughness Coefficient for the kinematic wave approximation method.

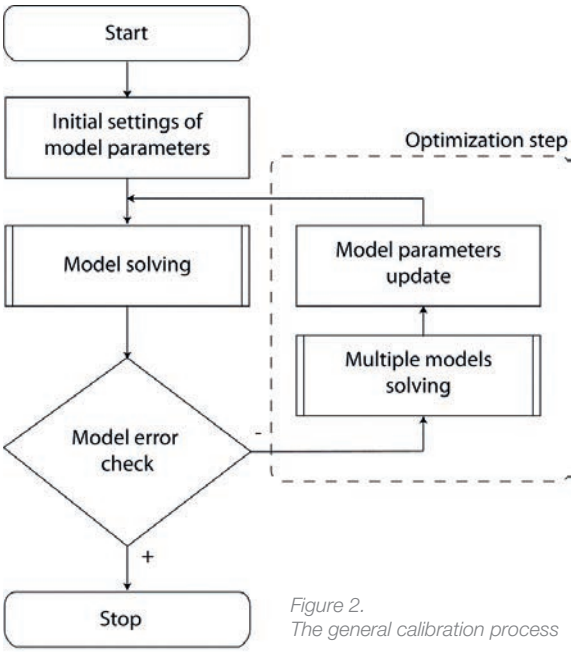


Figure 2.
The general calibration process

The calibration process depicted in Figure 2 can be described by the following minimization procedure:

- Set the initial parameters of the model.
- Run a simulation.
- Evaluate the model error using a combination of metrics. If the error is less than a set tolerance, then stop.
- Perform an optimization algorithm step comprising many simulation runs to obtain new calibration parameter values.
- Update calibration parameters of the model with the new values and continue from step 2.

We have implemented six common local (Linesearch, Interior point) and global (Pattern search, Global search, Multi-start, Genetic algorithm) optimization methods, and compared them using numerical experiments and statistical analysis. Several combinations of metrics focused on calibration efficiency, including an adapted Nash-Sutcliffe model efficiency coefficient (1), Normed peak error (2) and Normed volume error (3) were used to evaluate the accuracy of our model.

(1) $$NSC = \frac{\sum_{t=1}^{T}(Q_o(t)-Q_m(t))^2}{\sum_{t=1}^{T}(Q_o(t)-\overline{Q_o})^2},$$

(2) $$NPE = \frac{|Q_o(t_p)-Q_m(t_p)|}{Q_o(t_p)},\ where\ t_p = arg\max_{t\in\langle 0,T\rangle}(Q_o(t)),$$

(3) $$NVE = \frac{\sum_{t=1}^{T}|Q_o(t)-Q_m(t)|}{\sum_{t=1}^{T}|Q_o(t)|},$$

where Q0(t) denotes observed discharge at time t and Qm(t) denotes modelled discharge at time t.
The statistical results of the comparison are presented in Table 1 and the parallel scalability of the methods is shown in Figure 3. Based on a simple and efficient parallel implementation, good scalability and acceptable rate of convergence, we chose Multi-start and Linesearch methods as global and local minimization algorithms, respectively.
These methods were used for the development of the calibration module of the HPC execution framework for rainfall runoff model calibration and uncertainty modelling.
The framework was implemented in C/C++ using a hybrid parallel programming model consisting of MPI communication for multi-node execution and OpenMP (OMP) for multi-core support aiming for optimal utilization of computing resources. The scalability of the calibration process is ensured by distributing the parallelization into several levels as shown in Figure 4.
For the combined Multi-start Linesearch method used in our approach, the process begins with the sampling of initial values of calibrated parameters, which will be used as starting points (SP) for the individual Linesearch calibrations. The initial value of each parameter is sampled randomly from the uniform distribution with its minimal and maximal values constrained by the physical limits of the parameter. The number of generated SP is configured by a user. The total number of SP can be divided into chunks and their processing, including generation of parameters, can be effectively parallelized by MPI on multiple nodes.
The next step in the algorithm is to obtain an initial error value for each SP. This value is obtained by running a simulation with the sampled parameter values and comparing its results with measured values.
Each MPI process further divides its chunk of SP between a defined number of OMP threads. This is possible due to the fact, that optimization runs (calibrations) for separate SP are independent. The next level of parallelization by OMP threads is allowed by an internal implementation of the optimization loop. It is limited by the number of sub-basins in the schematization for which separate model simulations can be executed in parallel. Also, the rainfall-runoff model itself can be parallelized if it is allowed by the model implementation. Currently, the Math1D model allows several OMP threads to run concurrently in one node.

Optimization methods	$\|NSC\|_2$			$\|NSC\|_\infty$		
	Mean	Median	SD	Mean	Median	SD
Linesearch	0.811	0.855	0.183	0.796	0.848	0.215
Interior point	0.852	0.864	0.099	0.848	0.860	0.105
Pattern search	0.892	0.907	0.080	0.882	0.904	0.097
Global search	0.802	0.840	0.159	0.806	0.826	0.145
Multistart	0.677	0.769	0.304	0.637	0.776	0.337
Genetic alg.	0.768	0.772	0.132	0.751	0.755	0.152

Table 1.
Statistical results for improvement of Nash-Sutcliffe model efficiency coefficient

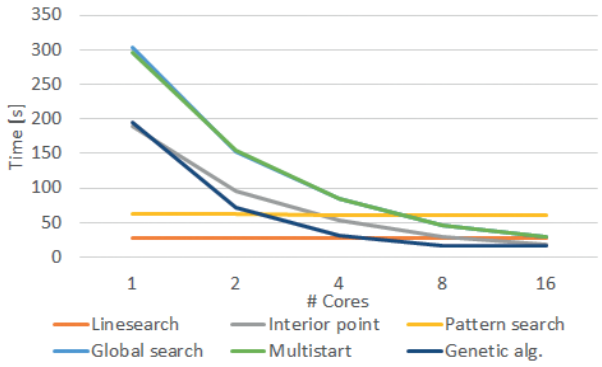


Figure 3.
Scalability of calibration methods

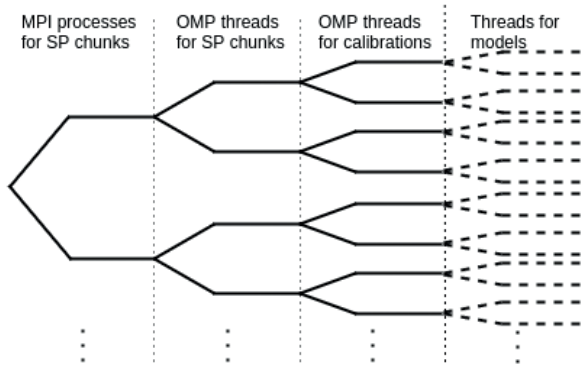


Figure 4.
Parallelization levels of the rainfall-runoff model calibration algorithm

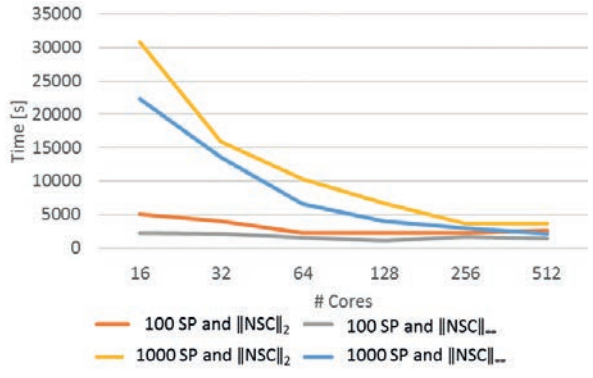


Figure 5.
Scalability of rainfall-runoff model calibration with 100 and 1000 starting points

Actual calibration of the model is then performed in the optimization loop which is executed as the next step of the algorithm. The loop runs until a change in the error value for the current iteration is below a certain threshold. However, the calibration process for a high level of precision (small thresholds) can be a time-consuming task, so we have also introduced a time constraint to the loop condition. If the change in error value is still above the threshold after a given amount of time, the optimization loop ends. Current parameter values are then passed as the best optimized values in a given time span.

After all the SP in one chunk finish their optimization, the best parameter set of the node is determined by the local lowest error value. Those values are then gathered from all running MPI processes (nodes) and compared to each other. Finally, the parameter values that produced the global lowest error are selected and the model is recalibrated with these parameters.

With the calibration software implementing the described algorithm, we performed two types of numerical experiments. The aim of the first type was to verify the scalability of the implemented algorithms. Figure 5 presents the experiment on 100 and 1000 randomly generated but constrained starting points (SP). These points were divided into chunks and uniformly distributed using MPI to 1, 2, 4, 8, 16 and 32 processes. During these experiments, the calibration process searched for the minimum value of the objective function within a set threshold, without any other computation limitations. We measured the execution time of the calibration with the Nash-Sutcliffe model efficiency coefficient combined with L2-norm and L∞-norm as a metric of the model accuracy. An improvement of the calibration against the non-calibrated results was comparable in all the experiments.

The second type of experiments was calibration with a strictly declared time limit for the calibration execution time. This constraint is very useful for cases when the calibration has to be run before every simulation to provide the best possible

Time limit [s]	600			
SP	N		2N	
CPU cores (N)	Imp^2	Imp^∞	Imp^2	Imp^∞
16	0.520	0.595	0.502	0.664
32	0.583	0.662	0.538	0.604
64	0.592	0.680	0.548	0.671
128	0.611	0.687	0.603	0.654

Table 2.
Improvement of rainfall-runoff model calibration with a time limit

results in a short time frame (e.g. during critical situations). We limited the total calibration time to 10 minutes and tried two configurations of starting points – 16 and 32 starting points per MPI rank. As each MPI rank is only able to run 16 calibrations at the same time, the time limit for each of the 32 calibrations had to be halved (i.e. set to 5 minutes) to fit in the overall time limit. The final improvement of the calibration with these limitations for both norms executed on a different number of nodes is presented in Table 2.

All numerical experiments were executed on the Anselm HPC cluster at IT4Innovations.

On-going Research / Outlook

Future research into rainfall-runoff model calibration will be focused on enabling other rainfall-runoff models including HEC-HMS model in the HPC execution framework. Another aim will focus on optimization of execution runtimes and efficiency of the calibration process, and particular optimization methods.

Conclusion

In this text, the rainfall-runoff model calibration process was introduced and several phases of the development were overviewed. Using the calibration software implemented according to the selected methods and calibration algorithm, two types of experiments were carried out. The first one showed that the solution scales properly with the increasing number of starting points and also improves the precision of the results. The second experiment showed that even with limited time, a significant improvement of the calibration can be achieved.

References

[1] Martinovic, J.; S. Kuchar, I. Vondrak, V. Vondrak, B. Sir and J. Unucka. Multiple Scenarios Computing in The Flood Prediction System FLOREON. In 24th European Conference on Modelling & Simulation, Malaysia, June 2010, pp. 182-188. Doi: 10.7148/2010-0182-0188

[2] Beven, K. J. Rainfall-Runoff Modelling: The Primer. John Wiley & Sons, 2012. E. W. Wohl, Inland Flood Hazards. Cambridge, Cambridge University Press, 2000.

[3] Vieux, B. E. Distributed Hydrological Modeling using GIS. 4th edition. Addison-Wesley Professional, May 2013.

Publications

[1] Vavrik, R., Theuer, M., Golasowski, M., Kuchar, S., Podhoranyi, M. and Vondrak. Automatic calibration of rainfall-runoff models and its parallelization strategies. AIP Conference Proceedings. 2015, 1648, 830014-. Doi: 10.1063/1.4913040.

[2] Vondrak V., Kuchar S., Golasowski M., Vavrik V. and J. Martinovic. Scalable Algorithms for Inverse and Uncertainty Modelling in Hydrology. In J. Kruis, Y. Tsompanakis and B.H.V. Topping, (Editors), Computational Techniques for Civil and Structural Engineering, Saxe-Coburg Publications, Stirlingshire, UK, Chapter 20, pp 467-486, 2015. Doi:10.4203/csets.38.20

[3] Theuer M., Vavrik R., Vondrak V., Kuchar S., Sir B. and A. Portero. Efficient methods of automatic calibration for rainfall-runoff modelling in the Floreon+ system. Submitted to Neural Network World.

HYDROLOGICAL MODEL REMOTE EXECUTION AND HPC AS A SERVICE

Research institution:
IT4Innovations National Supercomputing Center

Principal investigators:
Jan Martinovič, Štěpán Kuchař, Václav Svatoň, Vít Vondrák

Researchers:
Lukáš Vojáček, Martin Golasowski, Aleš Ronovský

Project partners:
DHI
David Bezděk, Thomas Bech, Johan Hartnack, Jesper Carlson, Ole Rene Sørensen

Project IDs:
IT4I-1-18, IT4I-2-7, IT4I-3-7, OPEN-8-21

Introduction

Risk and uncertainty have become fundamental aspects of most decision making processes. Qualitative risk assessment is usually based upon perceptions, opinions, judgment, public or political consensus (or a combination thereof), whereas quantitative risk assessment is based on the results of modelling techniques. Modelling is used to express various components of the risk while considering the probability of whether loss or damage will occur.

The analysis of disaster risk phenomena often requires modelling using on-line monitored data. In this context, an important aspect is the relation between computing time and real time of the modelling run. Powerful personal computers struggle to run the fine scale models fast enough to get results in time. The results are needed for prediction and forecast but mainly for the response phase.

"HPC as a Service" is a new area in hydrologic modelling that enables running hydrologic models remotely on HPC clusters to significantly decrease their computing time. The near real-time results from hydrologic simulations may be used to support decision making in critical situations and to provide suggestions for preventive actions in reducing the risks and magnitude of future critical situations. Hydrologic model parallelization and remote execution on HPC also allows users to increase the precision of results by increasing the computing mesh resolution while maintaining manageable runtime.

HPC as a Service further lowers the entry barrier for users who are interested in utilizing massive parallel computers for modelling. Through this service, small and medium enterprises can take advantage of the technology without an upfront investment in hardware.

Results and Methods

The proposed solution consists of two main parts that create the whole simulation process – an HPC as a Service framework that ensures execution of models on the HPC infrastructure and a parallelized MIKE21 FM model that utilizes the HPC infrastructure for 2D hydrodynamic simulations.

HPC as a Service framework

To provide the possibility to execute simulations on the HPC infrastructure using hydrologic applications, and for automated flood monitoring and prediction systems, we have developed the HPC as a Service framework. The framework enables integration of HPC computations with internal and external applications. The HPC as a Service framework unifies access to different HPC systems through a simple object-oriented client-server interface using standard web services. It also ensures the availability of required supporting functions such as job management, monitoring and reporting, user authentication/authorization, result file transfer,encryption, and various notification mechanisms. The framework brings ideas from cloud infrastructures to high performance computing, while still guaranteeing high quality and accessibility of service along with much higher and predictable computational performance. The latter is ensured by the underlying supercomputing infrastructure.

An overview of the HPC as a Service framework architecture is shown in Figure 1.

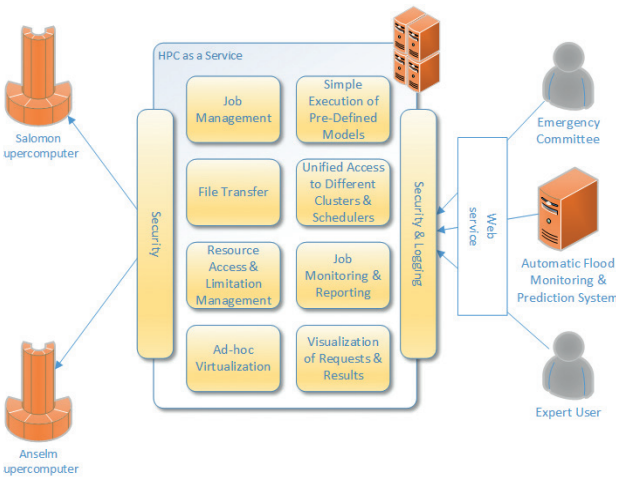


Figure 1.
Overview of the HPC as a Service framework architecture

MIKE21 FM

DHI's Flexible Mesh (FM) modelling system, which was developed for applications within ocean, coastal, estuarine, and riverine environments, was parallelized based on the message passing paradigm to enable utilization of distributed-memory systems. This extension makes it possible to utilize massively parallel computers and clusters. The distribution of work and data is based on the domain decomposition concept, and Message Passing Interface (MPI) is used for communication between processors. The computational mesh is partitioned into several physical subdomains, and the work associated with each subdomain is processed by an individual processor. The data exchange between processors is based on the halo-layer approach with overlapping elements.

The use of massive parallel computers makes it possible to use the MIKE 21 Flow Model FM (see DHI [1]) for real-time flood inundation forecasting and calculations of high resolution two-dimensional (2D) flow solutions over irregular floodplain topography. The MIKE 21 Flow Model solves the 2D shallow water equations. The spatial discretization is performed using a cell-centered finite volume technique and an unstructured mesh approach. An approximate Riemann solver (Roe's scheme, see Roe [2]) is used to calculate the convective fluxes. Second-order spatial accuracy is achieved by employing a linear gradient-reconstruction technique. The average gradients are estimated using the approach by Jawahar et al. [3]. To avoid numerical oscillations, a second-order total variation diminishing (TVD) slope limiter (Van Leer limiter, see Hirsch [4]) is used. An explicit Euler method or a second-order Runge-Kutta method is used for the time integration.

Performance of MIKE21 FM

The MIKE21 FM simulations were executed using HPC as a Service architecture on several standard compute nodes on the Anselm HPC cluster to see how the problem and models scale on a different number of cores. This model was tested on a processor range of between 16 and 2048 cores and the resulting speedup is shown in Figure 2. (The speedup experimentation was done for a 2D hydrodynamic model simulating the tides in the Mediterranean Sea.)

The results show that the speedup is linear to the number of cores for up to 512 cores used. For higher numbers of cores, the speedup shows a slower increase due to the high amount of communication between individual processes and the relatively small size of each subdomain of the mesh. Hence the greater the number of elements the more the scalability will be aligned with the ideal red line.

To achieve a good speedup, it is important that the subdomains have enough wet elements to calculate. Dry elements use negligible calculation time and a full subdomain of dry elements would be idle most of the calculation time. To get optimal performance, the subdomains should have an even distribution of the wet elements. The inherent changes in the number of dry/wet elements in the flood forecast models can cause an ideal partitioning to become suboptimal during subsequent simulated time steps.

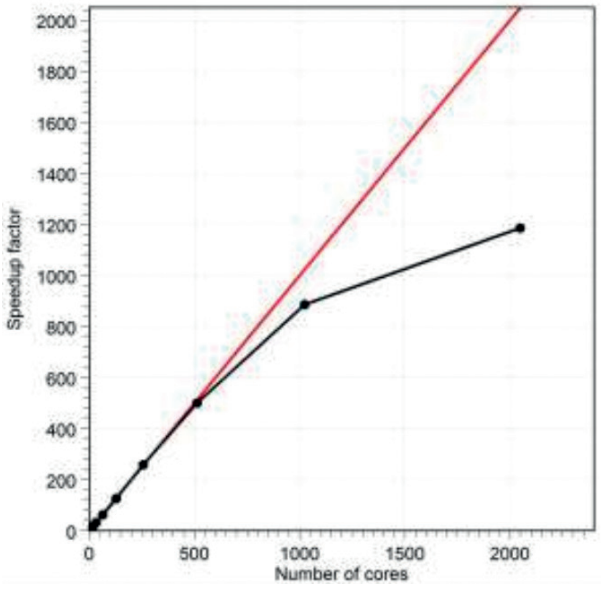


Figure 2.
Scalability of MIKE21 FM model

On-going Research / Outlook

Currently, our strongest focus is on finalizing the production version of the remote execution of hydrologic models and to successfully develop it into a functional system for users of MIKE Powered by DHI models. The proposed HPC as a Service framework is not limited to hydrology and flood prediction, but can also be applied in other domains. We are currently expanding it to different areas like air pollution modelling, traffic modelling and even earth observation analysis for urban environment modelling.

In the future, the HPC as a Service middleware could also serve as an integration platform for dynamic distribution of jobs to different HPC clusters based on their utilization and supported functions.

Conclusion

The HPC as a Service architecture provides a generic framework for submitting hydrological and other simulations to the HPC clusters, managing their data, results, limitations, users and statuses. This enables the running of different types of hydrological simulations on the HPC cluster that will lead to shortening the time needed to complete and get results from such simulations. This is very useful for users like fire departments and flood warning committees during critical disaster situations when fast simulation results and accurate information are essential for making good decisions, and implementing countermeasures appropriate for the situation.

References

[1] DHI, “MIKE 21 & MIKE 3 FLOW MODEL FM, Hydrodynamic and Transport Module, Scientific Documentation”, DHI, (2013).

[2] Roe P. L. Approximate Riemann solvers, parameter vectors, and difference schemes. Journal of Computational Physics. 1981, 43, 357-372. Doi: 10.1016/0021-9991(81)90128-5

[3] Jawahar P. and Kamath H. A high-resolution procedure for Euler and Navier-Stokes computations on unstructured grids. Journal Comp. Physics. 2000, 164, 165-203. Doi: 10.1006/jcph.2000.6596

[4] Hirsch, C. Numerical Computation of Internal and External Flows. Computational Methods for Inviscid and Viscous Flows, Wiley. 1990, vol. 2. ISBN: 978-0-471-92452-4

Publication

[1] Zeman E., Tachecí P., Štrof P., Hartnack J.N., Sorensen O.R., Carlson J., Vondrák V., Martinovič J. and Š. Kuchař. HPC for DRM - Operational Flood Management in Urban Environment. 11th International Conference on Hydroinformatics, HIC 2014. August 2014, New York, USA. ISBN: 978-1-5108-0039-7

LARGE-SCALE GEOSIMULATIONS USING GEM AND TRILINOS

Research institution:
Institute of Geonics of the Czech Academy of Sciences

Principal investigator:
Jiří Starý

Researchers:
Radim Blaheta, Ondřej Jakl, Rostislav Hrtus, Alexej Kolcun, Roman Kohut, Erhan Turan

Project ID:
IT4I-1-20

Introduction

The project deals with efficient solution of singular symmetric positive semidefinite problems. Our motivation arises from the need to solve special problems of geotechnics, e.g. to perform upscaling analysis of geocomposites. In that and other applications we have to solve boundary problems with pure Neumann boundary conditions. We show that the stabilised PCG method with various preconditioners is a good choice for systems resulting from the numerical solution of Neumann problems, or more generally problems with a known small dimensional null space. The work was also taken as an opportunity to compare parallel implementations of the corresponding solvers, namely implementations in the in-house finite element software GEM and implementations employing components of the general Trilinos library. The studies show that the solvers based on GEM are highly competitive with its recognised counterpart.

Results and Methods

High resolution FEM systems

In analysis of elastic properties of geocomposites [4], the sample Ω representing a cube with a relatively complicated microstructure. The FEM mesh is constructed on the basis of CT scans. As benchmarks, we shall use FEM systems arising from CT scanning of coal-resin geocomposites at the CT-lab of the Institute of Geonics. The characteristics of three benchmarks can be seen in Table 1.

Benchmark	Discretisation	Size in DOF	Data size
GEOC-1	232 x 232 x 38	6,135,936	1.0 GB
GEOC-2s	257 x 257 x 57	50,923,779	8.5 GB
GEOC-2l	257 x 257 x 1025	203,100,675	33.5 GB

Table 1.
Benchmarks represent microstructures of three geocomposite samples. Notation, discretisation meshes and sizes of resulting linear systems.

The elastic response of the sample (a representative volume) Ω is characterised by homogenised elasticity C or compliance S tensors ($S = C^{-1}$), determined from the relations

$$C\langle\varepsilon\rangle = C\varepsilon_0 = \langle\sigma\rangle \quad \text{and} \quad S\langle\sigma\rangle = S\sigma_0 = \langle\varepsilon\rangle$$

respectively. Here ⟨σ⟩ and ⟨ε⟩ are volume averaged stresses and strains computed from the solution of elasticity problem

$$-\mathrm{div}(\sigma) = 0, \quad \sigma = C_m\varepsilon, \quad \varepsilon = (\nabla u + (\nabla u)^T)/2 \quad \text{in } \Omega$$

with Dirichlet and/or Neumann boundary conditions,

$$u(x) = \varepsilon_0 \cdot x \quad \text{on } \partial\Omega \quad \text{and} \quad \sigma \cdot n = \sigma_0 \cdot n \quad \text{on } \partial\Omega,$$

respectively. Above, σ and ε denote stress and strain in the microstructure, Cm is the variable local elasticity tensor, u and n denote the displacement and the unit normal, respectively. The use of pure Dirichlet and pure Neumann boundary conditions allows us to get upper and lower bounds for the upscaled elasticity tensor, see e.g. [4].

Iterative solvers

The domain is discretised by linear tetrahedral finite elements. The arising singular system is then solved by stabilised PCG method implemented in different software and using various preconditioners:

GEM-DD is a solver fully implemented in GEM software. It uses the one-level additive Schwarz domain decomposition preconditioner with subproblems replaced by displacement decomposition incomplete factorisation described in [2]. The resulting preconditioner is symmetric positive definite.

GEM-DD-CG solver differs in preconditioning, which is a two-level Schwarz domain decomposition arising from the previous GEM-DD by additive involvement of a coarse problem correction. The coarse problem is created by a regular aggregation of 6 × 6 × 3 nodes with 3 DOF's per aggregation. In this case, the coarse problem is singular with a smaller null space containing only the rigid shifts. The coarse problem is solved

only approximately by an inner (not stabilised) CG method with a lower solution accuracy – relative residual accuracy $\varepsilon 0 \leq 0.01$.

Trilinos ILU is a solver running in Trilinos, where the system from GEM is imported. The preconditioner is similar to GEM-DD, i.e. one-level Schwarz with the minimal overlap and working on the same subdomains as in GEM-DD are used. The subproblems are replaced by ILU without displacement decomposition, using a drop tolerance and a fill limit.

Trilinos ML-DD is again running in Trilinos and uses a multilevel V cycle preconditioner exploiting smoothed aggregations with aggressive coarsening. Six DOF's translational plus rotational are used per aggregation. ILU is applied as a smoother at the finest level, other smoothing is realised by symmetrised Gauss-Seidel. The coarsest problem is solved by a direct solver.

Note that in the computational experiments described in the next section, we solve the problems with pure Neumann boundary conditions.W

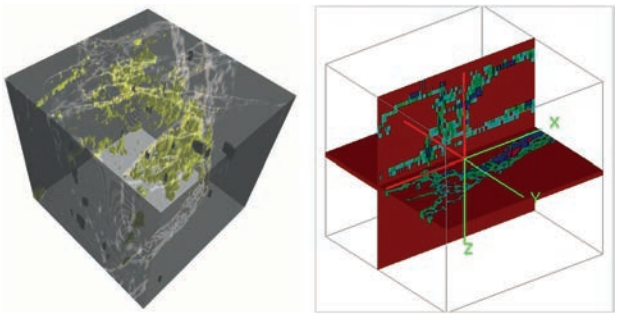


Figure 1.
Inner microstructure of a geocomposite material

Parallel computers

The computational experiments are performed on:

Enna – 64-core NUMA multiprocessor at the Institute of Geonics CAS: eight octa-core Intel Xeon E7-8837 / 2.66 GHz processors, 256 GB of DDR2 RAM, CentOS 6.3, Intel Cluster Studio XE 2013.

Anselm – multicomputer (cluster) with 207 compute nodes at the IT4Innovations National Supercomputing Center. We employed the computing nodes equipped with: two octa-core Intel E5-2665 / 2.4 GHz processors, 64 GB of memory and 500 GB of local disk capacity, Infiniband QDR interconnection, fully non-blocking, fat-tree, Bullx Linux OS (Red Hat family), Intel Cluster Studio XE 2013.

Computational experiments

The following tables summarise the results of computational experiments. For various GEM and Trilinos solvers running on both the computer platforms, the table columns report the number of subdomains (# Sd), iteration counts (# It), wall-clock times of the solution (Titer, in seconds) and the corresponding performance ratio Anselm/Enna (A/E). The stopping criterion was always $\| r \| / \| b \| \leq 10\text{-}5$.

Table 2 gives the performance comparison of the GEM and Trilinos solvers processing the smallest benchmark GEOC-1. The timings shows that both GEM solvers are much faster (4-1.5x) than their Trilinos counterparts, but less scalable. The involved coarse grid problem speeds up GEM-DD by a factor of 1.5-3.

Table 3 lists the timings of GEM solvers (without and with coarse grid problem applied) obtained for GEOC-2s, i.e. a problem of more than 50 million DOF's, where the performance up to 64 processing elements on Enna and up to 128 processing elements on Anselm could be compared. DD-CG solver made use of a coarse problem with aggregation factors 9 x 9 x 9 (81,000 DOF's).

For greater numbers of subdomains, the results confirm the advantage of systems with distributed memory, when the multiprocessors in general suffer from the memory-processor bandwidth contention. Thus, while on Enna the scalability fades out at about 32 cores, the turning point on Anselm is around 128 processing elements, when the small size of subdomains deteriorates the computation/communication ratio.

In absolute figures, we were able to solve the benchmark 3-4x faster on Anselm than on Enna. The advantage of Anselm can be seen from the fact that its newer Intel Sandy Bridge CPU

# Sd	GEM				Trilinos			
	DD		DD + CG		ILU (IFPACK)		ML-DD (ML)	
	# It	T_{iter}	# It	T_{iter}	# It	T_{iter}	# It	T_{iter}
1					345	1943.7	(fail)	
2	293	325.3	137	182.9	472	1442.6	53	734.9
4	302	187.3	124	88.6	436	714.6	61	386.0
8	300	127.6	115	56.5	441	372.6	74	202.2
16	350	149.4	116	54.3	387	219.0	93	127.1

Table 2.
Results of the GEOC-1 benchmark achieved by the various GEM and Trilinos solvers on the multiprocessor Enna

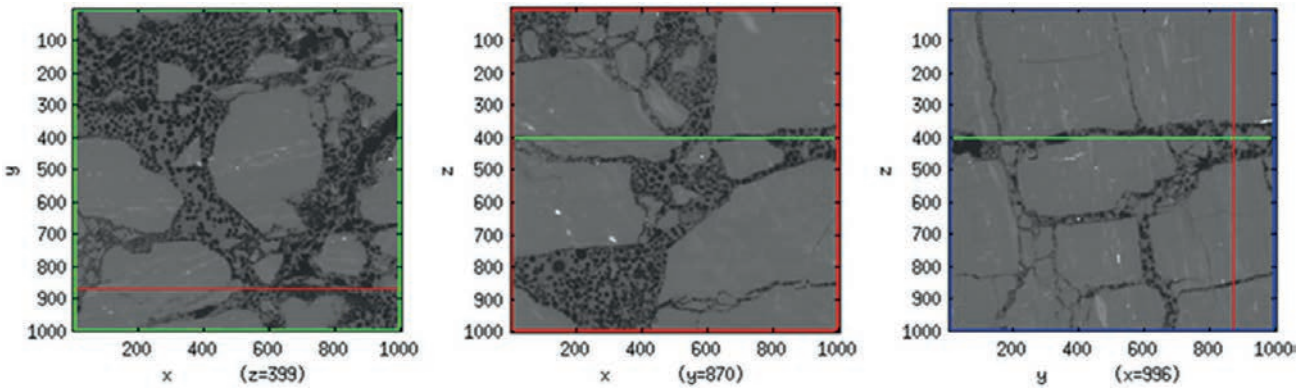


Figure 2.
Several cuts through the cubic sample of geocomposite material, consisting of a coal matrix injected by a polyurethane resin.

architecture as such outperforms Enna's Westmere one, in our application by 20-40%, which can be estimated from the test up to 8 processing elements (one socket) when the processors work in similar conditions.

Table 4 reports computations with the largest benchmark GEOC–2l (about 200 million DOF) and demonstrates the impact of the coarse grid size on the time of the solution. We can observe that very aggressive aggregation leads to the best results.

# Sd	Enna				Anselm			
	DD		DD + CG		DD		DD + CG	
	# It	T_{iter}	# It	T_{iter}	A/E	T_{iter}	A/E	T_{iter}
2	914	8,461.2	437	3,523.1	0.67	5,644.2	0.79	2,785.4
4	1,129	4,973.3	428	1,923.6	0.59	3,526.2	0.72	1,383.4
8	1,421	2,942.5	416	922.9	0.82	2,422.6	0.79	725.7
16	1,655	1,994.6	376	415.8	0.64	1,325.8	0.84	348.7
32	1,847	1,923.5	329	**348.3**	0.42	798.3	0.56	194.8
64	2,149	3,074.9	295	505.9	0.20	620.8	0.23	117.6
128					n/a	515.7	n/a	**107.1**

Table 3.
Timings of the GEOC-2s benchmark achieved by the GEM solvers (without and with a coarse grid problem involved) on the multiprocessor Enna and cluster Anselm

# Sd	Enna						Anselm	
	DD + 9x9x9		DD + 9x9x18		DD + 9x9x27		DD + 9x9x27	
	# It	T_{iter}	# It	T_{iter}	# It	T_{iter}	# It	T_{iter}
4	751	13,719.0	858	15,737.6	997	18,518.4	997	12,671.4
8	690	6,237.7	800	6,960.8	917	8,062.9	917	5,803.9
16	585	2,717.4	674	4,010.6	777	4,815.6	777	2,576.6
32	585	2,483.6	622	2,923.8	708	3,452.5	708	1,157.5
64					627	3,637.0	627	558.8
128							652	358.5
256							631	**299.6**
512							649	333.5

Table 4.
Timings of the GEOC-2l benchmark obtained by the GEM-DD-CG solver on the multiprocessor Enna and cluster Anselm. The coarse grid problem of different sizes is involved in computations.

We could confirm this observation on Anselm, where the best time in Table 4 (299.6 s with 256 processing elements and aggregation 9 x 9 x 27) was surpassed by an experiment with the coarser aggregation 15 x 15 x 31. The overall best GEOC-2l solution time of 249.8 s was achieved after 910 iterations on # Sd = 512 subdomains (32 compute nodes employed).

Conclusion

In this project, we aimed at the development of efficient parallelisable solvers required in many geotechnical applications including the micro-FEM analysis of geocomposites. We showed that the stabilised PCG method is a good choice for systems arising from the numerical solution of Neumann problems, or more generally problems with a known small dimensional null space.

We also compared parallel iterative solvers and development approaches – specific GEM solvers developed from scratch and solvers set up from components of the general Trilinos library to achieve the same functionality. We compared the performance of the Anselm supercomputer with an ordinary workstation.
The computational demands can be increased by nonlinearity or the solution of inverse problems. Thus the future steps will be oriented to a possible efficient use of thousands of cores, requiring a new strategy with the multilevel Schwarz method.

References

[1] Blaheta R., Jakl, O., Kohut, R., Starý, J. GEM – A Platform for Advanced Mathematical Geosimulations. In: R. Wyrzykowski et al. (eds.): PPAM 2009, Part I, LNCS 6067, Springer-Verlag, 2010, pp. 266-275. Doi: 10.1007/978-3-642-14390-8_28

[2] Blaheta R. et al. Material parameter identification with parallel processing and geo-applications. In: R. Wyrzykowski et al. (eds.): PPAM 2011, Part I, LNCS 7203, Springer-Verlag, 2012, pp. 366-375. Doi: 10.1007/978-3-642-31464-3_37

[3] Blaheta R., Jakl, O., Starý, J., Turan, E. Parallel solvers for numerical upscaling. In: P. Manninen, P. Oster (eds.): PARA 2012, LNCS 7782, Springer-Verlag, 2013, pp. 375-386. Doi: 10.1007/978-3-642-36803-5_27

[4] Blaheta R., Kohut, R., Kolcun, A., Souček, K., Staš, L. Micromechanics of geocomposites: CT images and FEM simulations. In: EUROCK 2013, Balkema, 2013, pp. 399-404. ISBN 978-1-138-00080-3

[5] Blaheta R., Jakl, O., Starý, J. Iterative solution of singular systems with applications. PPAM 2013, LNCS, Springer, 2014. Doi: 10.1007/978-3-642-55224-3_12

[6] Blaheta R., Kohut, R., Starý, J., Sysala, S. Computational and reliability aspects of micro-geomechanics. IACMAG 2014, accepted.

ADJOINT TOMOGRAPHY OF SURFACE WAVES: CHOICE OF REGULARISATION BASED ON SYNTHETIC TESTS

Research institution:
Charles University in Prague

Principal investigator:
Ľubica Valentová

Researchers:
František Gallovič, Bohuslav Růžek, Josep de la Puente, Peter Moczo

Project ID:
OPEN-4-11

Introduction

With increasing computational power, seismic tomography based on the so-called adjoint calculation of sensitivity kernels is becoming a common tool for improving our knowledge of the Earth's structure (e.g. Tromp et al. 2005, Fichtner et al. 2006, etc.). The greatest advantage of the adjoint method is the usage of numerical solution of the elastodynamic equation (EDE), meaning the least amount of simplifications (approximations) in the forward problem. To obtain the sensitivity kernel (Figure 1) one needs only two calculations solving the EDE, which makes the method computationally feasible. Inversion is then performed by an iterative procedure of improving model parameters based on the kernel calculation and misfit between data and synthetics.

In this study, the adjoint tomography is combined with traveltime measurements of surface waves originating from the ambient-noise cross-correlations. It has been shown that by the cross-correlation of wave fields between two receiver points, the Green's function may be extracted, (e.g. Shapiro & Campillo 2004, Bensen et al. 2007).

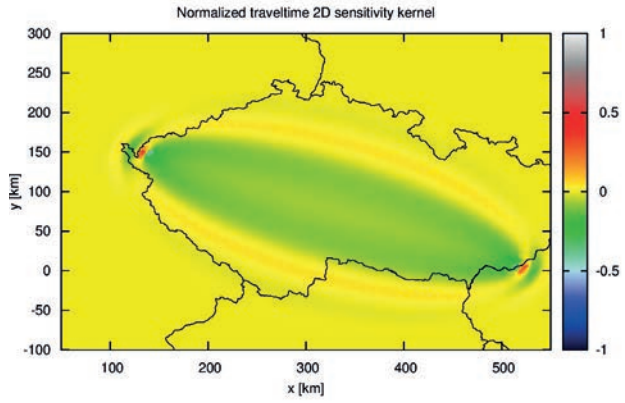


Figure 1.
2D traveltime sensitivity kernel for 20 s surface wave between 2 stations in the Czech Republic

We present synthetic tests of 2D adjoint tomography of traveltimes of 20 s surface waves across the Czech Republic. The data coverage may be considered perfect for tomography due to the density of the station distribution. Nevertheless, artefacts in the inferred velocity models arising from the data noise may be still observed when weak regularisation (Gaussian smoothing of the misfit gradient) or too many iterations are considered. To examine the effect of the regularisation and iteration number on the performance of the tomography in more detail we performed extensive synthetic tests. Instead of the typically used (although criticised) checkerboard test, we suggest performing synthetic tests with two different target models – simple smooth and complex realistic models. The test with the simple model demonstrates the possibility of obtaining false small-scale structures. Contrarily, the test with the complex model reveals the possible resolving power of the dataset. For various noise and Gaussian smoothing levels, we analysed the convergence towards (or divergence from) the target model with an increasing number of iterations. Based on the tests we identified the optimal regularisation, which we then employed in the inversion of Love-wave traveltimes.

In our calculations, we employed the adjoint version of the software package SeisSol2D. The forward calculation is carried out by the Discontinuous Galerkin method with the Arbitrary High Order Time Derivatives (ADER-DG) on unstructured meshes (Käser & Dumbser 2006, de la Puente et al. 2007).

Results and Methods

Using the synthetic tests we investigate two effects on the convergence towards a correct model: the effect of i) noise level added to synthetics and ii) spatial Gaussian smoothing of the gradient.

In the following tests, we use three values of the noise level specified in terms of the standard deviation. One of them corresponds to the noise-level denoted as σ_t, which was estimated from the real dataset. The other two levels are chosen for analysing the effect of smaller and larger data noise ($1/3\sigma_t$ and $3/2\sigma_t$).

We smooth the calculated misfit gradients by means of convolution with a 2D isotropic Gaussian function. We consider three widths (denoted as σ_x), 50 km, 100 km and 150 km, representing different strengths of the smoothing. The smallest width corresponds to the wavelength of the 20 s data. This might be considered the natural choice, because it prevents the smaller-than-wavelength structures without oversmoothing. The other two smoothing levels represent two different degrees of over-regularisation.

We use two models to generate synthetic data: Model I represents a smooth structure, Model II contains strong small-scale heterogeneities, see Figure 2. The station configuration is kept the same as for the real inverse problem. Using the first model we investigate occurrence of artefacts due to data noise. Using the second model we test the reliability of inferring small-scale structures. Combining all synthetic tests we achieve the best relation between the quality of the obtained model and the regularisation in terms of (i) the amount of the smoothing and (ii) number of iterations.

To quantify model improvement, we define the model misfit as the L2 norm of the difference between the obtained model m_n at iteration n and target (i.e. true) model m_{targ} normalised by the L2 norm of the initial model m_0:

$$\text{misfit} = \frac{m_n - m_{targ}}{m_0} \cdot 100\%$$

Usually the curve has a local minimum. We denote the model corresponding to the minimum as the optimal model.

The results of the synthetic tests are shown in Figure 3, for Test I and II in the left and right column, respectively. The individual rows show results for the three gradient smoothing levels σ_x with distinct colours and symbols for different noise levels. The decrease of data misfit (traveltime residual RMS) for the three noise levels is plotted in grey using the respective symbols. Generally, during the first iteration steps of the inversions the structural model is improved at the longest wavelengths and both the model and the data misfit decrease considerably. During further iterations, the shorter wavelength structures of the model are revealed. Therefore, a higher number of iterations are needed to achieve the optimum for Test II rather than for Test I, namely 1-3 for Test I and 6-8 for Test II. A common characteristic of both tests is the behaviour of the model misfit after achieving the optimum: aside for the cases with the lowest noise level, the divergence from the target model (expressed by the increase in the model misfit) emerges, although the data misfit continues to decrease (i.e. the inversion starts to explain the data noise by false structures). With an increasing level of the gradient smoothing the smaller-scale heterogeneities are suppressed and the resulting model contains less detailed structure. Therefore, the results stabilise in the optimum after

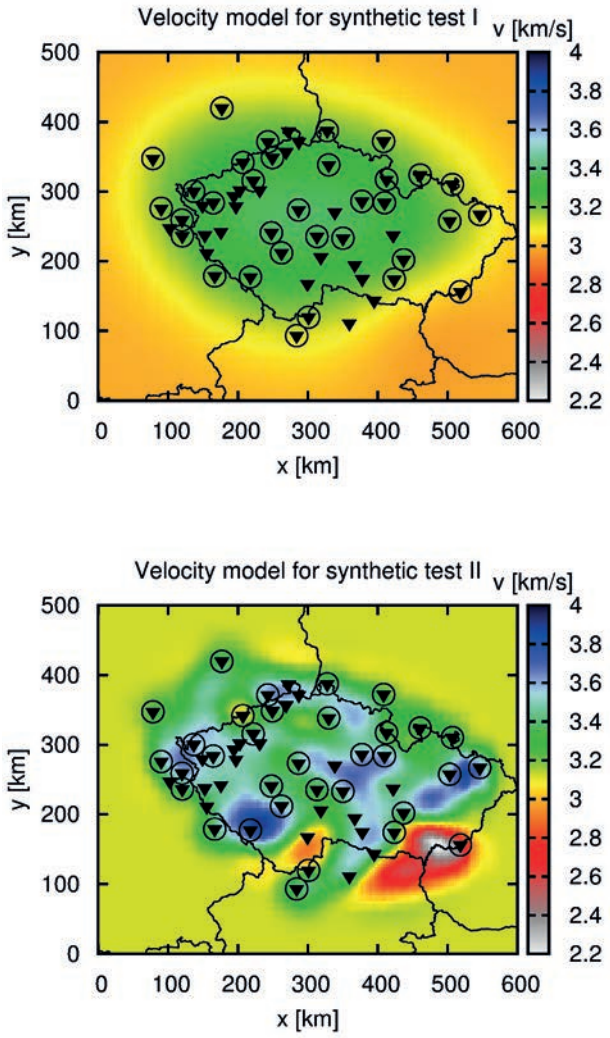


Figure 2.
Velocity (target) models I (up) and II (down) for synthetic tests. Stations and sources are shown by inverted triangles and circles, respectively.

a larger number of iteration steps and remain stable for several more iterations. According to Test I, the most reliable results are obtained with the smoothing width of at least 100 km and up to the 6th iteration. The best results for Test II are achieved using gradient smoothing of maximally 100 km in 6-8 iteration steps. A similar test where we changed the position of stations confirms that the conclusions are independent on the source-receiver coverage.

The inferred optimal regularisation parameter setting (for 20 s data: 100 km Gaussian smoothing and 6th iteration) scaled for different periods has been applied to inversion of Love waves in a 4-20 s period range. In this way we have obtained dispersion maps (i.e. 2D velocity maps for each period separately) of the Czech Republic with the highest resolution and reliability possible, for the given data and method (see Figure 4).

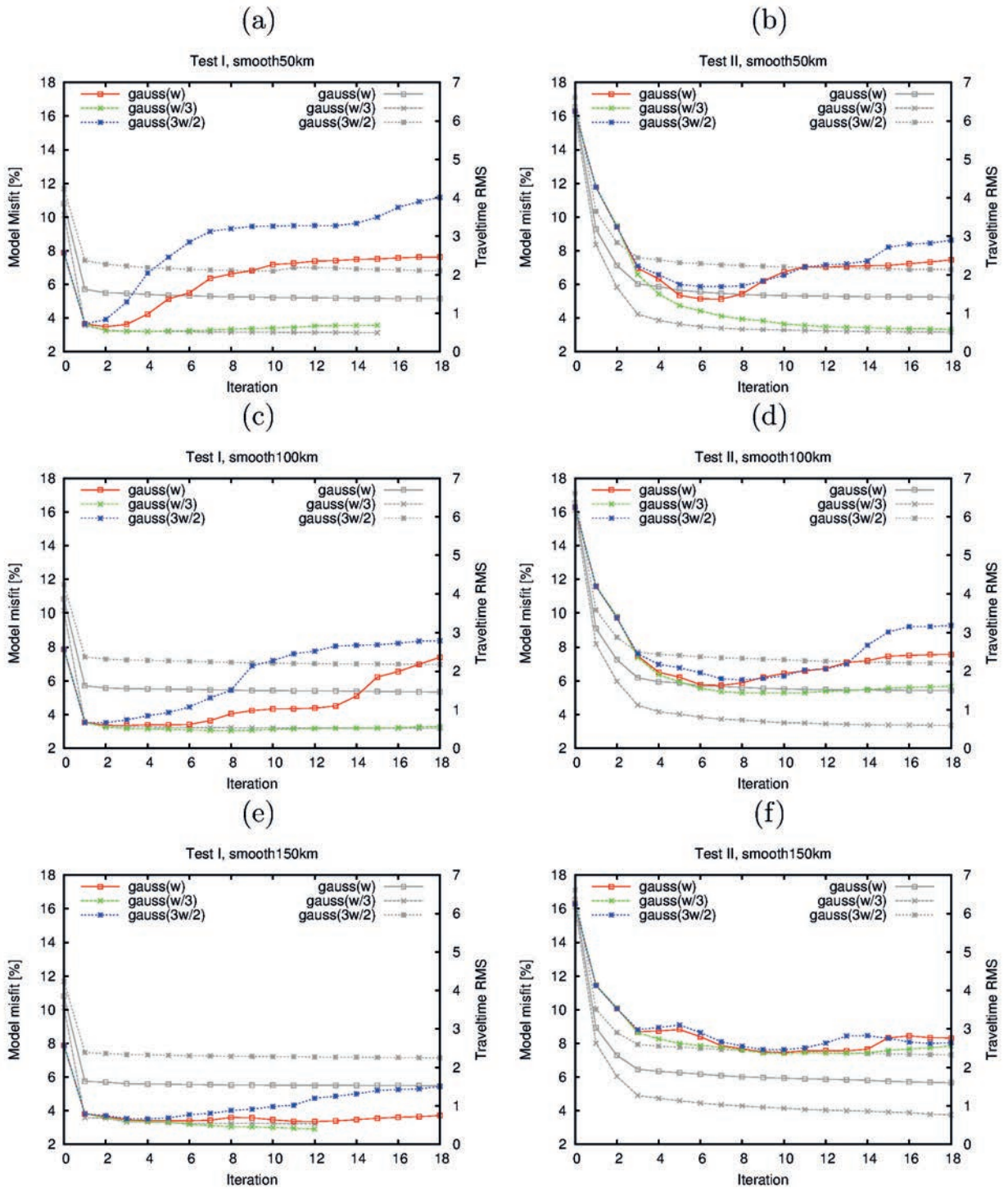
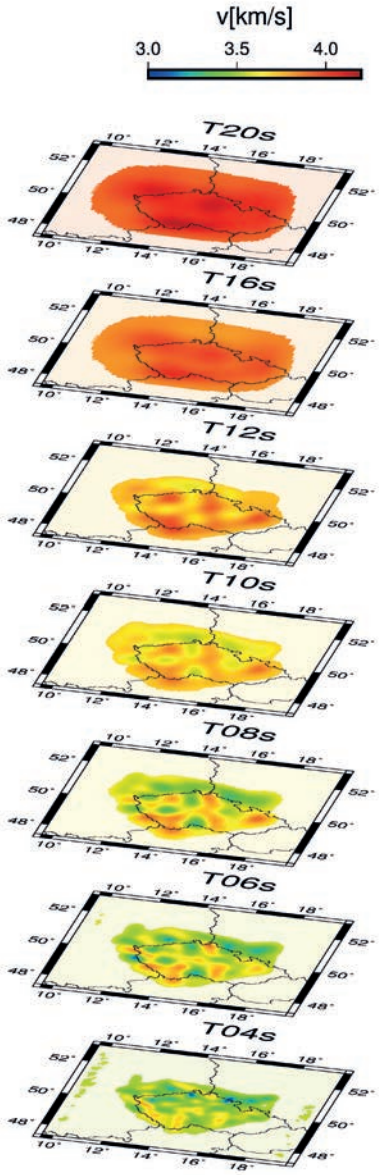


Figure 3.
Evolution of model misfit and data misfit with iteration for Test I (left) and Test II (right). Each row corresponds to different level of the gradient smoothing: 50 km (top), 100 km (middle), 150 km (bottom). Individual lines correspond to the different levels of noise applied to the synthetic data.



On-going Research / Outlook

Currently, we are working on the inversion of inferred dispersion maps into a 3D S velocity model of the crust. For this problem, we employ the MCMC method (parallel tempering) to obtain full Bayesian information on the model parameters.

Conclusion

We have used the synthetic tests to identify the appropriate smoothing strength and the optimal number of iteration steps so that the inferred model contains not only large-scale but also small-scale heterogeneities without false structures. The inferred regularisation setting was applied in real data inversion of Love wave traveltimes in the Czech Republic.

We believe that although the presented tests are feasible only in 2D (since the full 3D adjoint inversions are computationally extremely expensive), they provide an important insight into the method itself and reveal its main problems and limitations in general. From the presented numerical experiments, one may infer that the regularisation (e.g. in the form of gradient smoothing) is recommended to be greater than the wavelength considered to prevent from the bold structural artefacts. The proper number of iterations cannot be easily generalised because it strongly depends on the choice of the initial model. For applications similar to ours, the tests suggest that the number should be rather low (~5-10). Otherwise, the model may be spoiled by the data noise artefacts.

Figure 4.
Love wave dispersion maps for 20-4 s periods (top to bottom)

References

[1] Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, Geophys. J. Int., 169(3), 1239-1260. Doi: 10.1111/j.1365-246X.2007.03374.x

[2] de la Puente, J., Käser, M., Dumbser, M. & Igel, H., 2007. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – IV. Anisotropy, Geophys. J. Int., 169(3), 1210-1228. Doi: 10.1111/j.1365-246X.2007.03381.x

[3] Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, Phys. Earth planet. Inter., 157(1), 86-104. Doi: 10.1016/j.pepi.2006.03.016

[4] Käser, M. & Dumbser, M., 2006. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes I. The two-dimensional isotropic case with external source terms, Geophys. J. Int., 166(2), 855-877. Doi: 10.1111/j.1365-246X.2006.03051.x

[5] Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, Geophys. Res. Lett., 31(7), Doi:10.1029/2004GL019491.

[6] Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, Geophys. J. Int., 160(1), 195-216. Doi: 10.1111/j.1365-246X.2004.02453.x

Publication

[1] Valentova, L., Gallovic, F., Ruzek, B., de la Puente, J., Moczo, P. (2015). Choice of regularisation in adjoint tomography based on 2D synthetic tests, Geophys. J. Int. 202, 787-799. Doi: 10.1093/gji/ggv183

PARALLEL MODELING OF OCEAN FLOWS AND THEIR MAGNETIC SIGNATURES

Research institution:
Charles University in Prague

Principal investigator:
Jakub Velímský

Researchers:
Zdeněk Martinec, Libor Šachl, David Einšpigel

Project ID:
OPEN-5-8

Introduction

Sea water contains a high concentration of dissolved salts. Therefore, the oceans act as an ionised fluid where electric charges are carried by dissolved cations and anions, making sea water highly conductive. As oceans flow through the Earth's main magnetic field, the ion content in the oceans induces an electric current. As these electric currents flow around the globe, they, in turn, generate secondary magnetic fields, which evolve in terms of both magnitude and geometry through time. This effect, called motional induction, depends on the geometry and spatial scales of the ocean flow, as well as the electrical conductivity of sea water and the Earth's underlying crust and mantle. Observations of the ocean-induced magnetic field have the potential to be used as a constraint when examining ocean dynamics.

In general, ocean circulation models describe the response of the ocean system to atmospheric momentum, heat forcing, tidal forcing, fresh-water influx, etc. This response can simply be represented in terms of the eigenmodes of a linearized system of equations. The zeroth mode is equivalent to the vertically-averaged component of the motion, also known as the barotropic mode. The higher modes are called baroclinic modes and are associated with higher order components of the vertical density profile.

The renaissance of interest in the magnetic field generated by ocean flow arose from the launch of the European Space Agency Swarm mission. The mission is designed to measure the magnetic signals originating in the Earth's core, mantle, crust, oceans, ionosphere and magnetosphere. Swarm's goal is to investigate all of these components, but extracting the small part produced by the ocean flow is probably its greatest challenge. Compared to the other contributions, the ocean-generated magnetic field reaches amplitudes not larger than 2 nT at Swarm satellite altitudes, and cannot be recognized by a visual inspection of the Swarm magnetic data.

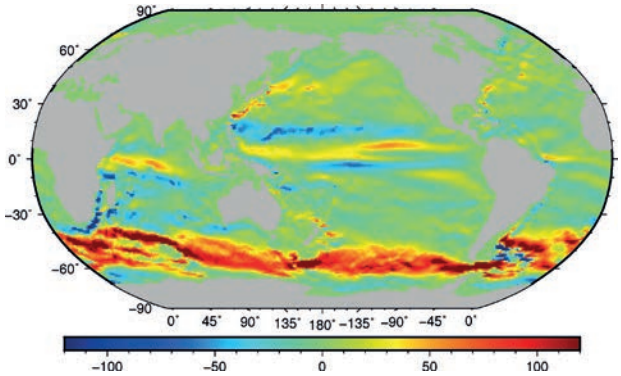


Figure 1.
Snapshot of the vertically integrated zonal velocity (in m2/s) from the LSOMG wind-driven model

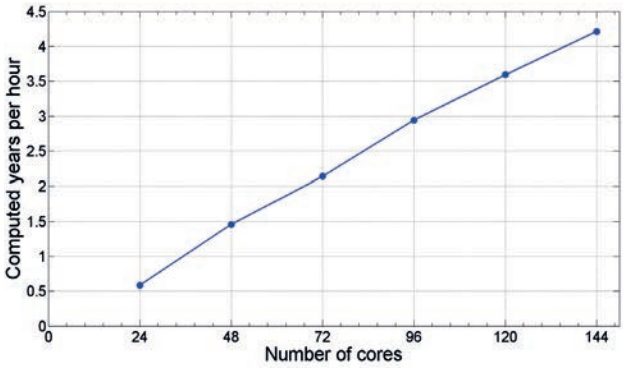


Figure 2.
Scaling of the LSOMG code

Results and Methods

Two different codes are used to model the ocean flows. The DEBOT model, working in the barotropic approximation, and the LSOMG model, predicting the baroclinic flows. Then, using suitable a-priori models of the electrical conductivity of the sea-water, the underlying Earth's crust and upper mantle, and the spatiotemporal characteristics of the Earth's main magnetic field, the magnetic signatures are calculated for each ocean model by solving the electromagnetic induction equation (ELMGIV).

DEBOT

The DEBOT model [1] is based on the shallow water equations which are approximated in space by finite differences on the standard Arakawa C-grid. The time-stepping scheme is a stable and second-order accurate generalized forward-backward scheme, using a combination of a third-order Adams-Bashforth step with a fourth-order Adams-Moulton step. The code has been tested by numerical experiments based on conservation of integral invariants, and a comparison with deep-ocean tide gauge data has been performed. The DEBOT code is written in the free-format Fortran language with implemented OpenMP parallelization.

LSOMG

The modeling of the baroclinic ocean flow is the most demanding task. It is modeled using the z-coordinate baroclinic ocean model which solves the primitive equations under the Boussinesq approximation. A snapshot of the vertically integrated zonal flux is shown in Figure 1.

The governing equations are discretized on the Arakawa C-grid using the finite difference method. Time stepping is based on the split-explicit method. The advection of temperature and salinity is performed using either the third-order Quick scheme or the flux limited version of the Lax-Wendroff algorithm. The state equation of sea water is formulated in potential temperature. The code is written in free-format Fortran language, making use of LAPACK subroutines (through the Intel Math Kernel Library) and MPI parallelization. The code scales well with the number of cores, as shown in Figure 2.

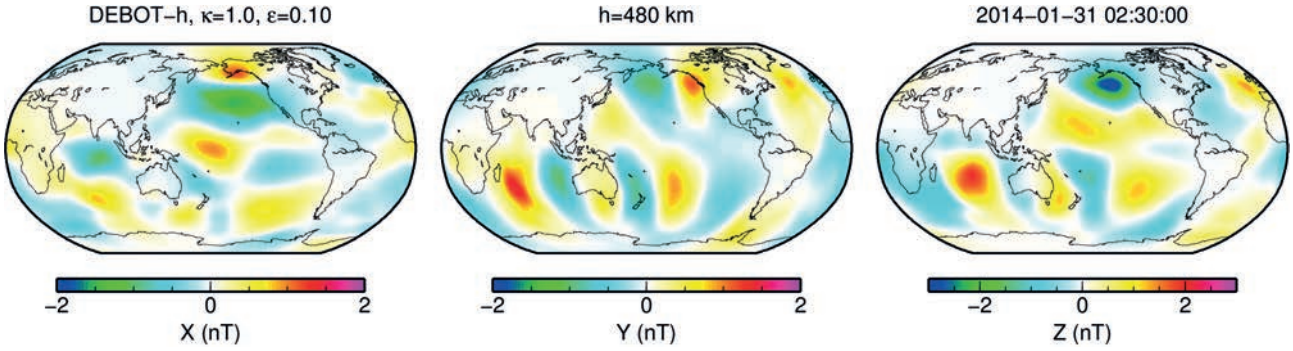


Figure 3.
Snapshot of the tidally induced magnetic field at satellite altitude

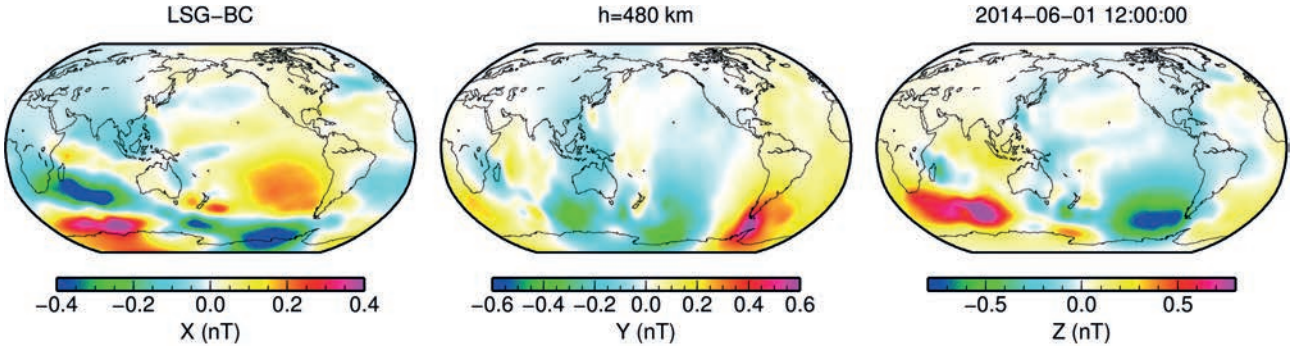


Figure 4.
Snapshot of the magnetic field induced by wind-driven flows

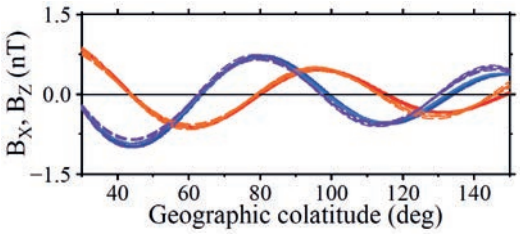


Figure 5.
Vertical (blue) and horizontal (red), medium-wavelength magnetic field along Swarm B track 7902. Solid lines show the ocean signature reconstructed from satellite data, dashed lines show the predictions for different settings of DEBOT and ELMGIV models.

ELMGIV

The induced magnetic field is modeled using the time-domain, spherical harmonic-finite element approach. Recently it has been modified to include the excitation by ocean flows represented by spherical harmonics, and to use an unconditionally stable Crank-Nicolson integration scheme. It is written in free-format Fortran, relying on BLAS and LAPACK libraries for linear algebra tasks, using FFT libraries for fast transformation between spatial and spherical-harmonic domains. Depending on the size of the problem, the stiffness matrix is solved either directly by a modified block-tridiagonal algorithm, or iteratively, using a preconditioned, stabilized bi-conjugate gradient technique.

RESULTS

We calculated an extensive set of runs, studying the magnetic signatures of both tidally and wind-driven ocean flows, and their dependency on various thermo-mechanical and electrical parameters. Figures 3 and 4 show respectively typical snapshots of the magnetic signatures of the barotropic, tidally induced flows, and baroclinic, wind-driven flows, at the satellite altitude. In the first case, we have detected a significant sensitivity to two physical phenomena: the gravitational self-attraction of the seawater column, and the drag caused by the energy dissipation of internal waves. The complicated physics of these phenomena is usually described by a simplified set of parameters in the barotropic ocean flow models. As the match between the predicted fields, and signals reconstructed from Swarm satellite data is not complete (see Figure 5), the magnetic field offers a potential to further improve the self-attraction and internal wave drag description.

In the case of wind-driven, baroclinic flows, we have shown the need for fully three-dimensional modeling of the magnetic fields, which correctly resolves the poloidal and toroidal parts, and their mutual interaction.

On-going Research / Outlook

In the future we plan to adjust the parameters of the DEBOT model, in particular the self-attraction and internal wave drag coefficients, to improve the match between the predicted and observed signals. This must be preceded by an automatization of Swarm data processing, in order to reconstruct the ocean signals from a large set of tracks. We will also proceed with high-resolution runs of the baroclinic LSOMG model and their induced fields to further study the toroidal magnetic field in the oceans.

Conclusion

The computational resources of the IT4I National Supercomputing Center allowed us to explore the parameter space of the numerical models of the ocean flows, and their magnetic signatures observed by the Swarm satellite mission.

Publication

[1] Einšpigel, D. & Martinec, Z., 2016: Time-domain modelling of global ocean tides generated by the full lunisolar potential. Accepted to Ocean Dynamics. Doi: 10.1007/s10236-016-1016-1.

Project website: geo.mff.cuni.cz/SwarmOceans

04 | ENGINEERING AND COMPUTATIONAL FLUID DYNAMICS

AXIALLY AND RADIALLY COOLED BRAKE DISC WITH COVER

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Marek Gebauer

Researchers:
Tomáš Blejchař, Petr Horyl

Project IDs:
OPEN-7-3, DD-16-2, IT4I-1-1, IT4I-5-3

Introduction

The proposed G-Cooling System (GCS) concept of active brake cooling in racing cars is designated for the Formula One racing cars. Since the new concept is focused on this limited group of racing cars, it was necessary to take into account the set of conditions for having to only cool the brake discs with a gaseous medium, i.e. air. The proposed cooling system can also be used for other competitive racing cars (Nascar, Le Mans, S2000, etc.). In these cases, cooling by liquid medium would theoretically be possible as well, depending on compliance with the terms of the races. Another group/type/area of races where you can apply GCS is high performance racing sport cars (Porsche, Lamborghini, Mustang, ...). The cooling system is based on the assumption of operation at high speeds. Whether it is possible to deal with the brake cooling system using GCS in passenger or goods vehicles is an issue; however, it certainly cannot be ruled out.

The background solves the cooling of the disc using one-way flow. The main stream of cold air is fed via a line into the centre of the disc and consequently non-heated air is blown around the perimeter of the free space. The main idea of G-Cooling System is in its dual cooling.

The principle of cooling is shown in Figure 1. A blue arrow indicates the cooling air being input into the centre of the disc. It is subsequently supplied into the centre of the disc and the primary radial channels forcedly directed to its circumference. Using the directional blades of the diffuser, air is discharged into the space below it, which converts part of the kinetic energy into pressure energy. Compressed air is discharged from the space of the diffuser through the secondary transverse channels of the disc to the outside of the braking system. As you can see, the air streams are isolated from each other in the disc.

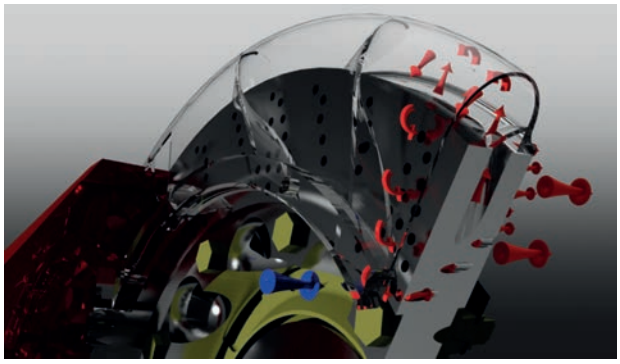


Figure 1.
G-Cooling System

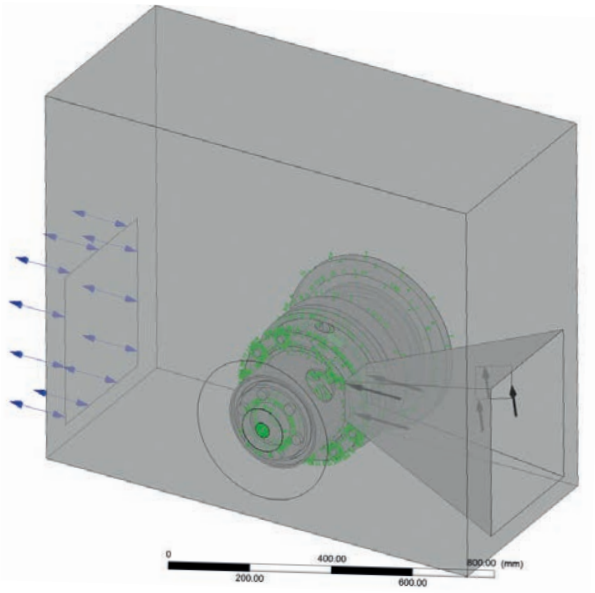


Figure 2 .
3D model of the classical braking system support inside the dynamometer

Results and Methods

Thanks to the IT4Innovations project, it is possible to perform mathematical simulations of flow solving the prediction of the turbulent behaviour of the newly proposed cooling system. Based on these findings, the prototype has been produced.

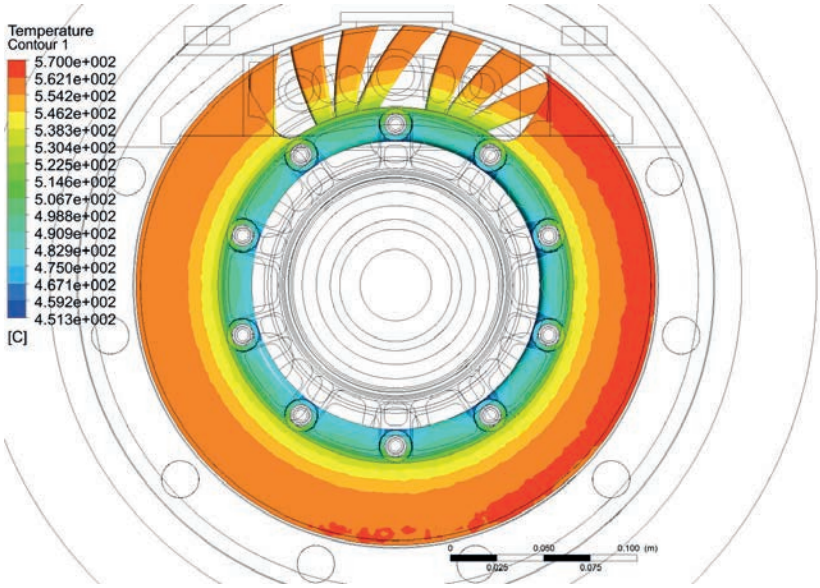


Figure 3.
Temperature distribution of the classical brake disc using ANSYS CFX software

At the beginning phase of the project a big question mark of the simulations was the heat transfer from an internally ventilated brake disc using a rotating thermal field boundary condition in the areas of the brake pads. The current situation only solves these conditions for a full brake disc. A partially simplified assembly imposing a classical brake system inside an airtight dynamometer was modelled, including its surroundings (Figure 2). The entire mathematical model including mesh was created in the programming environment of the ANSYS 16 module CFX. The mathematical mesh contained a total of approx. 80 million cells. Multi reference frame analysis has been set including transient effects. The turbulent model k-ω SST was chosen as it is suitable for blade devices. In the simulation it was necessary to configure the total of thirteen domains (e.g. calliper, cover, area of the air flow, etc.), and thirty-six interactions among them. The brake disc was divided into several domains. These domains in the form of rings were connected using interpolation functions. The rotating thermal field was applied in place of contact with the brake disc and pads (see Figure 3). Convergence of simulation was positively influenced by that procedure. The other boundary conditions were set according to the experimental procedure.

Comparison of the real experiment and computer simulation turned out inside the dynamometer in accordance with the assumption. The result of the measured values on the brake disc have not corresponded with the course subtracted results from the simulation (see Figure 4). The main reason for these differences is the inaccurate setting of initial values in the calculation. More accurate outputs from the simulations would need more positional measurements on the real assembly or allow the simulation to run in several braking cycles, which would be very time consuming. However, although the maximum difference between the reality and numerical calculation was about 50°C, the next step will similarly set a simulation of the brake discs (GCS) in a dynamometer. When selecting the same boundary conditions, the results of both simulations will only refer to the relative difference values and the sensitivity to a change of geometry, which ultimately determines the relative efficacy between the two brake systems.

According to the experiment results, it is obvious that the newly designed system has reached satisfactory results. Despite the limiting conditions of the air flow, inaccuracies in the manufacture of the bladed diffuser, and the stationary rim, the brake disc and calliper temperatures were reduced by about 10%, and even 20% respectively (see Figure 5). In the racing car, the braking components were cooler in the range of 10-25%. The GCS brake disc system, after finishing the equivalent of one race, was subject to metallographic analysis and comparison with conventional brake discs. The results were surprising. The GCS brake disc system displayed an approx. 200°C lower temperature.

On-going Research / Outlook

At present, the primary steps are made in order to sell the remaining licenses. In the near future, it is planned to submit the project to facilitate the production of the brake discs casting method and offer a more affordable version of the GCS for companies and racing teams.

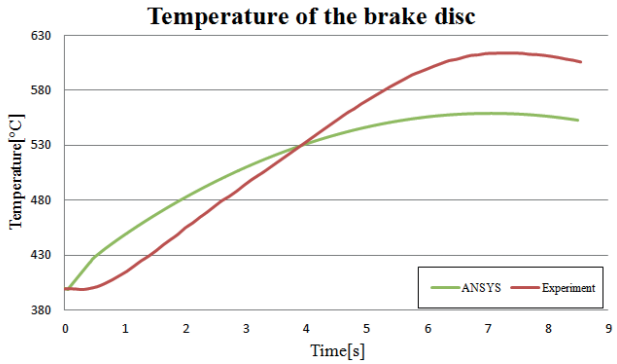


Figure 4.
Temperature distribution of the classic brake disc

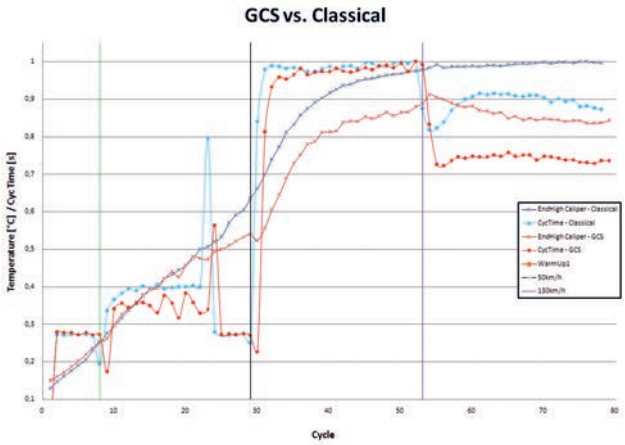


Figure 5.
Comparison of the temperature distribution of a classic and the GCS system under dynamometer conditions

Conclusion

The main benefit is the vastness of computing tasks and its comparison with real measured experiments. It provides an overview of individually set tasks and its evaluation, depending on many parameters (geometry, speed, etc.). Above all, the geometric analyses will help assess the current state of the results in rotary vane machines with an emphasis on heat dissipation.

This work may serve as a detailed guide to set calculations in the ANSYS CFX software environment. The experimental verification procedures and methods can serve as the initial testing background for future jobs in non-commercial software, e.g. OpenFOAM or developed algorithms using HPC resources.

Publications

[1] Vysoká škola báňská - Technická univerzita Ostrava. Axially and Radially Cooled Brake Disc with Cover. Marek GEBAUER, Zbynek LORENCUK. IPC: F16D 65/12. EU. Patent, EU2596264.

[2] Vysoká škola báňská - Technická univerzita Ostrava. Axially and Radially Cooled Brake Disc with Cover. Marek GEBAUER, Zbynek LORENCUK. IPC: F16D 65/12. United States. Patent, US 9,068,612 B2.

[3] Vysoká škola báňská - Technická univerzita Ostrava. Axiálně a radiálně chlazený brzdový kotouč s krytem. Marek GEBAUER, Zbynek LORENCUK. IPC: F16D 65/12. Czech Republic. Utility model, 22100.

[4] Vysoká škola báňská - Technická univerzita Ostrava. Axial und radial gekühlte Bremsscheibe mit Abdeckung. Marek GEBAUER, Zbynek LORENCUK. IPC: F16D 65/12. Germany. Utility model, 20 2011 110625.

[5] Vysoká škola báňská - Technická univerzita Ostrava. Brakes for vehicle (part of -). Marek GEBAUER, Zbynek LORENCUK. EU. Industrial design, 2013-35796.

SOLVING NON-LINEAR STRUCTURAL DYNAMICS PROBLEMS ON A SUPERCOMPUTER

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Tomáš Karásek

Researchers:
Pavel Maršálek, Petr Horyl

Project partner:
Laboratory of Mechanical Devices Testing at GIG Katowice

Project ID:
IT4I-2-1

Introduction

In many areas of engineering, physical testing was widely used to test the durability of designed structures or their parts in the past. This old-fashioned approach usually requires special testing facilities or laboratories. It also required a prototype to be built and tested. All of this makes this approach very costly and ineffective when it comes to bringing a new product to the market. Another essentially cheaper way is the creation of a computer model which exhibits a good correlation with respect to the existing data from equivalent laboratory conditions. Thanks to easier access to huge computational power, in the past decade a numerical instead of a physical model could be used for significantly more precise and very accurate results when compared to physical tests. For rapid prototyping, virtual testing is inevitable, and without numerical modelling and simulation rapid prototyping would be impossible.

In this project a numerical model of the steel arch supports was developed. This model was validated through physical testing and was used for determination of the influences of important factors including different materials, the number of stirrups in the yielding friction joints, and different values of tightening torque for the proper function of it. This parametrical study was created based on practical requirements from industry and the obtained results will be reflected in the design of new types of steel arch supports.

Figure 1.
FEM model of yielding connection

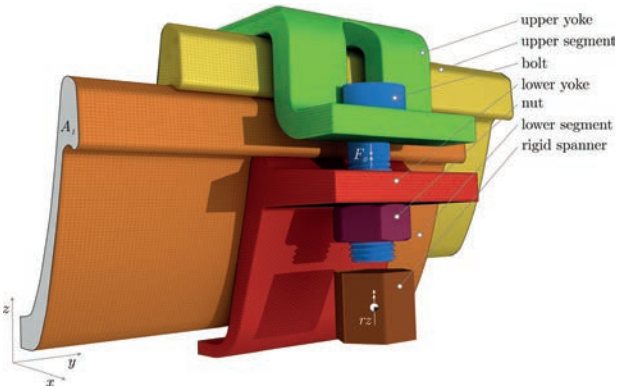


Figure 2.
Finite element model of the bolted connection

Results and Methods

The project dealt with the total load-bearing capacity of steel arch yielding supports of roadways, which is inevitable for the design of the roadway support. The results of a finite element (FE) model were compared with results of experimental tests of steel arch supports in a laboratory. Computer modelling of the load-bearing capacity of the steel arch support was performed using the massively scalable strongly nonlinear solver MARC. The finite element models of the arch supports and stirrups were created and assembled per drawings without shape simplification (Figure 1). To achieve accurate results, even details which would be otherwise omitted, such as bolted connection (Figure 2), were taken into consideration.

An experimental investigation was realised in the Laboratory of Mechanical Devices Testing at GIG Katowice. The computer model of the steel arch support captures the strength and deformation behaviour of the arch support very well.

For the mathematical description of the behaviour of the material properties of all the components, a multi-linear model (bilinear plus ideal plastic material from ultimate stress) was used. Calculation of the FE model was implemented on a supercomputer to 12 solver-threads. The solution took 29 hours.

The results show that the chosen modelling procedure reflects the considerable influence of the tightening of the friction joints on the load-bearing capacity of the yielding support (Figure 3 and Figure 4).

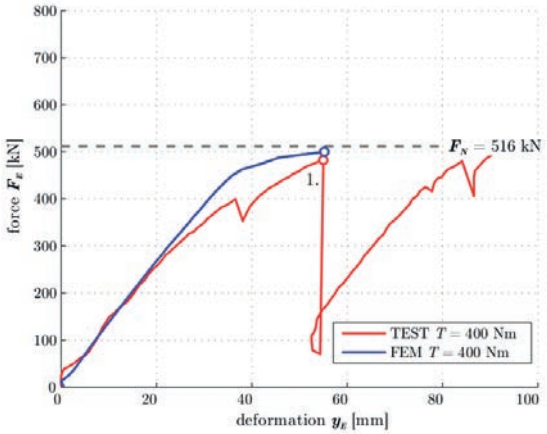


Figure 3.
Comparison of testing of yielding supports and FEM simulations T = 400 Nm

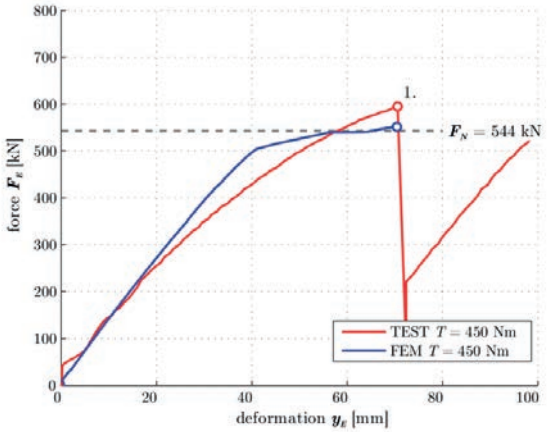


Figure 4.
Comparison of testing of yielding supports and FEM simulations T = 450 Nm

On-going Research / Outlook

Results show that bolted connections play an important role and should be modelled in full resolution. In future work the detailed investigation of forces, stresses and deformation of bolted connection will take place.

Conclusion

The computer model of the yielding support also plausibly characterises the behaviour of the support up to the time of the first slip in the clamped joints, consistent with experiments. The value of the loading capacity of the yielding arch supports at the moment of first slip represents with sufficient accuracy the total load-bearing capacity of arch supports at the specified loading conditions.

The mathematical model reflects the influence of the tightening bolts of yielding friction connections on the first regular slip and load-bearing capacity of the support. The procedure used for mathematical modelling with the results of representative laboratory tests facilitates reliable parametric studies of the behaviour of steel arch supports from the viewpoint of construction, shape, material, optimal yielding connections and a scheme of loading.

References

[1] Horyl, P.; Šňupárek, R.; Maršálek, P.; Paczesniowski, K. Load-bearing Capacity and Deformation of Steel Arch Supports of Roadways. At press in Arch. Min. Sci.

[2] Horyl, P.; Šňupárek, R.; Maršálek, P. Behaviour of Frictional Joints in Steel Arch Yielding Supports. Arch. Min. Sci. 2014, 59 (3), 723-734. Doi: 10.2478/amsc-2014-0054

[3] Horyl, P.; Šňupárek, R.; Hlaváčková, M. Loading capacity of yielding connections used in steel arch roadway supports. Proceedings of the Seventh International Symposium on Ground Support in Mining and Underground Construction. Perth: ACG – Australian Centre for Geomechanics (Potvin, E.; Brady, B.). 2013, pp. 461-470. ISBN 978-0-9806154-7-0.

[4] Brodny J. An analysis of frictional effect on the working of the yielding support frame. Gornictwo i geologia. 2010, Tom 5, Zeszyt 2 (in Polish)

[5] "Steel Qualities." Special Sections. ArcelorMittal Rodange & Schifflange, 24 Feb. 2015. Web. 24 Feb. 2015. <http://www.ares.lu/4_mining.htm>

Publications

[1] Horyl, P.; Maršálek, P.; Šnupárek, R.; Paczesniowski, K. Total Load-Bearing Capacity of Yielding Steel Arch Supports. Proceedings of the 2016 ISRM EUROCK 2016, Cappadocia, 29-31 August 2016, p. 897-900. ISBN 978-1-138-03265-1

[2] Maršálek, P. and P. Horyl. Modelling of Bolted Connection with Flexible Yokes Used in Mining Industry. ICNAAM2016, Rhodes, 16-20 September 2016

[3] Maršálek, P.; Poruba, Z.; Horyl, P.; Kováčik, M.; Kaláb, K. and P. Maršálek. Analysis of Force Effects on Bolted Connection with Parallel and Semi-Parallel Contact During Tightening Process. Advances in Mechanical Engineering (Sub. in Nov. 2016)

HPC FOR MODELLING STRONGLY NONLINEAR PROCESSES IN MECHANICS, PARTS I.-III.

Research institution:
IT4Innovations National Supercomputing Center

Principal investigators:
Pavel Maršálek, Petr Horyl, Tomáš Karásek

Researchers:
Marek Klemenc, Petr Ferfecki, Jan Grepl, Luboš Pečenka, Rostislav Čech, Marek Gebauer

Project partner:
Borcad cz, s.r.o.

Project IDs:
IT4I-8-6, IT4I-9-19, OPEN-7-20

Introduction

An accelerating of rail traffic brings negative consequences in the form of decreasing passive safety. Globally, the dynamic testing standards for interior equipment are not established in the countries of the European Union (EU), except for in the United Kingdom, where strict rules defined by GM/RT2100 [1] are applied. Currently, the politicians of the EU are making great effort to expand this standard to all countries of the EU.

GM/RT2100 prescribes for any type, or a new design, of the seats used in railway traffic (first class, standard class) dynamic structural integrity and potential injury tests. The testing seats are subjected to a specific acceleration impulse, see Figure 1. The maximum velocity of the seats during the test is approximately 48 km/h.

Both of tests are applied in the rearward and forward projection of the movement. In our research, we focus on the Injury prevention test in the forward direction because it is the most unfavourable test for the safety of passengers.

The testing process is carried out in special laboratories where two particular seats are positioned to an adjustable platform moving on the rail. One Anthropomorphic test device (ATD) – the crash test dummy [2] is placed in the second row of the seats according to GM/RT 2100. The platform is then accelerated and force, displacement, velocity, and acceleration time histories in the ATD are evaluated simultaneously. The goal is to meet all permissible values of more than twenty injury criteria [1]. In the case that the limits are exceeded, it is necessary to redesign the seat and perform the lab test again.

For this reason, it is appropriate to use the computer modelling method and a simulation of the entire test.

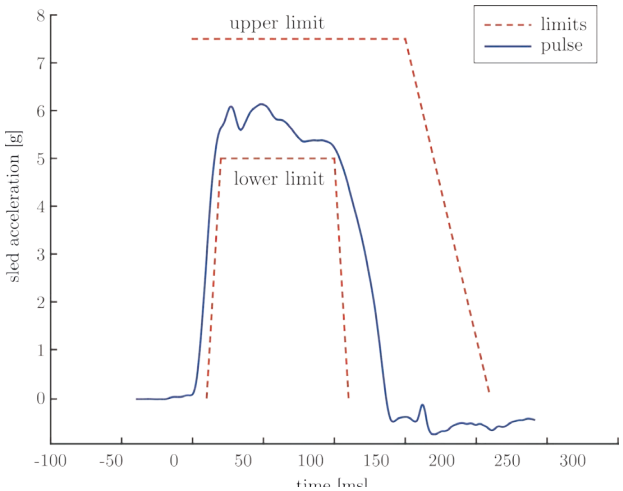


Figure 1.
Test impulse

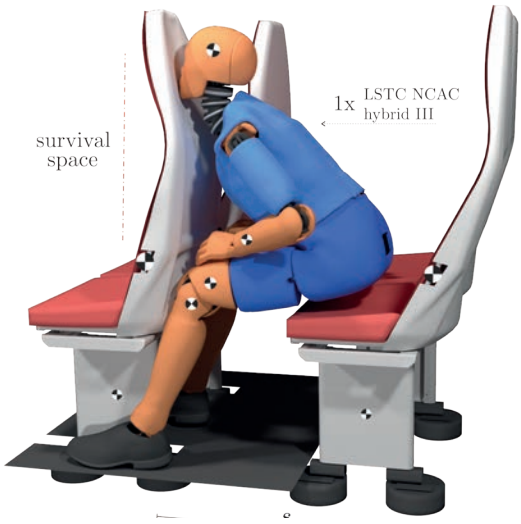


Figure 2.
Finite element model of Injury prevention test

Results and Methods

For mathematical description of this testing procedure, we use the Finite element (FE) method. The computational model of the injury prevention test is created and very detailed. For example, the seat frame structure is evaluated on the base of any spot-

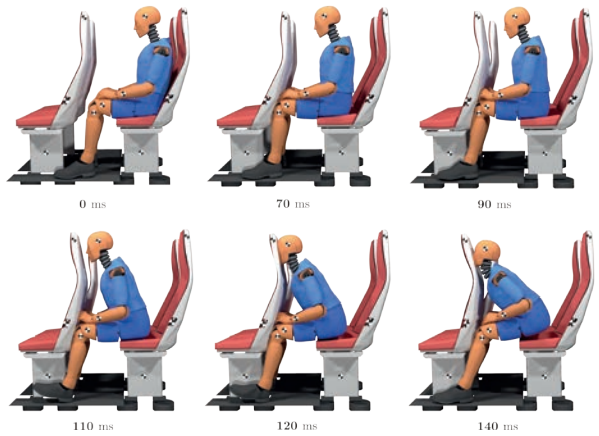


Figure 3.
Kinematic frames for particular time points

weld fracture, see Figure 2. A plastic material models, large deformations and non-linear description of friction between contact parts are considered.

The FE model consists of solid, shell and beam elements combination. The first load step is preloading of the bolted connections. A gravity load then starts to act, which settles the ATD to the seat. Finally, the acceleration pulse is applied to the mounting points (Figure 1). For implementation of these demanding simulations, we use the LS-DYNA solver. This analysis is very time consuming, and use of a supercomputer is highly recommended.

The simulated testing process is shown for particular time points in Figure 3. First, the ATD hits the seat plastic cover with the knees, then the hands and finally with the head.

Within the computational time projects and complimentary testing license, the scalability tests of this analysis were performed on the Anselm (Figure 4) and Salomon (Figure 5) clusters.

It was found that the LS-DYNA, LS-R7.1.1., mpp version solver is very sensitive to the type of connection between computational nodes. If the number of computational nodes is low, it is more suitable to use the Salomon cluster, because it includes 24 cores per single node. For higher numbers of computational nodes, communication among the nodes on the Anselm cluster is faster. Its nodes are interconnected by the fully non-blocking fat-tree Infiniband network.

For our application, it is most appropriate to use the Salomon cluster and utilise the entire computational node. Performing a calculation takes approximately 40 hours.

On-going research

We are currently working on development of a new type of seat platform, which is mounted directly to the train wall. This construction solution is very challenging because significant additional moments are acting in the wall during the testing procedure. At the same time the local buckling effect occurs, which is impermissible in term of GM/RT2100.

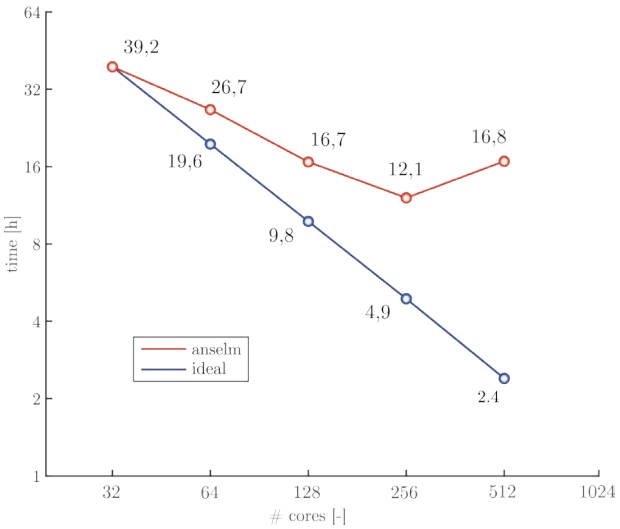


Figure 4.
Scalability test of LS-R7.1.1 mpp, dp – Anselm

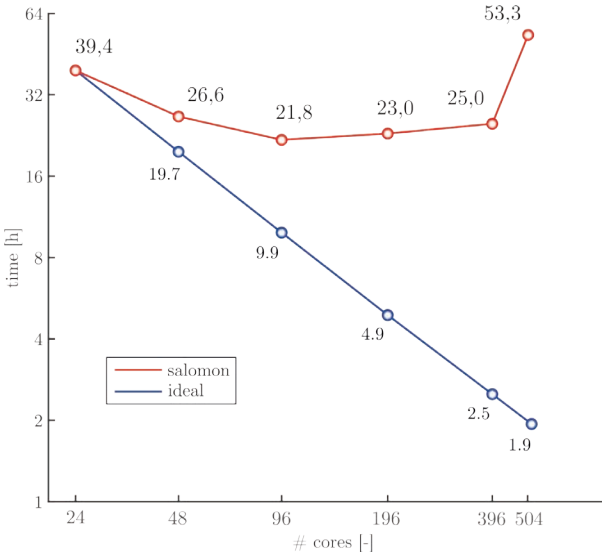


Figure 5.
Scalability test of LS-R7.1.1 mpp, dp – Salomon

Conclusion

The implementation of injury prevention tests to railway traffic plays a key role in passenger safety. The use of supercomputers for these application purposes offers manufacturers a new option in the terms of a safer seat design. Moreover, natural and financial resources are saved because there is no need to repeat more laboratory tests.

This work was created on the basis of practical requirements from industry and the obtained results will be reflected in the design of new types of regional and long distance passenger seats.

References

[1] GM/RT2100, Iss. 5: Requirements for Rail Vehicle Structures, http://www.rssb.co.uk/rgs/standards/GMRT2100%20Iss%205.pdf, 2012.
[2] Humanetics: Hybrid III 50th Male Dummy: http://www.humaneticsatd.com/crash-test-dummies/frontal-impact/hybrid-iii-50th; 2015
[3] LSTC: LS-DYNA Keyword User's Manual; http://ftp.lstc.com/anonymous/outgoing/jday/manuals/DRAFT_Vol_I.pdf; 2015.

Publications

[1] Marsalek, P.; Horyl, P.; Karasek, T. and P. Ferfecki. Influence of Seat Pitch on Passenger Injury Criteria in Regional Railway Traffic, ICNAAM2016, 16-20.9.2016, Rhodes.
[2] Marsalek, P.; Horyl, P.; Karasek, T. and P. Ferfecki. Folding Table and Safety in Railway Traffic. Proceedings of the 24th SVSFEM ANSYS User's Meeting and Conference 2016. 1-3.6.2016, Dolni Morava. ISBN 978-80-905525-3-1
[3] Marsalek, P.; Horyl, P.; Karasek, T. and P. Ferfecki. Increasing Passive Safety in Railway Traffic. PraceDays2016, Prague, 10-12.5.2016, poster section.
[4] Marsalek, P.; Horyl, P.; Karasek, T. and P. Ferfecki. Influence of Seat Pitch on Passenger Injury Criteria in Railway Traffic. CMSE 2016, Rožnov pod Radhoštěm, 25-27.5.2016, poster section.

CFD PROXY – STRONG SCALING IN CFD

Research institution:	Principal investigator:	Researcher:	Project partner:	Project ID:
T-Systems SfR	Christian Simmendinger	Jens Jägersküpper (DLR)	DLR	DD-16-20

Introduction

T-Systems and the DLR (EXA2CT SiB member) have agreed to develop a proto-app for a multigrid CFD solver, named CFD Proxy, within the scope of the EXA2CT project. The proxy will serve as a means to not only replicate the current programming model of TAU (SiB code), but to also implement, test and evaluate a large variety of hybrid programming models. The programming models will be generally applicable to most CFD applications that make use of unstructured meshes. T-Systems has developed a first version of this CFD Proxy. So far this version implements two hybrid programming models. The first version is an OpenMP/MPI model (used for reference), where a single MPI domain is further refined by means of a graph partitioner (Hendrickson 1995) into a fixed set of sub-domains. Every thread is then statically assigned to compute one of these sub--domains. Potential race conditions are handled via computing the cross-cutting edges (edges between thread domains) twice. For these cross-cutting edges results are only partially written to the attached end points. Every thread then only updates the points which the thread actually owns. While this task model introduces some computation overhead (i.e. computing cross--cutting edges between different thread domains twice), there are some important advantages to this task model.

- The fixed assignment of mesh points to threads allows for a straightforward optimal data locality (spatial and temporal data locality).
- The code features very good readability and maintainability due to the exclusive access of threads to mesh points.
- Due to the fixed assignment this hybrid programming model will deliver results which are bit-identical in subsequent runs.
- The implementation avoids (potentially substantial) locking overhead.

The CFD Proxy will readily allow to implement, test and evaluate other task models (like e.g. divide and conquer or asynchronous threading models).
The CFD Proxy uses meshes of actual (publicly available) airplane configurations. Here, we are only interested in a strong scaling setup. Performance critical communication

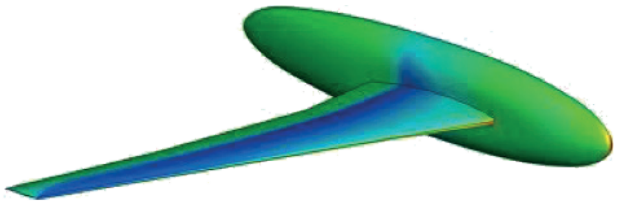


Figure 1.
The real world engine benchmark

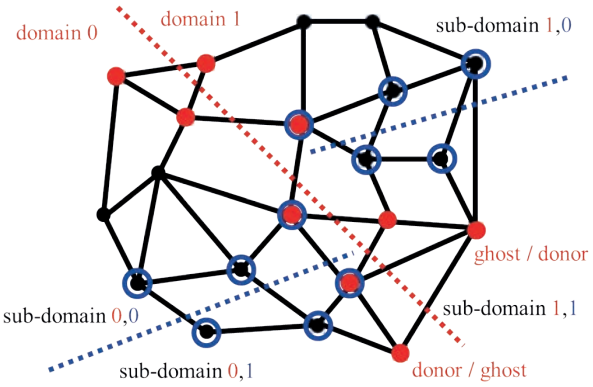


Figure 2.
Second-level decomposition (a vertex corresponds to a cell, an edge to a face)

in these codes is essentially next-neighbour communication. The respective applications hence typically scale linearly with increasing mesh sizes.

Methods

In this paper a single, but very prominent, use-case is considered: the ghost-cell exchange. Ghost-cell exchanges arise whenever some form of domain decomposition is used in combination with short-range interaction. Prominent examples are stencil codes for structured or unstructured meshes, particle-in-cell methods and many others. Although we here restrict

the discussion to the specific use case of ghost-cell exchanges for stencil codes for unstructured meshes, we believe that many of our findings will be readily applicable to other domains, like e.g. particle-in-cell methods.

In the graph in Figure 2, each vertex corresponds to a cell in the mesh so that each edge corresponds to a face between two cells. A vertex adjacent to three other vertexes corresponds to a tetrahedron, for instance. A mesh may be decomposed into two domains by a partition of the set of cells (i.e. vertexes in the corresponding graph) as indicated by the red dotted line in Figure 2. As (in this example) the stencil involves only direct neighbours, values at the red vertexes are needed at the respective other domain. So, each red vertex corresponds to a donor cell for which a copy exists as a so-called ghost cell at the respective other domain. As the number of cores grows, not only must the communication overhead (due to shrinking domain sizes and a correspondingly increased communication vs. computation ratio) be taken into account, but also the overhead due to ghost-cell access/computation. Some of the communication time can be hidden by overlapping communication with computation.

Results

We examine the underlying threading model and its suitability for strong scaling on Xeon Phi. All benchmarks have been performed on the IT4Innovations Salomon Cluster.

Figure 3 shows the scalabilty of our threading model for three meshes, running a single domain, i.e. NO ghost-cell exchange. The base for the depicted speed-up is one core running four threads. The speed-up of almost 49 for the fine mesh corresponds to an efficiency of 80%, which is a remarkably good value and mostly due to keeping the amount of shared data between threads at its bare minimum. For the second--level mesh, and more pronounced for the third-level mesh, load imbalance and (due to the shrinking mesh size per core) increased data transfer between neighbouring subdomains clearly affects scalability.

This is somewhat counterbalanced by the fact that at this scale the entire third-level mesh will fit in the L2 cache of the Xeon Phi. For a synthetic 3V multigrid cycle we define the runtime as T(level 1) + 2 *T(level 2) + T(level 3), resulting in a speed-up of 44.5, i.e. 74% core-level parallel efficiency for this 3V cycle. In Figure 4 192 Xeon Phi cards are used with a total of 46,080 threads. The GASPI implementation GPI-2 has been used as well as Intel's MPI implementation, version 5.0.1. In Figure 4 only the synthetic 3V cycle is depicted, where the comm_free scalability represents the theoretical limit – this is not linear due to load imbalances and additional computations involving ghost cells. For the third-level multigrid mesh we already are below our initial strong scaling target of 50 cells per thread. Certainly the most impressive result here is that GASPI async almost meets

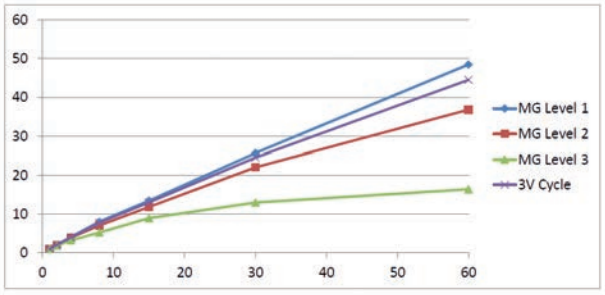


Figure 3.
Threading Model – Xeon Phi

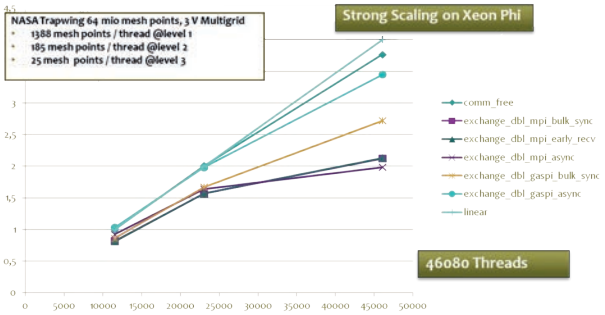


Figure 4.
Scalability of CFD Proxy on 46,080 threads on Intel Xeon Phi: Speed-up of 3V cycle over number of threads, 3V cycle, where the 3 levels have 1,388/185/25 cells per thread respectively.

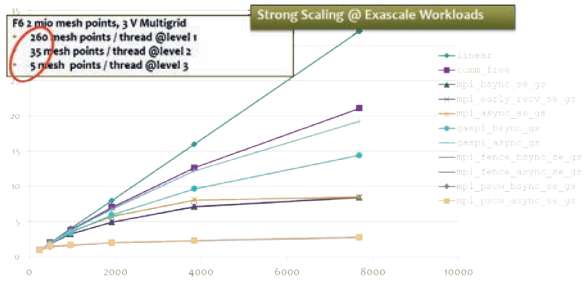


Figure 5.
Scalability of CFD Proxy on 7680 threads on Intel Xeon Phi: Speed-up of 3V cycle over number of threads, 3V cycle, where the 3 levels have 260/35/5 cells per thread.

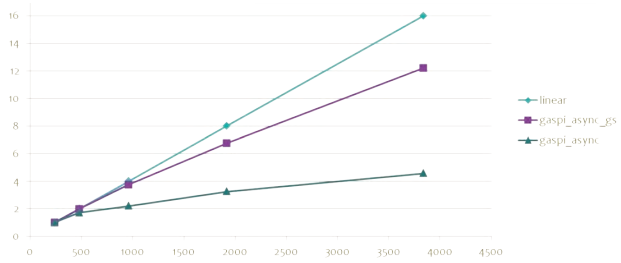


Figure 6.
Single-threaded pack/unpack vs. multi-threaded gather-scatter for GASPI

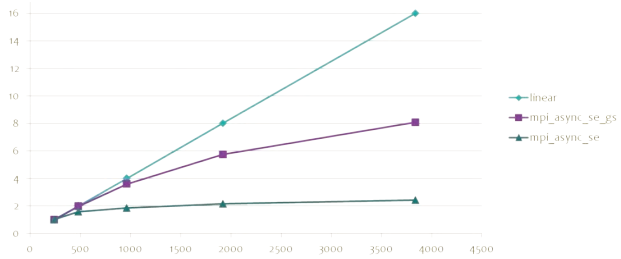


Figure 7.
Single-threaded pack/unpack vs. multi-threaded gather-scatter for MPI

comm free, implying that the ghost-cell exchange is entirely hidden, i.e. overlapped with computation. The asynchronous MPI implementation (which emulates the GASPI programming model, cf. above) performs almost as well. All other implementations scale poorly.

In Figure 5 we have used 32 Xeon Phi cards with a total of 7,680 threads. The mesh we have used however is much smaller. Per thread domain, here we are approaching the strong scaling regime of ~50 mesh point per thread we expect for exascale systems.

In order to highlight the remarkable relative performance impact of using a single thread for packing/unpacking the commbuffer vs multi-thread gathering we present Figures 6 and 7.

We observe a factor of up to three between serial packing (_async_se) and the full multithreaded gather-scatter implementation (_async_gs). This holds true for both the GASPI and the MPI implementation.

On-going Research / Outlook

The DLR has applied the findings in this work to their next-generation CFD Solver FLUCS. In doing so the DLR has achieved excellent scalability down to 400/200 mesh points per core/thread – an improvent of 10-100 over the existing SOTA.

Conclusion

In this work we have shown strong scalability down to less than a hundred cells per thread for a CFD Proxy application, for Intel Xeon Phi and Intel Xeon. Especially encouraging is the fact that the implementations appear as limited by the threading model rather than by the communication between neighbouring processes. We have found that, even though there are quite a few possible implementations for the ghost-cell exchange, only a single communication pattern appears to scale (exceedingly) well, both for multi-core and many-core. While that communication pattern originates from the GASPI PGAS API, it can be emulated using MPI. Even though we have used wimpy Xeon Phi cores for the first benchmark, the second benchmark demonstrates that the results carry over for brawny Intel Haswell nodes. The presented GASPI communication pattern is able to provide a significant improvement over the existing state of the art. We have published the CFD proxy source code under a liberal BSD license on GitHub and encourage the readers to improve this benchmark wherever possible.

References

GASPI Forum. URL <http://www.gaspi.de/>
PGAS-community-benchmarks/CFD-Proxy. URL <https://github.com/PGAS-community-benchmarks/CFD-Proxy>

SCALABLE SOLVERS FOR SUBSURFACE FLOW SIMULATIONS

Research institutions:
Institute of Mathematics of the Czech Academy of Sciences, Technical University of Liberec, Czech Technical University in Prague

Principal investigator:
Jakub Šístek

Researchers:
Jan Březina, Marta Čertíková

Project ID:
OPEN-1-3

Introduction

The main aim of the project has been performing detailed simulations of subsurface water flow in a candidate site for building a nuclear waste deposit and in a site surrounding an experimental underground tunnel. These simulations are performed using the finite element method (FEM). Due to very fine target resolution of the three-dimensional models, varying material coefficients, and geometrical inclusions of complicated geometry, the arising systems of linear equations are very large (up to tens of millions of unknowns) and ill-conditioned, and thus challenging for iterative solvers.

The project further built on a combination of the finite element package Flow123d (http://flow123d.github.io) developed for simulating various phenomena of the underground processes developed at the Technical University of Liberec. This software has been combined with the BDDCML solver (http://users.math.cas.cz/~sistek/software/bddcml.html) for solving the arising systems of linear algebraic equations.

This project has allowed us to optimise the solver and to solve large-scale problems of subsurface flows containing up to 15 million unknowns in a scalable way.

Methods and Results

The underground water flow through a fully saturated rock is modelled by the Darcy's law. The numerical solution is based on the mixed-hybrid finite element method, coupling the three-dimensional mesh with two-dimensional elements modelling fractures. While Darcy's law is also considered inside the fractures, the permeability coefficients are by orders of magnitude higher than those for the 3-D rock.

A state-of-the-art domain decomposition method – Balancing Domain Decomposition based on Constraints (BDDC) by Dohrmann [1] – has been used to solve the arising algebraic systems. The adaptive-multilevel version of the method is implemented in the open-source BDDCML library [2].

Within the project, we have combined Flow123d and BDDCML codes and significantly tuned the resulting algorithm and its implementation. The tuning was mostly done on benchmark problems. The resulting software was then applied to solving real-life problems of subsurface flow in an experimental site of the Bedřichov tunnel (Figures 1-3) and in a candidate site for building a nuclear waste deposit (Figures 4-6).

The results of the developed implementation have been published in a journal paper [3].

The developed method has been shown to provide both good robustness and scalability. An example of a strong scaling test on the candidate site for a nuclear waste deposit is presented in Figure 7. One can see that optimal scalability is achieved for a large range of the number of subdomains, i.e. computing cores.

As a subsequent goal of the project, new strategies for averaging discontinuous solutions at subdomain interfaces have been developed and tested [4]. The paper has shown that there is no single "silver bullet" in terms of averaging the discontinuous solution, and a good choice has to be made by the user based on properties of the problem.

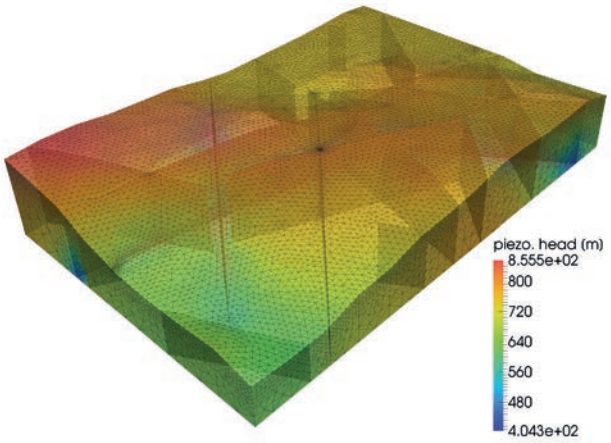


Figure 1.
Simulation of subsurface water flow around the Bedřichov tunnel. The computational model contains main geological fractures modelled as 2-D finite elements. Computational model by courtesy of Dalibor Frydrych (TU Liberec).

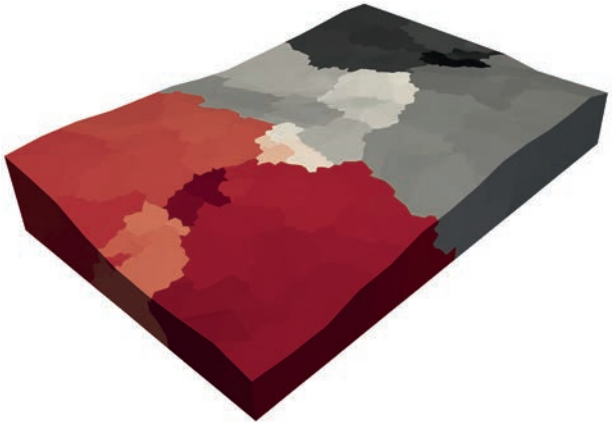


Figure 2.
Mesh around the tunnel divided into subdomains

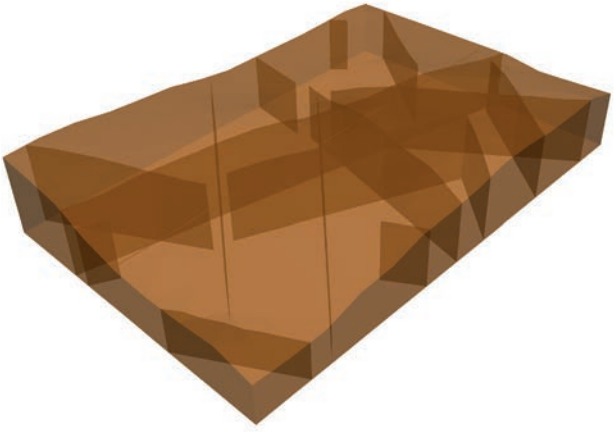


Figure 3.
Structure of fractures in the computational domain around the tunnel

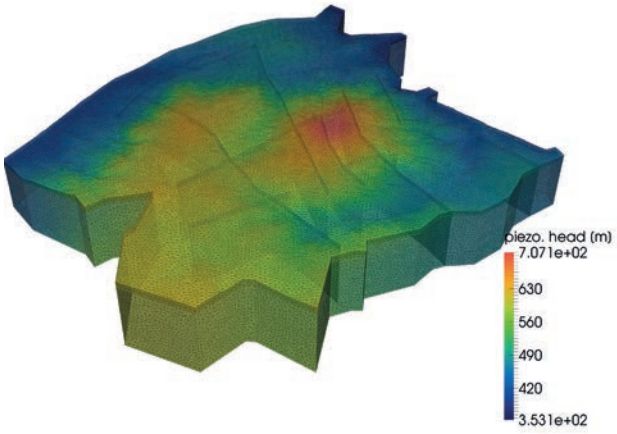


Figure 4.
Simulation of subsurface water flow in a candidate site for nuclear waste deposit. The computational model contains main geological fractures modelled as 2-D finite elements. Computational model by courtesy of Jiřina Královcová (TU Liberec).

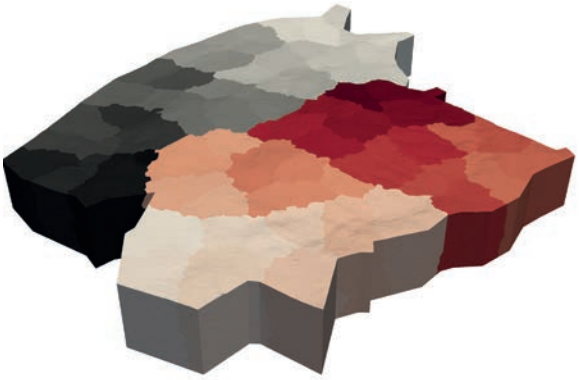


Figure 5.
Mesh divided into subdomains

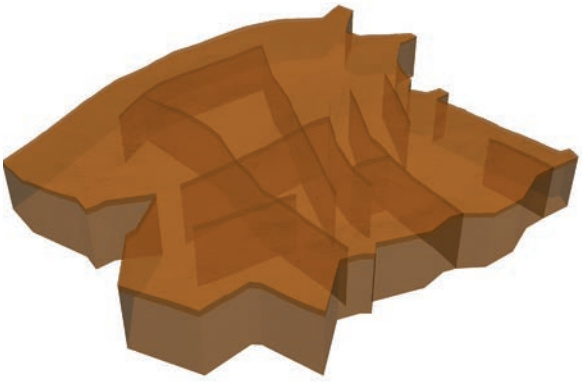


Figure 6.
Structure of fractures in the computational domain

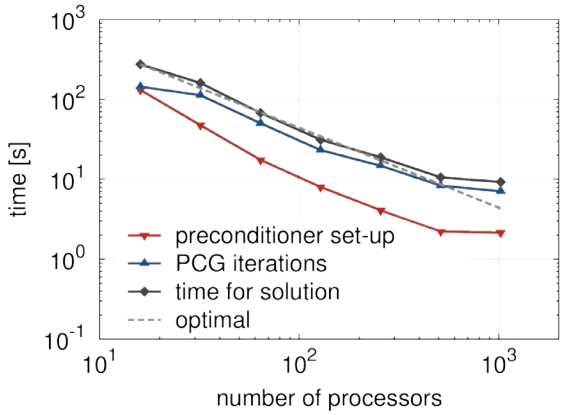


Figure 7.
Strong scaling test on the candidate site for the nuclear waste deposit

Conclusion

We have extended a state-of-the art domain decomposition technique, namely the BDDC method, to allow its use in connection to the mixed-hybrid finite element method and the combination of elements of different spatial dimensions. This allows modelling fractures in the porous media. Combining the finite element package Flow123d with the parallel implementation of BDDC in the BDDCML library has allowed us to solve the arising algebraic systems in a scalable way on up to several thousands of cores.

References

[1] Dohrmann, C. A Preconditioner for Substructuring Based on Constrained Energy Minimisation. SIAM J. Scientific Computing. 2003, 25(1), 246-258. Doi: 10.1137/S1064827502412887.

[2] Sousedík, B., Šístek, J. and J. Mandel. Adaptive-Multilevel BDDC and its parallel implementation. Computing. 2013, 95(12), 1087-1119. Doi: 10.1007/s00607-013-0293-5.

[3] Šístek, J., Březina, J. and B. Sousedík. BDDC for mixed-hybrid formulation of flow in porous media with combined mesh dimensions. Numer. Linear Algebra Appl. 2015, 22(6), 903-929.

Publication

[4] Čertíková, M., Šístek, J. and P. Burda. Different approaches to interface weights in the BDDC method in 3D. In J. Chleboun, P. Přikryl, K. Segeth, J. Šístek, and T. Vejchodský, editors: Proceedings of Programs and Algorithms of Numerical Mathematics 17. Dolní Maxov, Czech Republic, June 8-13, 2014. Institute of Mathematics AS CR, 2015, 47-57.

MULTILEVEL DOMAIN DECOMPOSITION SOLVERS FOR INCOMPRESSIBLE FLOWS

Research institutions:
Institute of Mathematics of the Czech Academy of Sciences, Czech Technical University in Prague

Principal investigator:
Jakub Šístek

Researchers:
Martin Hanek, Pavel Kůs

Project ID:
OPEN-4-10, OPEN-6-15

Introduction

The project has aimed at high-resolution computational fluid dynamics simulations of oil flow in hydrostatic bearings. Due to their complex geometries, three-dimensional simulations using fine meshes are required. The computations have been based on a novel domain decomposition algorithm, the multilevel BDDC method, with the aid of an existing parallel implementation of the method in the BDDCML library (http://users.math.cas.cz/~sistek/software/bddcml.html).

The two projects have allowed us to test the developed extensions of the domain decomposition methods in the massively parallel environment of the Salomon supercomputer. Weak scalability of the three-level BDDC method has been tested on up to 5,000 CPU cores of this machine, showing good parallel properties.

A subsequent goal has been further development of the multilevel method and optimisation of the BDDCML library for large numbers of computer cores. We have focused on incorporating parallel adaptive mesh refinement (AMR) into the workflow of non-overlapping domain decomposition. The p4est parallel mesh handler has been used by our prototype parallel implementation for mesh management and rebalancing. The second main ingredient of the developed implementation has been the BDDCML library for solving the arising systems of linear algebraic equations. The main challenges here included treatment of hanging nodes and incorporating disconnected subdomains provided by p4est.

Methods and Results

Hydrostatic bearings are used on heavy production machines. Oil is injected under high pressure into these bearings to provide low friction to the moving parts. However, understanding the lubrication mechanisms inside the bearing is an important task for simulations.

The flow of oil is described by the Navier-Stokes equations for incompressible fluids. The goal is to understand the flows inside a sliding bearing, which typically cannot by simplified by

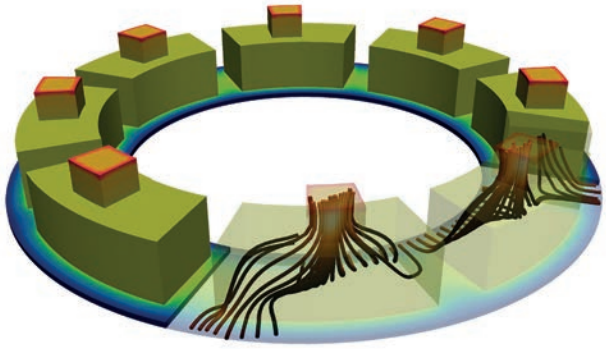


Figure 1.
Simulation of oil flow inside a 3-D computational model of a carousel hydrostatic bearing

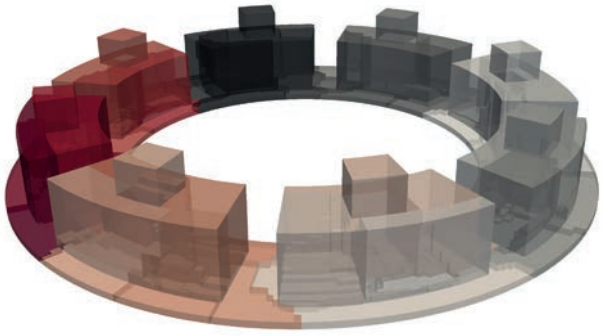


Figure 2.
Subdomains within the carousel hydrostatic bearing

considering axial symmetry. The equations are discretised by means of the mixed finite element method using Taylor-Hood finite elements. The main challenge for numerical modelling comes from the very low height of the throttling gap, which is only about 50 micrometers high. This immediately leads to finite elements with very high aspect ratio resulting in poor convergence of the BDDC method. An example of geometry of a carousel hydrostatic bearing is shown in Figure 1, with a division into subdomains shown in Figure 2.

A state-of-the-art domain decomposition method – Balancing Domain Decomposition based on Constraints (BDDC) by Dohrmann [1] – has been applied to algebraic systems with non-symmetric matrices resulting from discretising the Navier-Stokes equations in [2]. The effect of an irregular interface on the convergence of BDDC and an approach to significantly improve the convergence are described in a recent paper [3].

The solution of the coarse problem becomes a bottleneck for scalability for more than about 1,000 cores. An approach to overcome this issue is using the multilevel extension of BDDC introduced in [4]. The adaptive-multilevel version of the method is implemented in the open-source BDDCML library [5].

One of the goals of the projects has been testing the scalability of the developed implementation. These weak scaling tests have been performed on the Salomon supercomputer on the lid-driven cavity benchmark problem. The results have demonstrated that using three-level BDDC is beneficial compared to the standard (two-level) method for more than 1,000 cores, in spite of requiring more iterations. In Figure 3, we present the times required for set-up of the BDDC preconditioner, the iterations and the overall solution time, all three for the two and three levels in the BDDC method.

Another achievement of the projects has been development of a prototype implementation of a massively parallel finite element solver combining adaptive mesh refinement with the BDDC method. The implementation has been tested on the Salomon supercomputer, which considerably accelerated the development of this method.

The benchmark problem used for testing the adaptive features was a unit cube with a prescribed analytical solution of the Poisson problem. An adapted mesh after a few steps of refinement is shown in Figure 4 together with the prescribed solution. At the end of the project, the code successfully solved problems on highly adapted unstructured meshes resulting in systems with over 109 linear equations using 2,048 cores. Finally, convergence of the adaptive scheme was compared with the approach based on uniform refinements. As can be seen in the convergence plot in Figure 5, about 100 times less degrees of freedom are needed for the adaptive method to reach the same accuracy. The results have been summarised in our paper [6] recently submitted for publication.

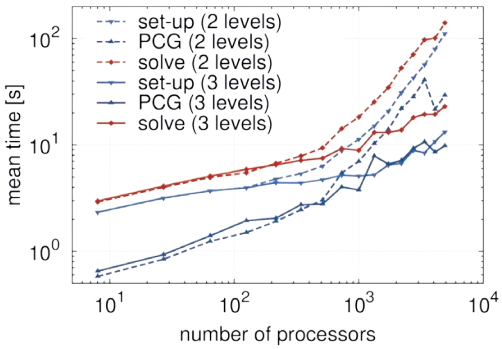


Figure 3.
Weak scaling test on the 3-D lid driven cavity problem. Adding the third level into BDDC results in better scaling behaviour beyond 1,000 cores.

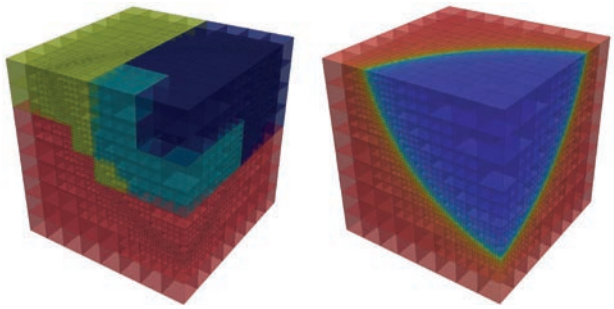


Figure 4.
A benchmark for parallel adaptive mesh refinement with the BDDC method. Mesh divided into four subdomains by p4est (left) and the prescribed solution (right).

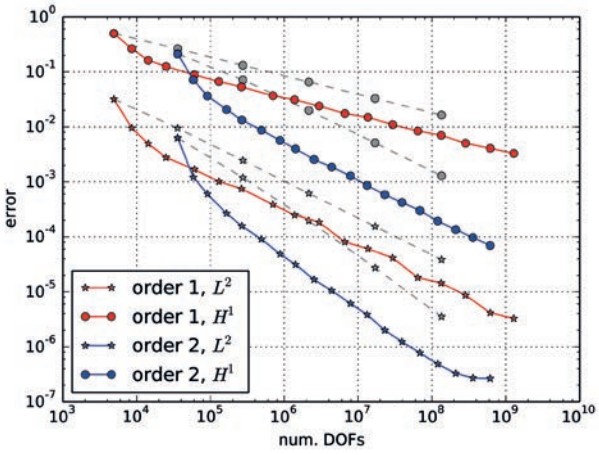


Figure 5.
Convergence orders using the adaptive and the uniform refinements with element of the first and the second order. The largest problem exceeds 109 degrees of freedom.

Conclusion

The BDDC method has been extended to non-symmetric problems arising in solving the Navier-Stokes equations. Within the project, the first computations using the multilevel BDDC method for these problems have been performed on a benchmark problem of the lid driven cavity showing significant benefits over the two-level method when using more than 1,000 cores. The developed parallel implementation was then applied to large-scale simulations of oil flows in hydrostatic bearings.

The BDDCML solver library has been subsequently optimised for its use in connection with the finite element method using adaptive mesh refinement. At the end of the projects, successful computations with as many as 109 degrees of freedom were performed on 2,048 cores. The BDDC, and in particular its three-level version, has been shown to provide a great technique for solving the arising systems of equations for Poisson as well as linear elasticity problems and using elements of up to the fourth order.

References

[1] Dohrmann, C. A Preconditioner for Substructuring Based on Constrained Energy Minimisation. SIAM J. Scientific Computing. 2003, 25, 246-258. Doi: 10.1137/S1064827502412887.

[2] Hanek, M., Šístek, J. and Burda, P. An application of the BDDC method to the Navier-Stokes equations in 3-D cavity. In J. Chleboun, P. Přikryl, K. Segeth, J. Šístek, and T. Vejchodský, editors, Proceedings of Programs and Algorithms of Numerical Mathematics 17. Dolní Maxov, Czech Republic, June 8-13, 2014, pages 77-85. Institute of Mathematics AS CR, 2015.

[3] Hanek, M., Šístek, J. and Burda, P. The effect of irregular Interfaces on the BDDC method for the Navier-Stokes equations. Proceedings of the 23rd international domain decomposition conference DDXXIII, Jeju Island, South Korea, July 6-10, 2015, to appear.

[4] Mandel, J., Sousedík, B. and Dohrmann, C. R. Multispace and multilevel BDDC. Computing. 2008, 83(2-3), 55–85. Doi: 10.1007/s00607-008-0014-7

[5] Sousedík, B., Šístek, J. and Mandel, J. Adaptive-Multilevel BDDC and its parallel implementation. Computing. 2013, 95(12), 1087-1119. Doi: 10.1007/s00607-013-0293-5

Publication

[6] Kůs, P. and Šístek, J. Coupling parallel adaptive mesh refinement with a nonoverlapping domain decomposition solver. Submitted 2016.

ACLA – ACTIVE LOAD ALLEVIATION

Research institution:
Výzkumný a zkušebný letecký ústav, Prague

Principal investigator:
Petr Vrchota

Researcher:
Aleš Prachař

Project ID:
OPEN-5-32

Introduction

The understanding of the aircraft's aerodynamics at the edge of the fight envelope is still not clear and has to be improved. These regimes are characterised by the massive flow separation at low-speed conditions on one side and buffeting at high-speed conditions on the other side of the flight envelope. The CFD simulations of these regimes are still challenging because of a massive flow separation on the wing and wake interaction with other parts of the aircraft. The ESWIRP [1] sub-project "Time-resolved wake measurement of separated wing flow and wall interference measurement" has been launched to help in better understanding these phenomena. The wind tunnel measurement of the NASA CRM has been done at ETW. Both low-speed and high-speed conditions have been tested and, among other values, the wake from the wing has been visualised by the time-resolved PIV method [2],[3]. The results will be used for improvement of the knowledge of the massive flow separation and the effect of the wake on the empennage, and also for further verification of the CFD methods. Due to very complex flow field, loads and deformation of the model, the evaluation of the applied CFD methods is necessary as the first step before analysing the effect of the wake.

During analyses of the wing tunnel test it has been found that a model deformation has an indispensable effect on the aerodynamic characteristics in cases where the ratio of dynamic pressure over the Young module is high. These cases correspond to the wind tunnel measurements with realistic Reynolds and Mach numbers corresponding to the flight conditions of an airliner [4]. Among other effects, the elasticity of a wind tunnel model has to be taken into account during CFD simulations to improve the accuracy [5]. The methods of how to implement the model deformation into CFD simulation ranges from the simple to the complex. Among the simplest methods belongs the deformation of the CFD model according to the measured deformation of the wind tunnel model during the test campaign [6]. The surface geometry is morphed manually to match the measured local wing's deformation for a particular angle of attack (AoA). On the other hand there is a coupling between the CFD and FEM solvers (Fluid Structure Interaction – FSI) [7]. Between these two methods is the use of the modal analysis, which can be done independently of the CFD simulations and it is not as demanding as FSI.

Results and Methods

The modal analysis has been used to obtain the model deformation. The structural model of whole airplane has been downloaded from the CRM website. It corresponds to the structural wind tunnel model. Due to hardware and memory restrictions, only the wing has been considered for this study. Modal analysis has been performed in MSC Nastran. Each mode is defined by modal frequency and mode shape, i.e., displacement of all nodes of the wing.

Since the model perturbation is obtained on the wing structural grid it is necessary to transform the mode shapes to the boundary of the CFD grid. The program, which is a part of the Edge package, based on the MLS (mean least squares) method has been used. The perturbation of the CFD surface node is obtained from the perturbations of the neighbouring structural nodes (the number of such nodes is limited by a given radius and/or maximum number of nodes) with linear fit.

The radial basis functions (RBF) approach [8] was used to extend the surface deformation onto the whole CFD volume grid. Several options are possible, however, for this case the deformed grids for each of the considered modes have been prepared prior to the run of the CFD solver by the mesh deform utility. The final deformed mesh considered in the CFD solver is derived from the baseline mesh plus linear combination of the mode shapes extended to the whole CFD domain. The final wing deformations for particular AoAs are depicted in Figure 1 and the shapes of the first six modes are depicted in Figure 2.

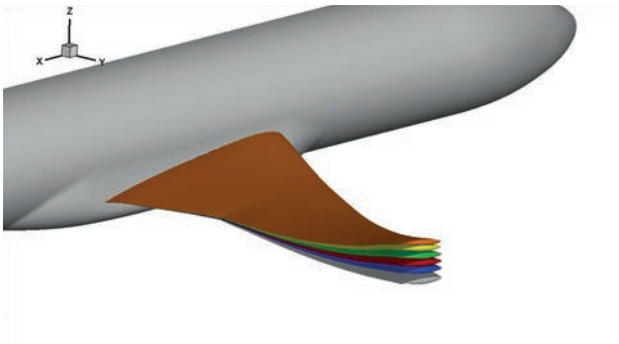


Figure 1.
Final wing deformations for particular AoAs

The CFD solver has been run in two modes, RANS and URANS, respectively. The solution from the RANS mode served as an initial solution for the URANS simulation. The modal coupling option has been switched on during the URANS simulations and the elasticity of the model has been taken into account. The final wing deformation has been obtained by solving equation 1 for the vector of generalised (modal) coordinates q, where the matrix A is the diagonal modal generalised mass matrix, D is a diagonal generalised damping matrix, E is the diagonal generalised stiffness matrix and Q is the vector of generalised forces (aerodynamic surface loads in this study).

$$Aq + Dq\dot{} + E\dot{q} = Q \qquad (1)$$

The lift and drag curves are depicted in Figure 3. It can be seen that the elastic model provides more accurate results in comparison with the rigid model for lift and drag coefficient, respectively. There are still some differences in C_L, mainly for smaller AoA's. These differences can be caused by the missing support system during CFD simulations. The support system affects (reduces) C_L, C_D and also lift curve slope (C_L vs. AoA) [5]. The wing's deformation of the CFD model, mainly twist angle, is slightly different in comparison with the wind tunnel model. It can explain the non-linear differences in C_L with AoA.

In order to ensure the wing's deformation during CFD simulations is the same as during the experiments, the wind tunnel model

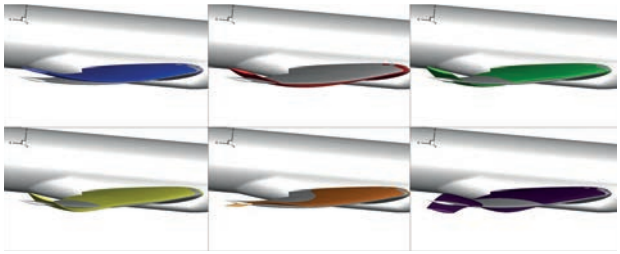


Figure 2.
Shapes of the first six modes

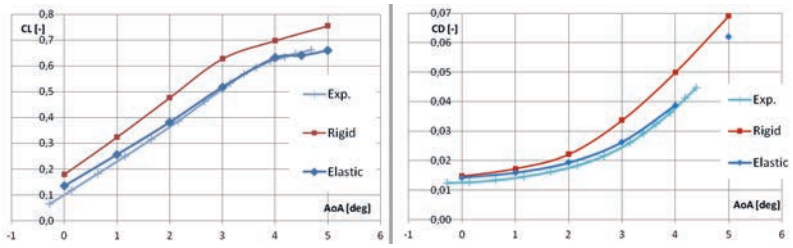


Figure 3.
Comparison of lift and drag coefficients with experimental results

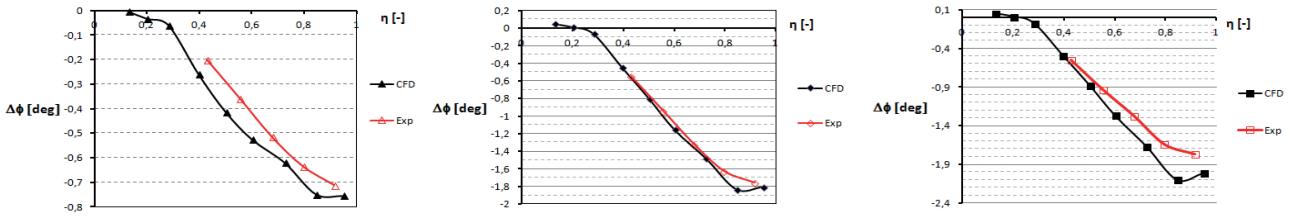


Figure 4.
Comparison of wing twist angle from the wind tunnel test and CFD for AoA=0° (left), AoA=3° (middle) and AoA=4° (right)

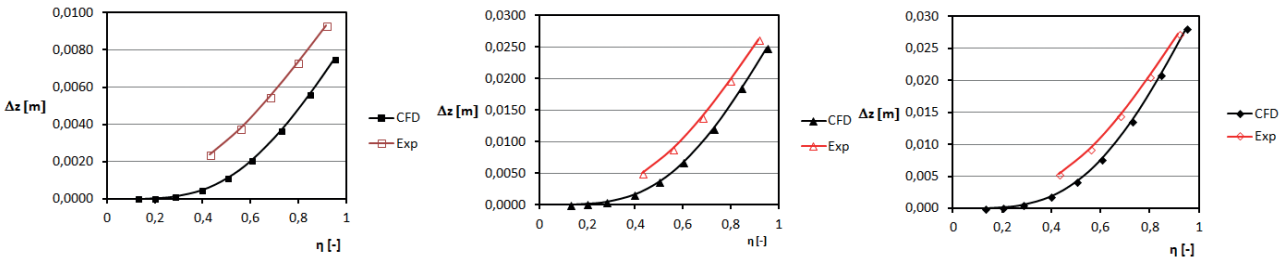


Figure 5.
Comparison of wing bend from the wind tunnel test and CFD for AoA=0° (left), AoA=3° (middle) and AoA=4° (right)

deformation data from the wind tunnel test for 4 AoA was utilised. Since the exact values of C_L were not obtained during the wind tunnel test, the wind tunnel model deformation data was interpolated to the same values obtained from CFD simulations.
The comparison of the wing twist distribution along a wing span for three AoA's is depicted in Figure 4. It is possible to see that that the wing's twist for AoA 0 and 3 is predicted by CFD simulations very well and almost within the error of the measurement. For the AoA=4 the wing twist is slightly higher in comparison with the wind tunnel model.
The effect of aeroelasticity is clearly seen on wing bend depicted in Figure 5. The wing deformation is related to the rigid model. Almost constant differences of vertical coordinates, in order of 0.002 m, along the wing span is observed for AoA=0 while for higher AoA's the differences of vertical coordinates are smaller towards the wingtip, from 0.003 m to 0.0015 m.
The effect of the wing bend is not as important to the aerodynamic forces and CP as wing twist. But on the other hand the better prediction of the wing bend together with wing twist is very important for a better match of the results and for development of the real flow behaviour.

On-going Research / Outlook

A follow on investigation will look at computational analyses of other Reynolds and Mach numbers, and also the effect of different setting of the empennage. The on-line mesh deformation (deformation of CFD grid every time step during simulation) with utilisation of all 20 modes of the wing and the rest of the model (especially empennage) will also be used and compared with presented results.
The effect of the wake from the wing on the empennage for low and high-speed conditions will be also investigated.

Conclusion

The main aim of this study was utilisation of the modal analysis of the wing for better prediction of the aerodynamic forces. The first six modes of the wing have been used during CFD simulations and the final wing deformation corresponding to particular AoA has been obtained. The wing's deformation (twist angle) is very well predicted and a very good agreement of the results has been achieved. In spite of consideration of the wing deformation some differences in CL, CD and the wing twist angle of the outer portion of the wing persist. The biggest differences in aerodynamic coefficients are related to the smaller AoA's. The reasons for these differences could be caused by the missing support system, by the simplification of the model used for modal analysis and by omission of the higher modes during CFD simulations.
This method can be used in case high fidelity methods, such as FSI, cannot be used. The only limit of this method is the hardware restriction because the perturbation grid of each considered mode has to be created and stored in the memory, but it has been shown that a quite small number of modes is sufficient for increasing the accuracy of the CFD results.

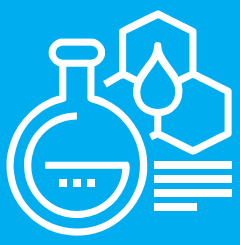
References

[1] ESWIRP project website: www.eswirp.eu.
[2] Lutz, T., et al. "Time-resolved Prediction and Measurement of the Wake past the CRM at high Reynolds number stall conditions", AIAA Paper 2015-1094, 2015. Doi: 10.2514/6.2015-1094
[3] Konrath, R., "High-Speed PIV Applied to Wake of NASA CRM Model in ETW Under High Re-Number Stall Conditions for Sub- and Transonic Speeds", AIAA Paper 2015-1095, 2015. Doi: 10.2514/6.2015-1095
[4] Rivers, M. B. and Dittberner, A., "Experimental Investigations of the NASA Common Research Model in the NASA Langley National Transonic Facility and the NASA Ames 11-ft Transonic Wind Tunnel (Invited)", AIAA Paper 2011-1126, 2011.
[5] Rivers, M. B., Hunter, C. A., and Campbell, R. L., "Further Investigation of the Support System Effects and Wing Twist on the NASA Common Research Model", AIAA Paper 2012-3209, 2012.
[6] Eberhardt, S., Benedict, K., Hedges, L., and Robinson, A., "Inclusion of Aeroelastic Twist into the CFD Analysis of the Twin-Engine NASA Common Research Model", AIAA Paper 2014-0251, 2014. Doi: 10.2514/6.2014-0251
[7] Keye, S. and Rudnik, R., "Validation of Wing Deformation Simulations for the NASA CRM Model using Fluid-Structure Interaction Computations", AIAA Paper 2015-0619, 2015. Doi: 10.2514/6.2015-0619
[8] Jakobsson, S. and Amoignon, O., "Mesh deformation using radial basis functions for gradient-based aerodynamic shape optimisation", Computers & Fluids, Vol. 363, No. 6, 2007, pp. 1119- 1136. Doi: 10.1016/j.compfluid.2006.11.002

Publications

[1] Vrchota, P. and Prachar, A., "Improvement of CFD aerodynamic characteristics using modal deformation", Proceedings of the 55th Israel Annual Conference on Aerospace Sciences, Tel Aviv & Haifa, Israel, 2015.
[2] Vrchota, P. and Prachar, A., "Using wing modal deformation for improvement of CFD results of the ESWIRP project", Proceedings of the 5th CEAS Air & Space Conference, Delft, The Netherlands, 2015.
[3] Vrchota, P. and Prachar, A.," Improvement of CFD results of NASA CRM by on-line mesh deformation using modal analysis of a wing", 33rd AIAA Applied Aerodynamics Conference, Texas, USA, 2015, Doi: 10.2514/6.2015-3153
[4] Vrchota, P., Prachar, A. and Smid, M., "Improvement of CFD Results of NASA CRM by Wing Modal Analysis", Journal of Aircraft (to be published)

05 | CHEMISTRY AND MATERIAL SCIENCES



BENCHMARKING OF NEW ACCURATE MULTI-REFERENCE COUPLED CLUSTER METHODS

Research institution:
J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Principal investigator:
Jiří Brabec

Researchers:
Jiří Pittner, Libor Veis, Martin Srnec, Jakub Lang

Project partners:
Edoardo Apra and Karol Kowalski (Pacific Northwest National Laboratory, Richland, WA, USA)

Project ID:
OPEN-7-42

Introduction

Many important chemical processes involve strongly correlated systems, which are very difficult to describe. One often encounters this problem in studies of bond breaking, low-spin open-shell and excited electronic states, diradicals, transition states and in compounds containing transition metal atoms, so they inevitably arise in most applied research areas, like material science, photovoltaic or drug design. For a proper description of these problems we need an accurate approach capable of capturing both the static and dynamic correlation effects. The current state-of-the-art methods are based on the multi-reference coupled cluster approach. Recently, we have developed a new universal state-selective correction [1-2] for these methods, which for smaller systems significantly improves their accuracy with respect to full configuration interaction (FCI). The benchmark results will be very helpful for understanding the correction performance and its application to larger and more challenging systems.

Results and Methods

The highly scalable implementation of the USS correction was developed using the Tensor Contraction Engine (TCE) [3] available in NWChem [4]. NWChem is a suite of computational chemistry modules for large scale molecular simulations. The parallelisation of NWChem is largely built upon the Global Array (GA) library, which focuses on the requirements of distributed data algorithms in the computational chemistry area, with heavy emphasis on the use of dense matrices. GA forms an abstraction layer that alleviates the task of the developer by isolating most of the complexities involved in parallelising software that make use of dense matrices. It relies on several components: Message Passing library, the ARMCI one-sided communication library and a memory allocator (MA library). Most of the NWChem modules make very little direct use of MPI, since the majority of the communication is managed by the ARMCI one-sided communication library. Each process in a GA based parallel program can asynchronously access logical blocks of physically distributed dense arrays, without a need for explicit cooperation by other processes. GA is being actively developed and offers interfaces to parallel linear algebra libraries (such as PeIGS and ScaLAPACK).

The USS(2) formalism requires an inclusion of the subset of triple and quadruple excitations with respect to each reference, which need to be evaluated and stored. Due to storage requirements of the full USS(2) formalism, we also developed a perturbative version of the USS(2) formalism where three- and four-fold excited contributions are evaluated in an on-the-fly manner. The efficiency of the full USS(2) and perturbative USS(2) approaches in describing quasi-degenerate electronic states is evaluated by applying USS(2) corrections to the BWMRCC and Mk-MRCC approaches with singles and doubles (BW-MRCCSD and Mk-MRCCSD methods, respectively).

Benchmarks

We have performed several tests. The approaches based on the Jeziorski-Monkhorst ansatz are not invariant with respect to rotations of active orbitals. In particular, it is very important to qualitatively estimate how much active orbital rotations impact the USS energies. To address this question we have performed two benchmark calculations for the twisted ethylene using the cc-pVTZ basis set, where two active orbitals were continuously rotated/mixed using rotation angles falling into the [0°, 45°] interval.

As seen in Figure 1(a), the USS correction significantly improves the invariance of the BWMRCCSD method, where 9.2 mhartree deviation (calculated as a difference between the maximum and minimum values of the energy as a function of rotation angle α) is reduced to only 2.4 mhartree when using USS(full), or to 2.8 mhartree when using USS(pt) correction. Smaller discrepancies are a consequence of the fact that the USS(full) and USS(pt) curves are significantly “flatter”, which indicates that these methods are less sensitive to the orbital rotation than the original BWMRCCSD formulation. For comparison, the

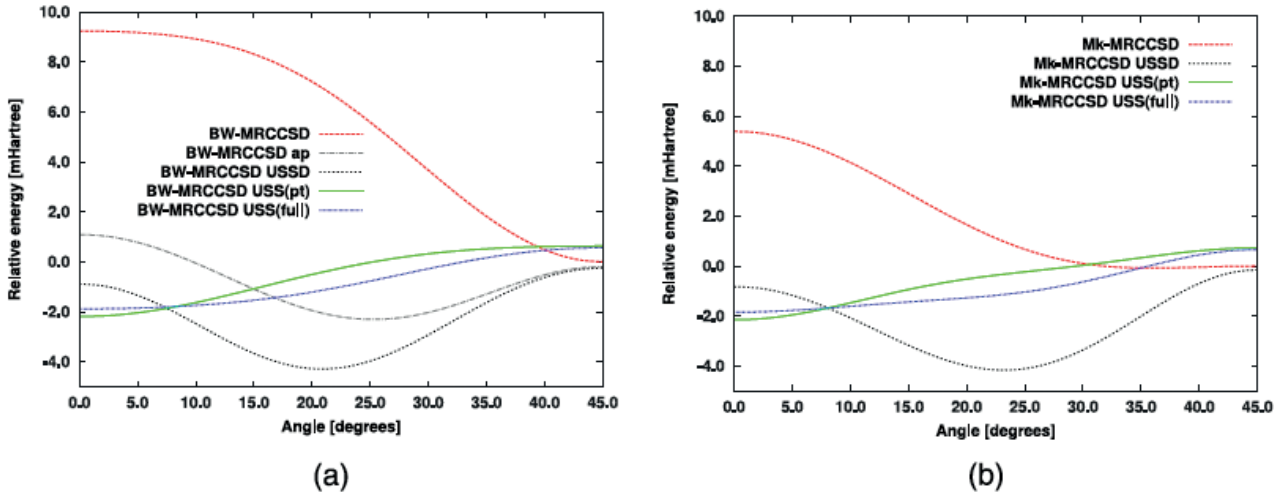


Figure 1.
Comparison of the singlet state energies in the twisted ethylene model, where active orbitals were rotated α degrees. The relative energies are calculated with respect to the uncorrected MRCCSD energy at α = 45.

USSD energy shows a minimum around 22° and the largest discrepancy is 4 mhartree. The a posteriori correction gives a similar non-monotonic shape to USSD correction with the largest difference 3.4 mhartree. As shown in Figure 1(b), the shape of Mk-MRCCSD is very similar to the BW-MRCCSD case and demonstrates a significant effect of mixing active orbitals on the resulting energies.

The medium-size α,3-dehydrotoluene diradical has experimentally confirmed the open-shell singlet as the ground state, which is in contrast to α,2- or α,4-isomers with the triplet ground state. The experimental data indicate that the S-T gap is smaller than 5 kcal/mol. In Table 1 we compare the USS-corrected and uncorrected BW-MRCCSD and Mk-MRCCSD energies for the singlet ground state of this diradical. For comparison, we also provide CASSCF and CASPT2 results for different active space sizes. In the first benchmark, we used canonical RHF orbitals, while in the second set, which is much more challenging, we rotated two active orbitals for 45° in order to show the non-invariance to active orbital rotations. Due to the numerical cost of the full USS corrections, we only calculated the USS(full) energies for the cc-pVDZ basis set.

As shown in Table 1, we find that the differences between the USS(pt) and USS(full) corrected MRCCSD energies (BW and Mk) for HF orbitals in the cc-pVDZ basis set are very small and amount to 0.1 kcal/mol, while for rotated active orbitals the analogous differences reach 0.7 kcal/mol (Mk) and 0.5 kcal/mol (BW). For both (cc-pVDZ and cc-pVTZ) basis sets, the differences between HF orbitals based BW-MRCCSD USS(pt) and Mk-MRCCSD USS(pt) results are negligible (less than 0.1 kcal/mol), whereas for rotated orbitals these differences grow to 0.7 kcal/mol and 0.8 kcal/mol for the Mk-MRCCSD and BW-MRCCSD USS(pt) formulations, respectively. However, this is much less than the 14.1 kcal/mol (cc-pVDZ) or 11.9 kcal/mol (cc-pVTZ) difference between uncorrected MRCCSD methods (Mk-MRCCSD-BW-MRCCSD) for α = 45. The effect of orbitals rotation on the singlet-triple separation is especially strong for the BW-MRCCSD approach. While for α = 0 the singlet triplet splitting obtained with uncorrected BWMRCCSD approach in cc-pVDZ basis set corresponds to 2.8 kcal/mol, for α = 45 the analogous singlet-triplet value reaches –15.5 kcal/mol. It is worth mentioning that despite these huge differences, the USS(pt) approach can “stabilise” the BW-MRCCSD formalism and yield reasonable results of 2.8 and 4.9 kcal/mol for α = 0 and α = 45, respectively.

These results provide a very good illustration of the universality of the USS corrections. The singlet-triplet MRCCSD/cc-pVTZ energy with the USS(pt) correction is approximately 2.6 kcal/mol, which is in agreement with the experimental observations and also with the values reported in Refs. [5] , where calculations were performed at the CI or MCSCF(8,8)

Basis set	cc-pVDZ	cc-pVTZ	cc-pVDZ	cc-pVTZ
Rotational angle	$\alpha = 0$		$\alpha = 45$	
CCSD (triplet)	−269.424 668	−269.756 928	−269.424 668	−269.756 928
BW-MRCCSD	2.8	2.6	−15.5	−14.2
BW-MRCCSD USSD	3.7	3.4	8.3	9.5
BW-MRCCSD USS(pt)	2.8	2.6	4.9	6.4
BW-MRCCSD USS(full)	2.9		5.4	
Mk-MRCCSD	3.3	3.0	−1.4	−2.3
Mk-MRCCSD USSD	3.6	3.3	6.7	7.7
Mk-MRCCSD USS(pt)	2.8	2.6	4.2	5.6
Mk-MRCCSD USS(full)	2.8		4.9	
CASSCF(2,2)	−0.5	−0.5	−0.5	−0.5
CASSCF(4,4)	3.0	3.1	3.0	3.1
CASSCF(8,8)	3.0	3.2	3.0	3.2
CASPT2(2,2)	0.4	0.4	0.4	0.4
CASPT2(4,4)	1.8	1.7	1.8	1.7
CASPT2(8,8)	1.9	2.0	1.9	2.0

Table 1.
The scalability of the USS(pt) and USS(full) implementations for the BW-MRCCSD calculations on α,3-dehydrotoluene using the cc-pVDZ basis set.

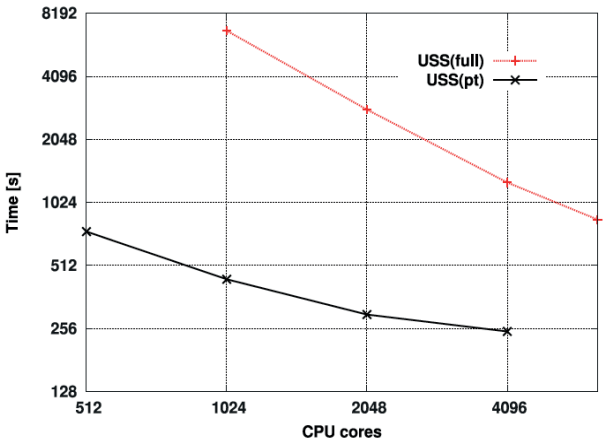


Figure 2.
The scalability of the USS(pt) and USS(full) implementations for the BW-MRCCSD calculations on α,3-dehydrotoluene using the cc-pVDZ basis set.

level using 6-31G** or cc-pVDZ basis set. The CASSCF(2,2) S-T energy corresponding to our CASPT2/ cc-pVTZ optimised geometry is -0.5 kcal/ mol, using the larger model space (4,4) or (8,8) the energy increases to 3 kcal/ mol. The CASPT2 energies behave more consistently with the size of the model space compared to the CASSCF method. The CASPT2(8,8) gap is approximately 1 kcal/ mol smaller than the corresponding USS-MRCCSD results. In contrast to BW- and Mk-MRCCSD approaches, for rotated orbitals these methods take advantage of the invariance to active orbital rotations.

Concerning the computational cost, the perturbative USS approach has two advantages. First, the required computational time for USS(pt) is much smaller than for the full USS calculation. Second, since the USS(pt) approach calculates the triple and quadruple projections in an on-the-fly manner it eliminates the global memory bottleneck of the USS(full) associated with the need of storing expensive recursive intermediates. In Figure 2, we report timings of the USS algorithms for α,3-dehydrotoluene. For example, calculation of the BW-MRCCSD USS(pt) energy on 1,024 cores takes 441 s compared to 6,762 s for the BW-MRCCSD USS(full) calculation, which means that the speed up factor is more than 15 times bigger. It is interesting to notice that the USS(full) correction requires more than 6400 cores in order to achieve the same timing as the USS(pt) calculation on 512 cores. The scalability of the USS(pt) code on more than 1,000 cores is still good even if the number of tasks is much smaller than in the USS(full) calculation.

On-going Research / Outlook

We plan to apply the developed methods on larger systems and also utilise MIC accelerators for the (T) correction, which has been efficiently ported for this architecture.

Conclusion

We proposed and implemented an efficient perturbative approximation of the USS correction for the BW-MRCCSD and Mk-MRCCSD methods. We demonstrated on several numerical examples that the perturbative USS approximation substantially reduces the computational cost of the full USS correction without any significant effect on the accuracy of resulting energies, compared to the full USS algorithm. For example, calculations of the S-T energy gap for twisted ethylene using the cc-pVTZ basis set demonstrated that the energy difference between USS(pt) and USS(full) energies is as small as 0.3 mhartree. For larger system, such as the α,3-dehydrotoluene, the analogous difference only corresponds to 0.16 mhartree.

On several examples, we demonstrated that the USS energies are significantly less sensitive to the rotations of active orbitals than the iterative BW-/Mk-MRCCSD formulations. The performance tests show a good scalability of the USS(pt) implementation and a significant speed up compared to the original USS(full) implementation, while preserving almost full accuracy.

References

[1] Brabec, J.; van Dam, H. J. J., Pittner, J. and K. Kowalski. Universal state-selective corrections to multi-reference coupled-cluster theories with single and double excitations. J. Chem. Phys. 2012, 136, 124102. Doi: 10.1063/1.3692969

[2] Banik, S.; Ravichandran, L., Brabec, J., Hubac, I., Kowalski, K. and J. Pittner. Iterative universal state selective correction for the Brillouin-Wigner multi-reference coupled-cluster theory. J. Chem. Phys. 2015, 142(11), 114106-. Doi: 10.1063/1.4914311

[3] Hirata, S. Tensor contraction engine: Abstraction and automated parallel implementation of configuration-interaction, coupled-cluster, and many-body perturbation theories. J. Phys. Chem. A, 2003, 107(46), 9887-9897. Doi: 10.1021/jp034596z

[4] Valiev, M.; Bylaska, E.; Govind, N.; Kowalski, K.; Straatsma, T.; van Dam, H.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. and W. de Jong. Nwchem: Open source high-performance computational chemistry. Comput. Phys. Commun. 2010, 181, 1477. Doi: 10.1016/j.cpc.2010.04.018

[5] Wenthold, P.G.; Wierschke, S.G.; Nash, J.J. and R. R. Squires. Experimental and theoretical studies of the mechanism and thermochemistry of formation of alpha.,n-dehydrotoluene biradicals from gas-phase halide elimination reactions. J. Am. Chem. Soc. 1944, 116(16), 7378-7392. Doi: 10.1021/ja00095a048

Publication

[1] Brabec, J., Banik, S., Kowalski, K., Pittner, J. Perturbative universal state-selective correction for state-specific multi-reference coupled cluster methods. The Journal of Chemical Physics. 2016, 145(16), 164106. Doi: 10.1063/1.4965826

ATOMIC-SCALE DESIGN OF OPTICALY ACTIVE NANOSTRUCTURED AMORPHOUS SILICA

Research institution:
Czech Technical University in Prague

Principal investigators:
Antonio Cammarata, Luisa Spallino, Paolo Nicolini, Benjamin Irving

Project ID:
OPEN-6-8

Introduction

To satisfy the requirements of ideal sensing, over affordability and sustainability, a luminescent probe must guarantee high sensitivity. In this context, nanostructured amorphous silica (SiO_2) is a promising material thanks to its peculiar features: low manufacturing costs, low toxicity, chemical stability, good resistance to harsh environment [1-3] and high photon emissivity in a wide range of the electromagnetic spectrum, from IR to Vis-UV [4-7]. A peculiar structured photoluminescence (PL) was singled out at room temperature [8] in the spectrum emitted in a vacuum by SiO_2 nanoparticles. The PL, due to intrinsic nearly-isolated defects [9], is characterised by very sharp vibronic lines, between 3.0 and 3.5 eV, attributed to the coupling of the electronic transition with two localised vibration modes of frequency 1370 and 360 cm^{-1}, respectively. These exceptional spectral features, combined with the pronounced sensitivity to the presence of molecules in the environment responsible for the luminescence quenching, are the key features to exploit SiO_2 nanoparticles as luminescent sensors of specific molecules. Although a clear spectroscopic characterisation is given, the origin of the structured PL is still unknown. This is a limit for the exploitation of such properties in sensing-related applications. Computational techniques can boost the interpretation of experimental data efficiently driving the design of novel SiO_2 based materials with improved or new optical properties.

Results and Methods

In a previous work [12], the vibronic progression of frequency 1370 cm^{-1} had already been observed in a vacuum, together with a broad blue PL spectrum also observable in air. The authors proposed that a common defect is responsible for both the blue and structured emission, consisting of two faced defects: a dioxasilyrane and a silylene. The vibrational progressions, observed under a vacuum treatment, was interpreted as the emission from the photo-induced excited state of O_2 that originates by the transient photolysis of the dioxasilyrane. The model of the defects pair was formerly proposed only to explain the origin of the blue PL band and it is based on the good agreement between ab initio calculations and the spectroscopic features of the blue band [13]. Thanks to a fine spectroscopic characterisation, our work has shown, for the first time, that the blue PL component and the vibronic progressions are disentangled, having different emission, excitation and decay properties [8]. Moreover, in the same work, we identify the vibrational progression with frequency 360 cm^{-1}, unresolved in previous experiments reported in literature. We have proposed that the observation of two vibration modes associated to the same electronic transition is consistent with a defect structure containing more than two atoms. However, the observed properties are not clearly tied with the electronic and geometric features of a specific optically active centre at the atomic level and their assignment to a specific group at the SiO_2 surface is not yet supported. For this reason, this computational project has the final goal to formulate a new model to support the recent experimental findings.

We performed ab initio DFT simulations on amorphous silica nanoclusters. In particular, in order to identify the interatomic interactions and atom dynamics relevant to the optical activity, we analysed the local electronic density of states, charges and orbital polarisation, to characterise the local electronic features.

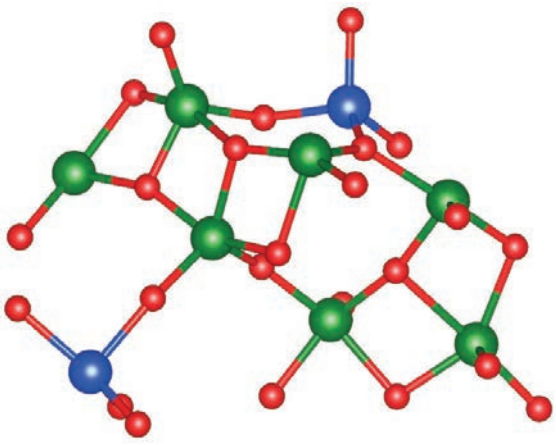


Figure 1.
Example of optically active SiO clusters.
Red and blue spheres represent oxygen and silicon atoms, respectively, while green spheres show the position of Si atoms with coordination different than 4.

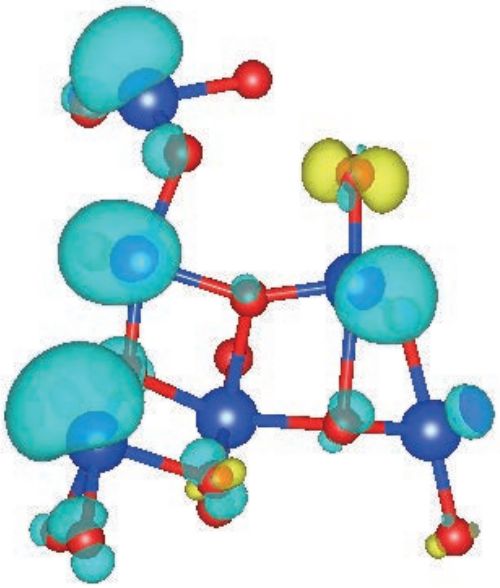


Figure 2.
Spin density isosurfaces (yellow = positive, blue = negative) of overcoordinated and undercoordinated Si cluster. Blue and red spheres represent silicon and oxygen atoms, respectively. Magnetisation excess is believed to be responsible for the optical activity.

We exploited standard and advanced electro-vibronic analysis techniques [10-11] that the PI recently formulated and successfully applied to the study of other classes of optically active materials [14].

Use of the IT4Innovations HPC facilities for the OPTOSILICA project has been crucial for enhancing our knowledge of optically active centres in amorphous silica. The reference system for DFT calculation was a fragment of amorphous silica obtained by means of MD simulations. Our computational investigations, based on both quantum and classical mechanics, in conjunction with group theory techniques, have focused on how interatomic interactions affect local atomic topology in order to determine the formation mechanism of the optically active defect and the electronic distribution responsible for the optical activity.

We identified small clusters of over- and undercoordinated silicon atoms, as optically active centres (Figure).

In particular, we noticed that optical transitions can be ascribed to the magnetisation excess found in the environment of such clusters (Figure). Si-overcoordinated clusters are characterised by an excess of positive spin magnetisation with smaller magnitude compared to the negative magnetisation found on Si-undercoordinated clusters. In Si-overcoordinated clusters, magnetisation excess is accounted by the bridging oxygen atoms, contributing with the 2p orbitals which are characterised by a non-isotropic spatial distribution, due to the distorted local geometric environment that favours interactions with first neighbouring silicon atoms. On the other hand, in si-undecoordinated clusters, we observe that the magnetisation effect is accounted by the silicon atom, showing a characteristic lone-pair.

On-going Research / Outlook

Other results obtained from the simulations performed within the present project are still under evaluation; such results must be reviewed within the framework of joint experimental analyses, which are still ongoing at the moment of the present report. These results will be reported in a manuscript in preparation together with those discussed above.

Conclusion

Luminescent activity in amorphous silica nanoclusters has been investigated. We identified over- and undercoordinated SiO clusters as possible optically active centres. Such clusters are characterised by an excess of spin magnetisation and anisotropic spatial distribution of the electronic density. The electronic anisotropy is due to the distorted environment of the optically active clusters; such distortion favours the interactions among first neighbouring silicon atoms.

References

[1] Lipani E.; et al. Langmuir. 2013, 29, 3419. Doi: 10.1021/la304689d
[2] Montalti M.; et al. Chem. Rev. Soc. 2014, 43, 4243. Doi: 10.1039/C3CS60433K
[3] Helle M.; et al. ACS Nano. 2013, 7, 8645. Doi: 10.1021/nn402792a
[4] Bonacchi S.; et al. Angew. Chem. Int. Ed. 2011, 50, 4056. Doi: 10.1002/anie.201004996
[5] Brites C. D. S.; et al. Langmuir. 2012, 28, 8190. Doi: 10.1021/la300288j
[6] Davies G. L.; et al. J. Mater. Chem. 2012, 22, 7358. Doi: 10.1039/C2JM16086B
[7] Zhang C.; et al. Chem. Soc. Rev. 2012, 41, 7938. Doi: 10.1039/C2CS35215J
[8] Spallino L.; et al. Phys. Chem. Chem. Phys. 2014, 16, 22028. Doi: 10.1039/C4CP02995J
[9] Spallino L.; et al. J. Phys.: Condens. Matter. 2015, 27, 365301. Doi: 10.1088/0953-8984/27/36/365301
[10] Cammarata A.; et al. J. Chem. Phys. 2014, 141, 114704. Doi: 10.1063/1.4895967
[11] Cammarata A.; et al. Inorg. Chem. 2015, 54, 5739. Doi: 10.1021/acs.inorgchem.5b00431
[12] Anjiki A.; et al. J. Phys. Chem. 2012, C 116, 15747. Doi: 10.1021/jp303359v
[13] Uchino T.; et al. Phys. Rev. B. 2006, 73, 233203. Doi: 10.1103/PhysRevB.73.233203
[14] Cammarata A.; et al. Chem. Mater. 2014, 26, 5773. Doi: 10.1021/cm502895h

QUANTUM SIMULATIONS OF DOPED HELIUM NANODROPLETS I. PHOTOABSORPTION SPECTRA OF CHARGED HELIUM CLUSTERS

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Rajko Ćosić

Researcher:
René Kalus

Project partners:
University of Ostrava,
Universite Paris EST Marne-la-Vallee

Project ID:
IT4I-10-5

Introduction

The present project represents the starting step of a broader research intention aiming at detailed investigations of helium nanodroplets doped with charged impurities. Cold helium clusters are very interesting objects due to their special properties, such as superfluidity. At sufficiently low temperatures, the helium can be used as an inert environment for observing other particles or molecules with almost no effect on the solvated matter. This behaviour is widely exploited in the so-called ultra-cold chemistry, where the special states of the reactants and products are studied. The need for good understanding of properties of doped helium clusters follows directly from the former.

The project focuses on the physical properties of selected small charged helium clusters, He_N^+ (N = 3, 4, 10), and their interaction with electromagnetic radiation is investigated. The convergence of the photoabsorption spectra of He_N^+ has been studied in terms of the number of so-called ghost particles (see below) and also in dependence on the temperature. The project also served as a test of the path integral Monte Carlo (PIMC) routines, which are currently developed in our group and which are used for quantum simulations at nonzero temperatures.

In this paper, the focus is not put to the results of the simulations in the physical sense (which are planned to be published elsewhere), but to the computational point of view of such simulations. As we see in the Results and methods section, PIMC simulations without any advanced features can lead to very serious difficulties resulting from the finite simulation length. This has been observed during our simulations and has lead us to implement new features, which will be utilised to mitigate the problem.

Results and Methods

The PIMC method is based on the Feynman path integrals [1] and the decomposition of the density matrix into kinetic and potential parts [2], which enable us to treat the kinetic and the potential part of the density matrix operator separately.

$$e^{-\beta(T+V)} = \lim_{M\to\infty}\left[e^{-\beta T/M}e^{-\beta V/M}\right]^M$$

where T is the kinetic part of the Hamiltonian operator, V is the potential part of the Hamiltonian operator and $\beta = 1/k_B T$ is the inverse temperature.

A classical – quantum [3] mapping scheme is then used to enable us to treat the system of quantum particles classically with the particles replaced by closed chains of so-called ghost particles introducing quantum effects into purely classical calculations (the lower the temperature, the more ghosts are needed). Thus, Monte Carlo (MC) methods of classical simulations on polymers can be used.

Unfortunately, the Metropolis algorithm used for the MC sampling does not allow us to parallelise the process of sampling. Two other possibilities to parallelise the code are a) to parallelise the calculation of the potential energy and b) to involve the parallel tempering algorithm [4] to enhance the simulation efficiency. The parallelisation of the potential energy is done using the MPI standard and its scalability depends on the computational complexity of the potential part of the energy. In our calculations, the diatomics in molecule (DIM) approach [5] with the computational complexity $O(N_3)$ is used. This complexity grants us very good scalability (the weak scalability is ≈ 80%) up to $N_c \approx N_{gp} / 4$ where N_c is the number of computational cores and N_{gp} is the number of ghost particles.

Another possibility, implemented during the project, is the parallel tempering algorithm [6], which enhances the ergodicity of the random walk and also speeds up the convergence. The idea is to simulate many systems simultaneously, each at a different temperature level. By periodic attempts to switch the configurations between selected neighbouring temperatures we enable the cluster to overcome barriers which would be otherwise difficult to cross in the finite number of steps. The attempt leading to crossing the barrier between two minima of a double-well po-

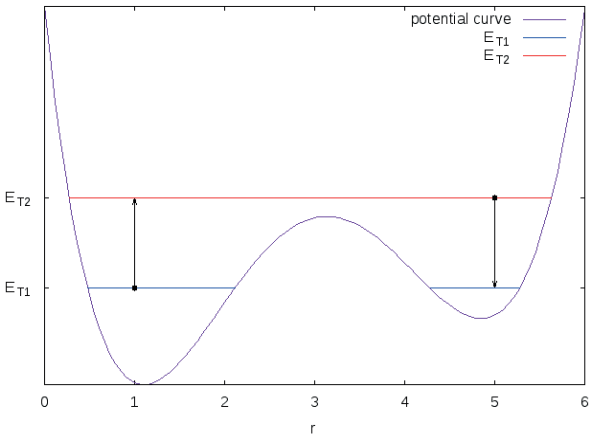


Figure 1.
Crossing the barrier between two distinct minima of the double-well potential

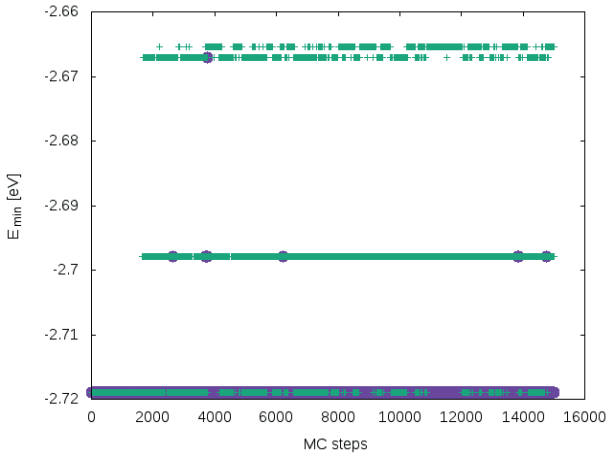


Figure 2.
Comparison of the simulation with broken ergodicity (violet) with the ergodic simulation (green). Violet marks stay at the lowest minimum energy level, while the green marks visit all the minima levels during whole simulation.

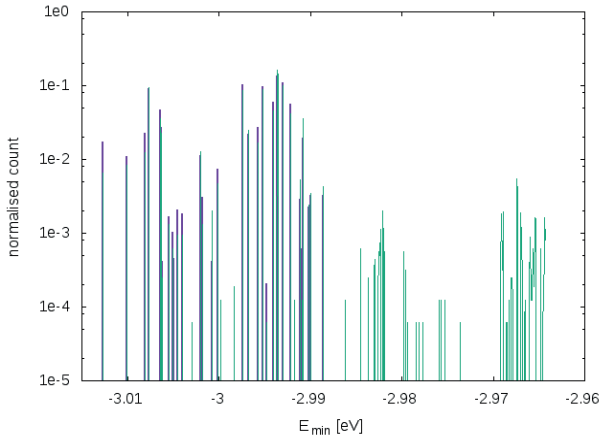


Figure 3.
Abundance of various minima found during the He_{10}^+ simulation via PIMC method (violet) and via the DMC method (green).

tential is depicted in Figure 1. The nature of this algorithm, which simulates the temperature levels almost independently, grants us almost linear scalability. Within this project, only implementation and a short preliminary test were done. This algorithm is ready to test and use within ongoing projects.

However, even if both the parallelisations scaled almost linearly, the computer time will increase rapidly for the increasing number of ghost particles. The number of ghost particles depends not only on the quantum nature of the cluster, but also on the temperature (which, in fact, affects the quantum nature of the system) at which the system is simulated. Our tests have shown that for smaller clusters (N = 3, 4) the number of 512 ghost particles will be sufficient to get converged photoabsorption spectra at the temperature level of 2 K. For medium cluster (N = 10), the sufficient number of ghost particles was 1,024 at the temperature level of 2 K.

Although the Metropolis sampling ensures the ergodic random walk in the limit of the infinite number of steps, our simulations have shown that the ergodicity is not always achieved globally. The basin of attraction of the local minimum which is closest to the starting point is explored well, whereas the other minima are seldom visited. A possible way to enhance the ergodicity of the random walk is the parallel tempering algorithm. According to our preliminary tests performed within the project, the standard parallel tempering algorithm, known from classical MC simulations, does not seem to be very helpful in the PIMC simulations if the paths are represented standardly in the Cartesian system.

The problem with the ergodicity has particularly occurred in the simulations of He_4^+ (see Figure 2), where the barrier between the potential minima happens to be too high to be overcome by the random walk in the finite number of steps. In Figure 2, almost all the violet dots correspond to the lowest minimum at -2.7189 eV, while the ergodic simulation (green) visits all the minima.

In Figure 2, almost all the violet dots are located at the energy level of the lowest minimum at -2.7189 eV. This means that during the simulation the other minima stay almost unvisited. The ergodic behaviour is depicted by green crosses. These crosses can be found at all the energy levels adjacent to the minima during whole simulation.

On the other hand, the ergodicity of He_3^+ and He_{10}^+ simulations seem to be all right. The potential energy surface of He_3^+ has only one minimum and hence there is no problem with achieving ergodicity. The minima found during the He_{10}^+ simulations have been compared to ones found by a DMC simulation (see Figure 3). The height of the bars which represents the rate at which the minima were sampled in the PIMC simulation may not correspond to the height of the bars representing the rates of minima sampled via the DMC approach as the methods sample the configurations with slightly different probabilities. The DMC calculation has found some minima that have not been sampled via PIMC calculation. However, the rate of sampling of these

minima is lower by an order of magnitude than the rate at which the minima found by both methods were sampled.
This good ergodicity is probably caused by the low barriers between the minima of the potential which are likely to be crossed during the simulation.

On-going Research / Outlook

A series of projects focused on the modelling of the behaviour of doped helium nanodroplets is planned in the future. The present project was dedicated to computing the photoabsorption spectra of selected He_N^+ clusters. A subsequent project, which is now being solved, focuses on structural changes in the ionised helium clusters. The next stage will naturally consist in extending the computations on the heterogeneous bosonic clusters, specifically the ones in which the charge carrier is one of the other rare gases.

Conclusion

Computations done within the project have clearly shown the need of advanced sampling methods, such as the Fourier path integral Monte Carlo (FPIMC) [7] method or multigrid based path integral Monte Carlo (MGPIMC) [8] method to speed up the sampling process and the parallel tempering approach to fix the broken ergodicity problem. Recently, the work on the implementation of the FPIMC method and its connection with the parallel tempering algorithm has been started. The FPIMC and MGPIMC approaches enable us to implement the parallel tempering algorithm very efficiently using the fact that for higher temperatures the number of ghost particles is significantly lower than for low temperatures. Both approaches also preserve the kinetic part of the action within the temperature level which is, according to our simulations, crucial for the efficiency of the parallel tempering algorithm in the PIMC simulations.

References

[1] R. P. Feynman, A. R. Hibbs, Quantum Mechanics and Path Integrals, Dover, Mineola & New York 2005. ISBN 978-0486477220
[2] H. F. Trotter, Proc. Am. Math. Soc. 10 (1959) 545. Doi: 10.1090/S0002-9939-1959-0108732-6
[3] Ceperley D.M. Rev. Mod. Phys. 67 (1995) 279, Doi: 10.1103/RevModPhys.67.279
[4] M. Falcioni, M. W. Deem, J. Chem. Phys. 110.3 (1999) 1754. Doi: 10.1063/1.477812
[5] Ellison F. O. J. Am. Chem. Soc. 85 (1963) 3540. Doi: 10.1021/ja00905a002
[6] C. J. Geyer, Proceedings of the 23rd Symposium on the Interface in Computing Science and Statistics, American Statistical Association, New York, 1991.
[7] J. D. Doll, L. E. Myers, J. Chem. Phys. 71 (1979) 2880. ISSN: 0021-9606
[8] J. Goodman, A. D. Sokal, Phys. Rev. D 40 (1989) 2035. Doi: 10.1103/PhysRevD.40.2035

Publication

[1] Ćosić, R., Karlický, F., Kalus, R.: Photoabsorption spectra of selected small He_N^+ clusters. In preparation.

MOLECULAR DOCKING ON HIGH PERFORMANCE COMPUTING CLUSTERS

Research institution:
University of Hradec Králové

Principal investigator:
Rafael Doležal

Researchers:
Teodorico Ramalho, Tanos Franca, Kamil Kuča

Project partners:
Biomedical Research Centre in Hradec Králové, Federal University of Lavras, Military Institute of Engineering in Rio de Janeiro

Project ID:
OPEN-5-7

Introduction

Molecular docking is a simulation process belonging to molecular mechanics that tries to specifically accommodate a small molecule (e.g. ligand) into a larger molecule of a receptor or enzyme (Figure 1). Prediction of the binding Gibbs free energy associated with the preferred binding mode can serve as an indicator of the ligand interaction strength, and thus as an estimation of biological activity of the ligand. Importantly, molecular docking can be used not only for modelling of non-covalent interactions between a single ligand and one receptor but also for large scale screening of virtual ligand libraries to discover novel lead structures.

In general, the molecular docking problem is solved as an optimisation task in which the global minimum in the potential energy landscape is searched for. Such minimisation of the scoring function is not a trivial task and has to be solved numerically. Generally, these demanding calculations involved in extensive virtual structure-based screenings utilise parallel and distributed high performance computing (HPC) approaches to achieve the solution in a reasonable time span. Fortunately, the molecular docking task can be ideally split into independent processes conforming to Markov chains and, thus, fully exploit the power of computer grids or supercomputers.

Within this study, which has already been reported [1], we compared the performance of the Anselm supercomputer in two types of calculations: 1) “pleasingly” distributed calculations using the AutoDock Vina 1.1.2 open source program along with a PBS Pro 12 scheduler and a self-developed distribution script, 2) message passing interface (MPI) based application of the VinaLC program, compiled from an open-source code implementing an advanced distribution scheme through the Infiniband QDR network (latency 1.3 μs, bandwith 40 Gbps, fully non-blocking fat-tree).

In fact, we reserved from Anselm only 8 nodes with 16 cores (Intel Sandy Bridge E5-2470, 2.3 GHz, 128 CPUs in total) and performed 200 similar molecular docking calculations differing only in random initialisation of the Monte Carlo based optimisation. The calculations were aimed at finding the top-scoring binding modes of one potential drug candidate for Alzheimer’s disease treatment in an X-ray model of human acetylcholinesterase (hAChE) (PDB ID: 4EY7). Comparing the designed calculations enabled us to discriminate which distribution scheme is more suitable for extensive structure-based virtual screening with utilisation of greater HPC systems. However, this study especially focuses on computational chemistry benefits rather than on detailed investigation of typical HPC parameters.

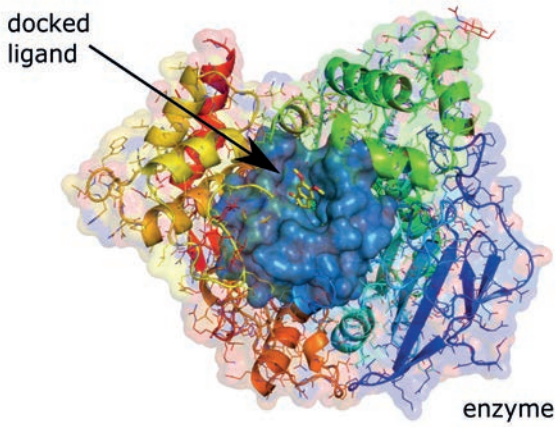


Figure 1.
Optimal binding pose of a ligand in an enzyme

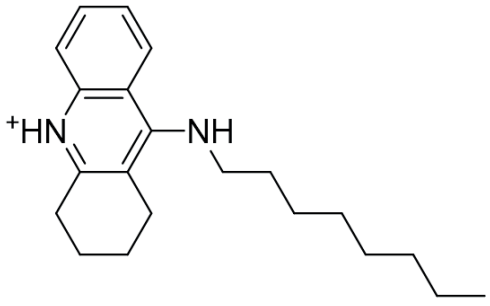


Figure 2.
An hAChE inhibitor used as the ligand in the docking studies

Results and Methods

The ligand molecule was constructed in a polarised mode corresponding to a physiological pH = 7.4, geometrically optimised with the PM3 semi-empirical method and Polak-Ribiére conjugate gradient algorithm in the HyperChem 8. program (Figure 2).

The ligand molecule was further processed in AutoDock tools and Open Babel 2.3.2 software. This pre-processing involved: definition of torsion angles, assigning Gastaiger partial atomic charges, condensation of non-polar hydrogens into heavy atoms and export of the model in PDQT format. Similarly, the X-ray model of hAChE (PDB ID: 4EY7) was downloaded from rcsb.org on-line database and prepared for docking in AutoDock tools software. The final enzyme model in PDBQT format contained only protein chain A with defined Gasteiger charges and condensed non-polar hydrogens. All water molecules and co-crystallised small molecules were removed from the original 3D model. Thus, the input files for the docking were: one ligand molecule and one target enzyme molecule. The ligand was treated as a flexible structure, while the enzyme as a rigid one. For all docking studies, a cube gridbox of 33 x 33 x 33 Å cantered at x = 10.698 Å, y = -58.115 Å and z = -23.192 Å was designed to encompass the catalytic and peripheral active sites of hAChE.

The first type of molecular docking calculation involved distribution of 200 as like calculations through the PBS scheduler over 8 nodes with 16 cores. Thus in one time, 8 jobs, each utilising 16 cores in parallel, run in the user-defined mini-cluster. Using simple time stamping, we monitored the start and end of each task. As the molecular docking engine, we utilised the AutoDock Vina 1.1.2. program with the exhaustiveness parameter set to 16. All 200 calculations were completed in 13 min 13 s, each task taking 20.245 s on average.

The second type of molecular docking was performed using the VinaLC program, which utilises MPI protocol for distributing the calculations over physically different nodes. From the computational chemistry point of view, VinaLC implements the same optimisation processes defined by the same scoring function. VinaLC, written in C++, adopts a master-slave MPI scheme which consists of three “for loops” for the master process and of one infinite “while loop” for each slave process. Briefly, the first master for loop imports every combination of the receptor and ligand, and distributes the docking jobs across different nodes through two pairs of MPI send/recv calls to the slave process. After receiving the data, the slave processes start the docking jobs. The second master loop performs the output data collecting and assigns the slave process another job. The third master loop is responsible for sending a finished job flag to free the slave processes, which finishes the MPI run. Within this master-slave MPI scheme, the master part takes care of job dispatching, whilst the slave processes control docking tasks until they are completed. Using the same setting and hardware, the 200 jobs were completed by VinaLC in 5 min 16 s. The mpirun parameters were set to split the task into 8 processes, each employing 16 CPUs.

By VinaLC, which significantly reduced idling processes and eliminated the time consumed by the PBS scheduler, a rough speed-up of 2.5 was achieved in comparison with

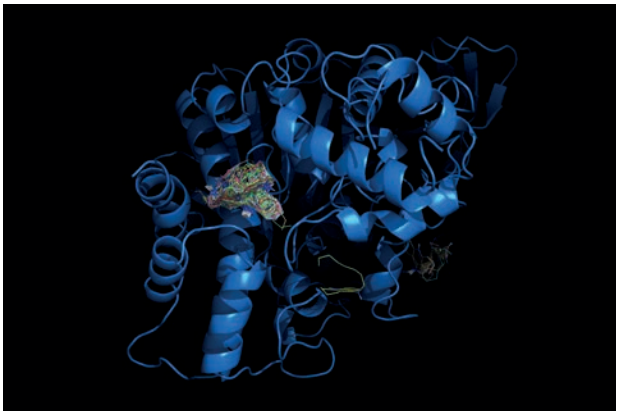


Figure 3.
200 superimposed top-scoring binding modes in hAChE obtained by Vina

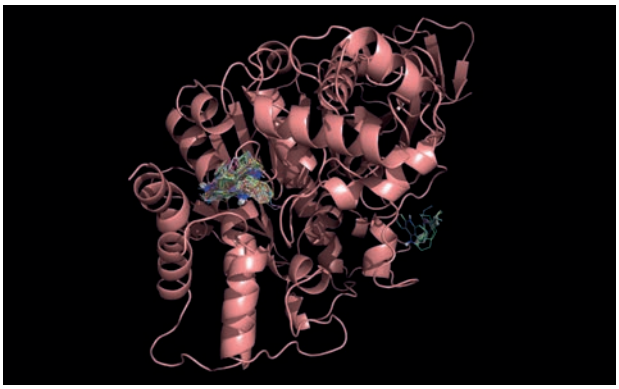


Figure 4.
200 superimposed top-scoring binding modes in hAChE obtained by VinaLC

Program	Minimum	Maximum	Mean	SD
Vina	-6.6	-4.9	-5.6	0.4
VinaLC	-6.5	-5.0	-5.6	0.4

Table 1.
Statistical analysis of predicted binding energies in the hAChE model by Vina and VinaLC programs

the parallelised calculations by Vina. Moreover, analyses of the top scoring docking poses resulting from both calculations revealed practically the same conformations and binding energies (Figures 3 and 4).

The basic statistics of the binding energy estimates provided by Vina and VinaLC are summarised in Table 1. It can be concluded that both methods predicted binding energy of -5.6 kcal/mol for the selected tacrine based hAChE inhibitor towards the 4EY7 model of hAChE.

Interestingly, VinaLC consumed less time to provide results of the same quality as Vina. This important feature can be utilised in subsequent studies that will focus on large structure-based virtual screening.

On-going Research / Outlook

In future research, we will focus on analyses of more complex computational chemistry problems, such as molecular docking with flexible amino acid residues delineating the enzyme active site or molecular dynamics. It will also be necessary to evaluate the scalability of such calculations depending on provided computational resources.

Conclusion

Molecular docking is a crucial investigation involved in computer aided drug design (CADD), which enables to reveal probable binding of drug candidates in biological targets. CADD methodology generally promotes the rationality of the process, and helps to reduce the time and costs spent on the development of new drugs. As such, these simulations have profound impact on large fields of medicinal chemistry and, therefore, we have to take care about their theoretical quality. HPC is as a promising way for shifting the simulations to a higher level of accuracy, which is associated with growing computational demands. Such approaches that especially minimise the intervention of unnecessary software layers can bring about significant acceleration to this area. MPI is one such means deserving proper attention in computational chemistry.

We proved in the study that MPI enables better utilisation of the hardware in comparison with the computational scenario controlled by a scheduler. A speed-up of 2.5 was achieved if AutoDock Vina, along with a pleasingly parallelised job deployment, was substituted by VinaLC which implements MPI. The approach demonstrated herein may be easily applied into similar in silico experiments and bring desirable benefits to medicinal chemists dealing with a variety of biochemical and bioinformatics topics [2].

Publications

[1] Doležal, Rafael; Ramalho, Teodorico; Franca, Tanos; Kuca, Kamil, “Parallel Flexible Molecular Docking in Computational Chemistry on High Performance Computing Clusters” (Conference), Lecture Notes in Computer Science, 9330, 2015. Doi: 10.1007/978-3-319-24306-1_41

[2] Dolezal, Rafael; Melikova, Michaela; Mesicek, Jakub; Kuca, Kamil, “Rational Discovery of GSK3-Beta Modulators Aided by Protein Pocket Prediction and High-Throughput Molecular Docking” (Conference), Lecture Notes in Computer Science, 9876, 2016. Doi: 10.1007/978-3-319-45246-3_41

THE SQM/COSMO SCORING FUNCTION RELIABLY DESCRIBES NONCOVALENT INTERACTIONS IN PROTEIN-LIGAND COMPLEXES

Research institution:
Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences

Principal investigator:
Pavel Hobza

Researchers:
Jindřich Fanfrlík, Adam Pecina, Susanta Haldar, Jan Řezáč, Martin Lepšík, Robert Sedlák

Project partners:
Institut de Génétique et de Biologie Moléculaire et Cellulaire (Alberto D Podjarny), University of Pardubice (Aleš Růžička)

Project IDs:
OPEN-1-7, OPEN-2-13, OPEN-3-2, OPEN-4-13, OPEN-5-12, DD-16-17

Introduction

Current docking/scoring tools are fast but of varying reliability. Here, we develop a novel scoring function featuring a quantitative semi-empirical quantum mechanical (SQM) description of all types of noncovalent interactions coupled with implicit COSMO solvation. We show unequivocally that it outperforms widely used scoring functions. The accuracy and chemical generality of the SQM/COSMO scoring function make it a perfect tool for computer-aided drug design.[1]

Results and Methods

To increase the fidelity of docking/scoring in identifying the native protein-ligand poses in cognate docking, we have been developing the "SQM/COSMO" scoring function, which combines semi-empirical quantum mechanical description of noncovalent interactions at the PM6-D3H4X or SCC-DF-TB3-D3H4 levels and COSMO implicit model of solvation. This approach outperformed standard scoring functions, even with the challenging metalloprotein featuring $Zn^{2+}\cdots S^-$ interaction. The new approach holds promise for diverse protein-ligand complexes including metalloproteins.

Our approach to the development of the SQM/COSMO scoring function is systematic. Using accurate calculations in small model systems as a benchmark, we have developed corrections for SQM methods that provide reliable and accurate description of a wide range of noncovalent interactions including dispersion, hydrogen- and halogen-bonding.[2] Coupled with the PM6 SQM method, the resulting PM6-D3H4X approach is applicable to wide chemical space and does not require any system-specific parameterisation.[1] We have demonstrated the generality of the SQM/COSMO scoring function in various noncovalent protein-ligand complexes, such as human enzyme aldo-keto reductases[3] and cyclin-dependent kinases.[4]

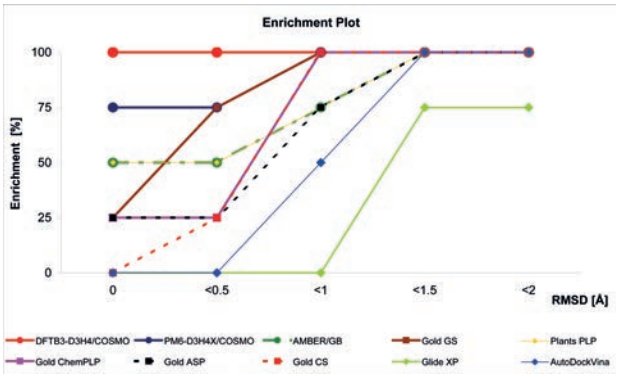


Figure 1.
Enrichment plot shows the % of cases in which the lowest energy ligand of a given SF has defined RMSD to the crystal pose. We can see that the SQM/COSMO scoring function at the DFTB3-D3H4 level makes a perfectly vertical curve. Figure adopted from Publication [1].

Using a known inhibitor scaffold of the potent aldose reductase inhibitor IDD388, we have designed a series of derivatives bearing the same halophenoxyacetic acid moiety and with an increasing number of bromine atoms on its aryl moiety. Next, by means of PM6-D3H4X based scoring function, IC_{50} measurements and X-ray crystallography, we have studied their Structure-Activity Relationship (SAR) against aldose reductase (AR) and aldo-keto reductase family member 1B10 (AKR1B10). We found inverse SAR: for AR, the introduction of Br substituents worsens the inhibition potency, while, for AKR1B10, their introduction improves it. The brominated inhibitor MK204

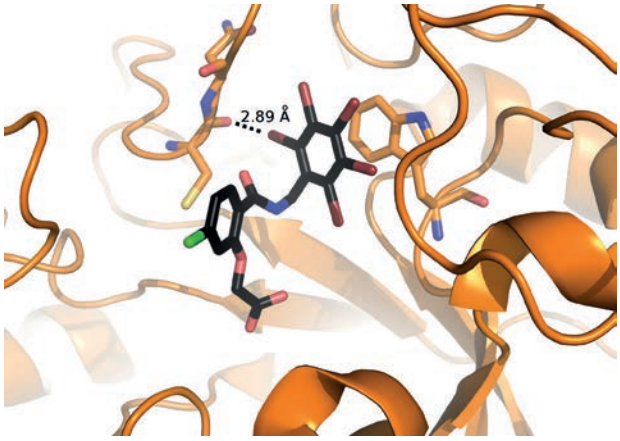


Figure 2.
Halogen bonding of MK204 to AKR1B10

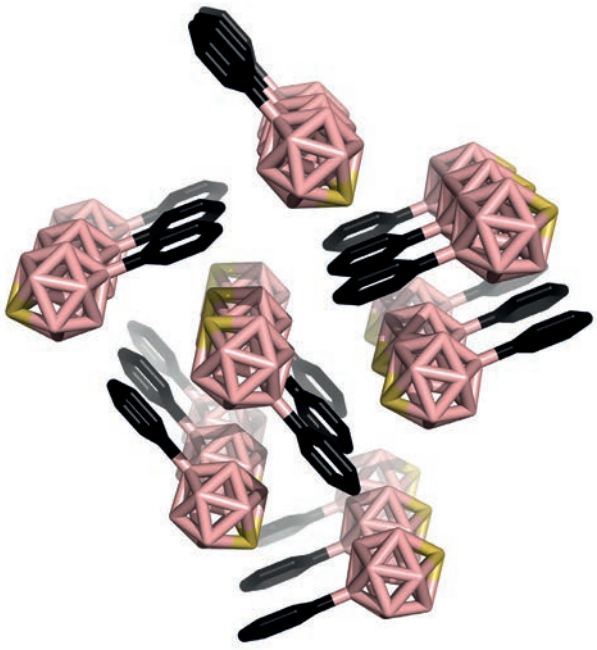


Figure 3.
The crystal structure of 12-Ph-closo-1-SB11H10. Coloured by element (pink: B, yellow: S, black: C). H atoms omitted for clarity.

was able to make a strong halogen bond (Br...O distance of 2.89 Å, C-Br...O angle of 169.3 degrees and calculated ΔE_{int} of the halogen bond of -4.4 kcal/mol, Figure 2) with the protein, while presenting the lowest desolvation cost among all the series, translated into the most selective IC_{50} for AKR1B10. Indeed, the finding of this novel AKR1B10 binding site conformer opens new perspectives into the structure-based drug design of selective AKR1B10 inhibitors targeting cancer.[3]

QM based methods can not only reliably describe noncovalent interactions of common organic compounds but also of unusual compounds like boron clusters. We performed systematic quantum mechanical studies of σ-hole (chalcogen, pnictogen, and halogen) bonding in neutral closo-heteroboranes. Chalcogens and pnictogens are incorporated in the borane cage, whereas halogens are considered as exo-substituents of dicarbaboranes. The chalcogen and pnictogen atoms in the heteroborane cages have highly positive σ-holes.[5] Consequently, these heteroboranes form very strong chalcogen and pnictogen bonds. Halogen atoms in dicarbaboranes also have a highly positive σ-hole, but only in the case of C-bonded halogen atoms.[6] In such cases, the halogen bond of heteroboranes is also strong and comparable to halogen bonds in organic compounds with several electron-withdrawing groups being close to the halogen atom involved in the halogen bond. We have already demonstrated the dominant role of chalcogen and halogen bonding in the crystal packing for 12-Ph-closo-1-$SB_{11}H_{10}$ and 1-Ph-2-Br-closo-1,2-$C_2B_{10}H_{10}$.[6,7]

On-going Research / Outlook

Based on our extensive experience with the development of QM methods, the SQM/COSMO scoring function has been accelerated. It only features two dominant terms to describe protein–ligand interactions and ligand and active site desolvation. The introduced SQM/COSMO scoring function has been compared to eight widely used scoring functions on four protein-ligand systems.[1] The SQM/COSMO scoring function exhibited superior performance. No ad hoc fitting against data sets has been involved in the development of our scoring function. Furthermore, it offers generality and comparability across the chemical space and no system-specific parameterisations have to be performed. The time requirements allow for calculations of thousands of docking poses, given the supercomputer power. We propose the SQM/COSMO filter as a tool for accurate medium-throughput refinement using docking/scoring or as a reference method for judging the performance of other scoring functions. Large-scale application of the method for diverse P-L complexes is now being set up.

Conclusion

We have been developing the SQM based scoring function at the PM6-D3H4X level that outperformed commonly used scoring functions in protein-ligand native pose recognition in cognate docking. We further show that the quality of the SQM/COSMO scoring function can be further improved by applying a more accurate DFTB3-D3H4 method combined with COSMO implicit solvation. The success of the SQM/COSMO scoring function is due to the high-quality description of both dominant terms and also probably due to fortuitous compensations of the neglected scoring terms. These very promising results are being further validated in our laboratory on a larger dataset of diverse protein-ligand complexes. Even though the DFTB3-D3H4 is more demanding than the PM6-D3H4X method, it is still orders of magnitude less time-consuming than the most economic DFT. Given the supercomputer power, the time requirements for the SQM/COSMO SF allow us to calculate thousands of docking poses in a reasonable time. We are showing here that the SQM/COSMO scoring function pushes the accuracy of scoring functions toward the experimental limits. Because of its generality, comparability across the chemical space and no need for any ad hoc system-specific parameters, we propose the SQM/COSMO scoring function be used as a computational tool for accurate medium-throughput compound refinement by docking/scoring or as a reference method to judge the performance of other scoring functions.

Publications

[1] Pecina, A.; Hadlar, S.; Fanfrlik, J.; et al. The SQM/COSMO Scoring Function at the DFTB3-D3H4 Level: Unique Identification of Native Protein-Ligand Poses. Journal of Chemical information and Modelling. 2017. Doi: 10.1021/acs.jcim.6b00513.

[2] Rezac, J.; Hobza. P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. Chemical Reviews, 116: 5038-5071, 2016. Doi: 10.1021/acs.chemrev.5b00526

[3] Cousido-Siah, A.; Ruiz, F. X.; Fanfrlik, J.; et al. IDD388 Polyhalogenated Derivatives as Probes for an Improved Structure-Based Selectivity of AKR1B10 Inhibitors. ACS Chemical Biology, 11: 2693-2705, 2016. Doi: 10.1021/acschembio.6b00382

[4] Hylsová, M.; Carbain, C.; Fanfrlík J.; et al. Quantum mechanical scoring with explicit solvent effects: Novel pyrazolo[1,5-a]pyrimidine CDK2 inhibitors point to dynamical effects. European Journal of Medicinal Chemistry, Doi: 10.1016/j.ejmech.2016.12.023.

[5] Pecina, A.; Lepsik, M.; Hnyk, D.; et al. Chalcogen and Pnictogen Bonds in Complexes of Neutral Icosahedral and Bicapped Square-Antiprismatic Heteroboranes. Journal of Physical Chemistry A, 119:1388-1395, 2015. Doi: 10.1021/jp511101n

[6] Fanfrlik, J.; Holub, J.; Ruzickova, Z.; et al. Competition between Halogen, Hydrogen and Dihydrogen Bonding in Brominated Carboranes. ChemPhysChem, 17:3373-3376, 2016. Doi: 10.1002/cphc.201600848

[7] Fanfrlik, J.; Prada, A.; Padelkova, Z.; et al. The Dominant Role of Chalcogen Bonding in the Crystal Packing of 2D/3D Aromatics. Angewandte Chemie-International Edition, 53, 10139-10142, 2014. Doi: 10.1002/anie.201405901

THEORETICAL STUDY OF THE STATE-SELECTED REACTIONS OF ELECTRONICALLY EXCITED OXYGEN IONS WITH METHANE RELEVANT FOR TITAN'S ATMOSPHERE

Research institution:
J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Principal investigator:
Jan Hrušák

Researchers:
Ivana Paidarová, René Kalus

Project partner:
IT4Innovations National Supercomputing Center

Project ID:
OPEN-5-20

Introduction

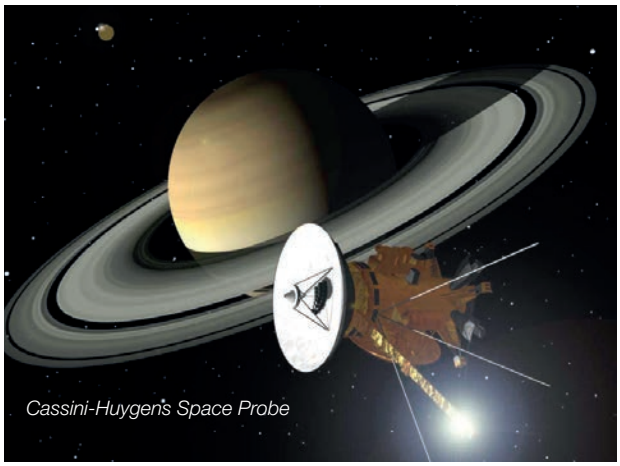
The motivation of the project is to understand chemistry in Titan's atmosphere based on one important milestone in the field of planetary exploration: the Cassini-Huygens mission (1997-2005) to Saturn's moon Titan. The presence of oxygen atoms on Titan [1] potentially provides the basis for pre-biological chemistry that could be helpful for explaining Earth's organic chemical evolution. Therefore, it appears very important to fully understand the reactions of O^+ with the two most abundant neutrals on Titan, namely N_2 and methane.

We have focused on the O^++CH_4 reactions. Contrary to the reaction of the methane with the ground state O^+, many more products arise when electronically excited reactants $O^+(^2D,^2P)$ are included. The experimental evidence of different reaction mechanisms [2] acting for individual O^+ electronic states cannot be explain without quantum chemistry computations, which enables discussing important product channels. The complexity of the task is seen from the energy diagram in Figure 1. Over 40 different reaction products have to be considered based solely on the simple thermodynamics consideration.

To encompass the energy window at least 19 states of A' symmetry have to be treated simultaneously. Therefore, we have to look for a computational strategy that allows through proper atomic manipulations to approach, in a chemically reasonable manner, the description of elementary processes of this complex system.

Methods and Results

A feasible approach, called "chemically reasonable intuitive reaction coordinate" (CRIRC), was proposed in [3]. It deals with the large dimensionality of the O/CH_4^+ system and indicates the positions of avoided crossings. Within the CRIRC concept only a limited set of reaction specific internal coordinates is needed to guide the reactants to products.



Cassini-Huygens Space Probe

This set can be chosen so that it enables the description of all the main features of the reaction paths for the whole variety of reactions at a still very moderate dimension of the energy surfaces. The multistate potential energy surfaces are divided in subspaces corresponding to different chemical reactions domains, and these are calculated separately in the corresponding subsets of internal coordinates. Through intelligent strategies, these fractal energy profiles can be coupled at the bifurcation points and the different chemical reaction channels can be followed.

The pilot scans along with the CRIRC by state-averaged multi-configuration self-consistent-field approach (SA-MCSCF) in conjunction with the augmented correlation consistent triple zeta basis set (aug-cc-pVTZ) were published in [3]. The coupled perturbed MCSCF (CP-MCSCF) gradients based on state averaged (SA) natural orbitals were used for geometry optimi-

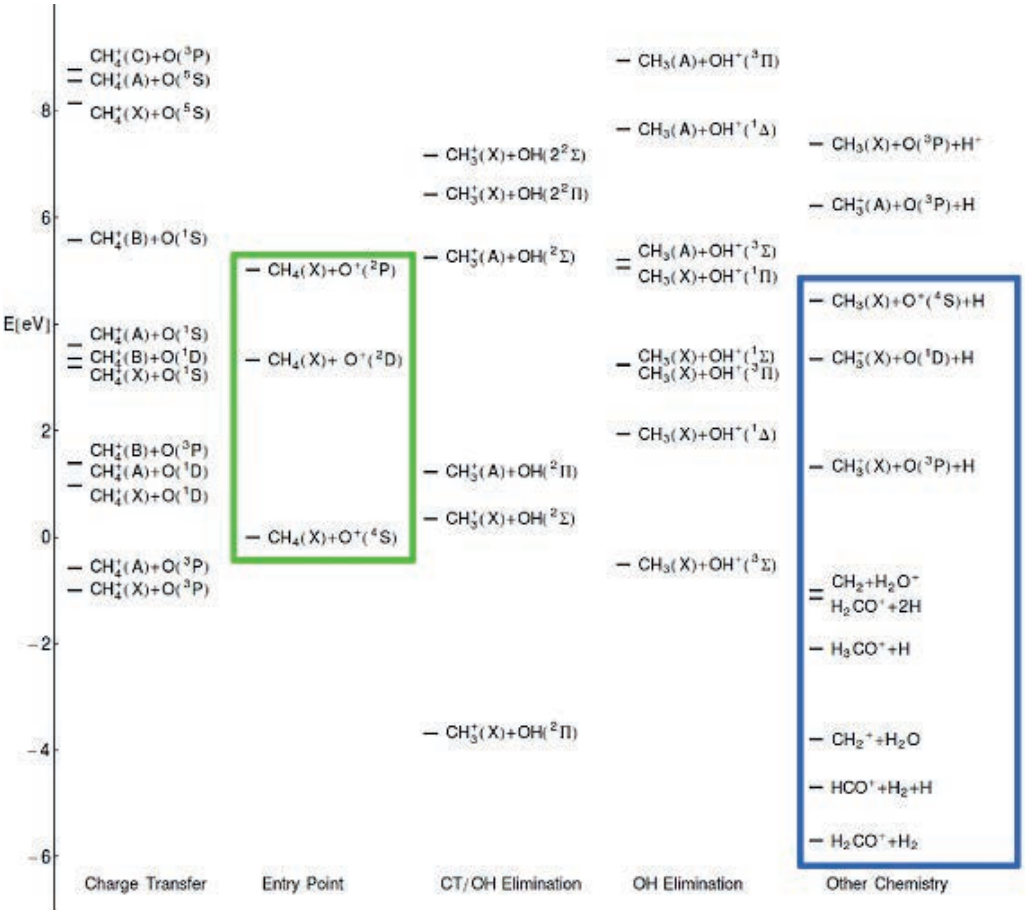


Figure 1.
Energy diagram of the thermodynamically accessible products of CH_4/O^+ reaction. The green frame corresponds to the reactants.

sation at each of the 56 points along the CRIRC. The obtained energy profiles revealed the necessity to refine the grid, namely at the avoided crossing regions (RCII and RCIII in Figure 3). On the other hand, they were sufficiently indicative to encourage further research.

In the present stage of the project we have focused on the long-range part of the surfaces characterised by the two coordinates relevant to the methane-oxygen ion distance and "impact parameter", OX and AX, respectively (see Figure 2). The quantum chemistry program package MOLPRO [5] and the methods described in [3] were used for the calculation of the multiple potential energy and electric dipole transition moments surfaces (see examples in Figures 4 and 5, respectively).

First qualitative investigation of the electric dipole transition moments between the states corresponding to the reactants $O(^2D,^2P) + CH_4$ and all the energetically lower states published in [1] aimed to answer the question important for the dynamic studies: Do the excited O^+ ions reach the "avoided crossings" region, or does the charge transfer prevailingly occur in the long-range part of the surfaces?

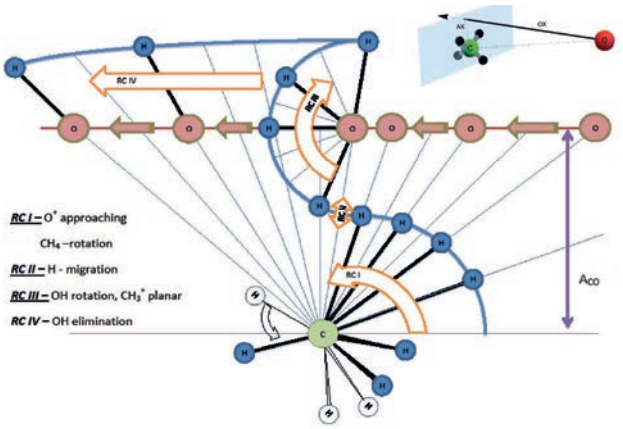


Figure 2.
Schematic diagram of the CRIRC

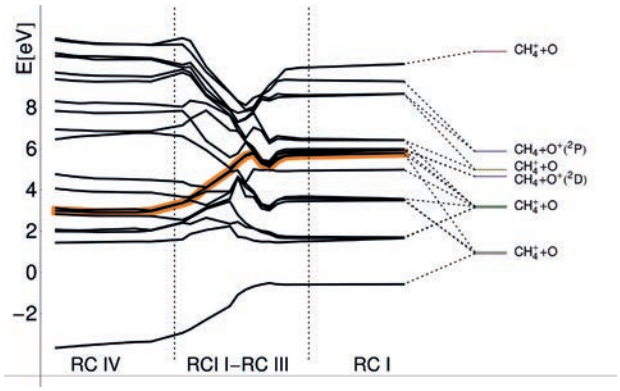


Figure 3.
Energies of 19 states along the CRIRC

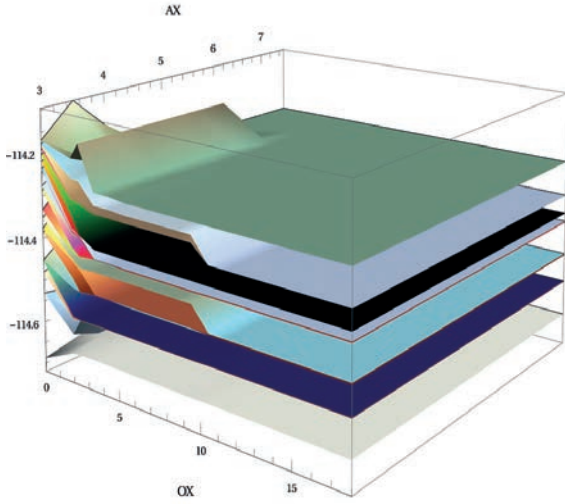


Figure 4.
Cut through the PES's corresponding to 19 A' states of CH_4/O^+ as functions of OX an AX coordinates. Calculations were performed using the SA-MCSCF method with energy optimisation at each step

Conclusion and Outlook

The concept of CRIRC [3] is rather straightforward and it shall be further developed. We intend to explore the possibility of other dimensionality reduction techniques for these purposes. In addition, until now we have only included the bound electronic states of OH, OH^+, CH_3, CH_3^+, and CH_4^+ at the product site and we did not consider excited states for the secondary reaction products CH_3O^+, CH_2O^+, HCO^+, and COH^+, etc. (blue rectangle in the 5th column in Figure 1). However, even though these species only represent a small fraction of the thermal reaction products, they are essential for the reactions leading to the creation of more complex molecules, the building elements of life, and will therefore represent the next steps on our way towards the theoretical description of the $[CH_4/O]^+$ chemistry. In spite of indicative results the complexity of the chemical reactions cannot be solved without dynamics on the multilevel surfaces. Being able to explore all important reaction channels we have to speed up the energy and energy gradients evolutions for extended dynamical calculations with the highly parallelised program of the Ostrava Group [5]. Therefore, in parallel with the surfaces calculations we tested the possibility of using the TT-DFT method.

The results obtained within the project OPEN-5-20 were published in [6] and presented as a conference contribution [7].

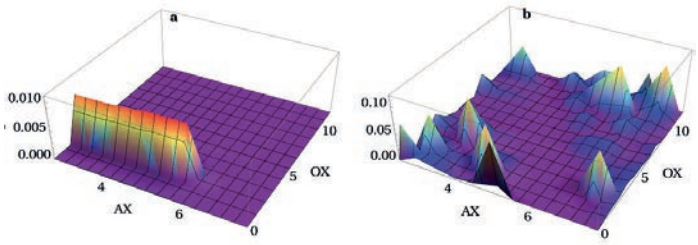


Figure 5.
Square value of the transition moments from the states corresponding to asymptotic limit $O^+(^2D)+CH_4$ and $O^+(^2P)+CH_4$ to all the lower states of CH_4/O^+.

References

[1] S.M. Hörst, V. Vuitton, R.V. Yelle. Origin of oxygen species in Titan's atmosphere. J. Geophys. Res, 113:E10006, 2008. Doi: 10.1029/2008JE003135
[2] B. Cunha de Miranda, C. Romanzin, S. Chefdeville, V. Vuitton, J. Žabka, M. Polášek, C. Alcaraz. Reactions of State-Selected Atomic oxygen ions $O^+(S^4, D^2, P^2)$ with methane. J. Phys. Chem. A, 119:6082-6098, 2015. Doi: 10.1021/jp512846v
[3] J. Hrušák and I. Paidarová. The way toward theoretical description of state-selected reactions of O^+ with methane. Int. J. Mass Spec, 354:372-377, 2013. Doi: 10.1016/j.ijms.2013.07.009
[4] Molpro Quantum Chemistry Software. Available at: http:/www.molpro.net
[5] Molecular Dynamics Group. URL: http:/moldyn.vsb.cz/index.php/software-menu/2-uncategorised/5-multidy

Publications

[6] Hrušák J. and I. Paidarová. Step towards modelling the atmosphere of Titan: state-selected reactions of O^+ with methane. Orig Life Evol Biosph. 2016, 46(4), 419-424. Doi 10.1007/s11084-016-9503-4 2016
[7] Paidarová I. and J. Hrušák. Transition moment surfaces for the CH_4^+ O^+. Conference contribution, COST TD1308, Liblice, 12-15.9. 2016

CONFORMATIONAL TRANSITIONS AND MEMBRANE BINDING OF THE NEURONAL CALCIUM SENSOR RECOVERIN

Research institution:
Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences

Principal investigator:
Pavel Jungwirth

Researchers:
Štěpán Timr, Roman Pleskot, Jan Kadlec

Project ID:
OPEN-7-43

Introduction

In living cells, a large number of processes are controlled by the intracellular concentration of calcium ions. The concentration of calcium is sensed by specialised sensor proteins that change their molecular shapes upon binding of one or more calcium ions, initiating a cascade of biomolecular events. Despite the tremendous physiological and medical importance of calcium signalling, molecular details of how calcium sensors work are not yet fully understood.

This project focuses on recoverin, a member of the Neuronal Calcium Sensor (NCS) family of proteins, which are expressed in neuronal cells and which, by serving as molecular switches dependent on intracellular calcium concentration, play an essential role in neuronal calcium signalling [1]. Recoverin, in particular, is found in photoreceptor cells of the vertebrate retina, where it allows regulation of light sensitivity and adaptation to low-light conditions [2]. The molecular structure of recoverin (Figure 1) contains a myristoyl lipid chain, sequestered in a hydrophobic cavity inside the N-terminal domain of recoverin at resting intracellular calcium concentrations. When the concentration of calcium is increased, two Ca^{2+} ions bind to recoverin, resulting in a structural rearrangement of the two domains of the protein. This structural rearrangement is believed to expose the hydrophobic myristoyl group, which is postulated to act as a membrane anchor, reversibly attaching the activated recoverin to retinal disc membranes, where this calcium sensor can interact with its target, the enzyme rhodopsin kinase.

Although the myristoyl group is assumed to be necessary for anchoring the activated calcium sensor to the membrane, other factors are likely to be important as well. Notably, experiments have shown that the association of myristoylated proteins with the membrane can be enhanced by the presence of negatively charged lipids in the membrane [3]. Moreover, specific interaction with negatively charged lipids might be required for the membrane-bound protein to attain a proper orientation compatible with its physiological role.

In order to investigate the fine-tuned interplay of the membrane anchoring via the myristoyl group and the recoverin-membrane electrostatic interactions, we perform molecular dynamics simulations. These simulations provide us with atomistic insight into the binding of the calcium-activated recoverin to the membrane and allow us to identify critical components necessary for its function as a reversible membrane-binding calcium sensor.

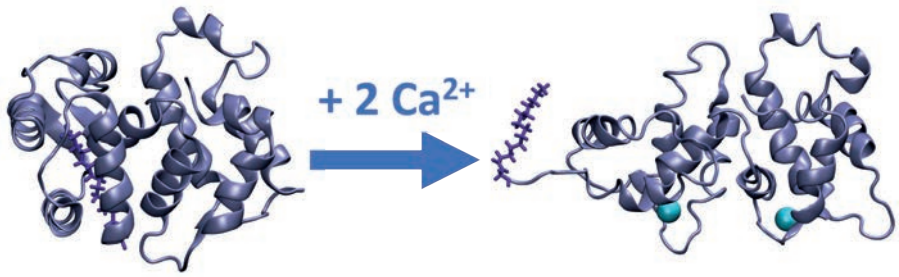


Figure 1.
Calcium-myristoyl switch of recoverin

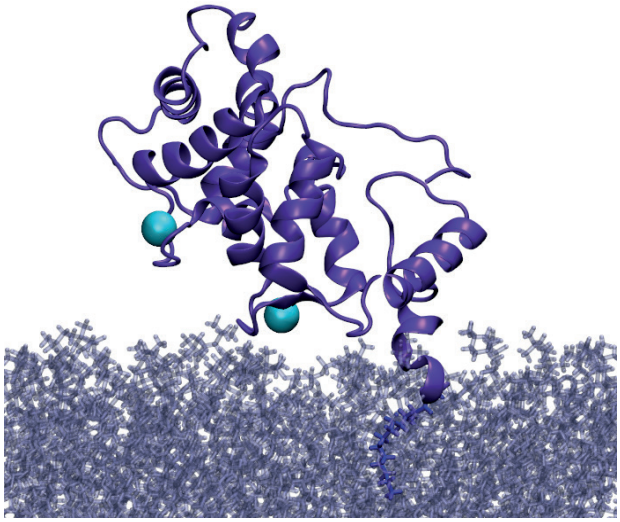


Figure 2.
Membrane-bound recoverin:
Snapshot from molecular dynamics

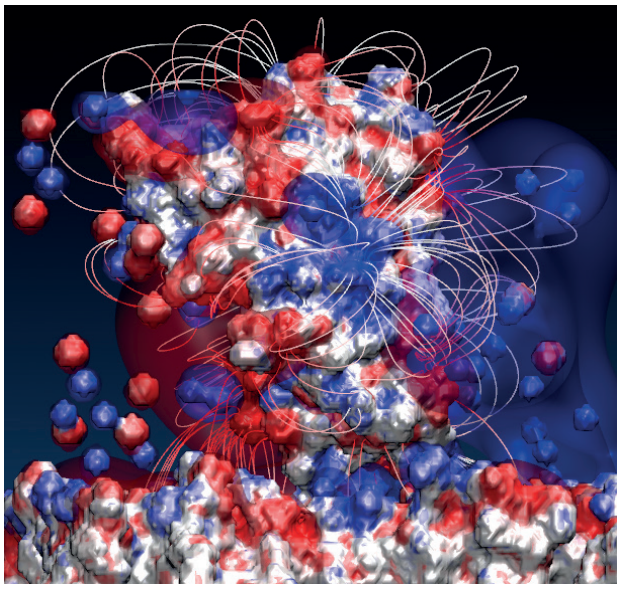


Figure 3.
Membrane binding of recoverin:
Depiction of electrostatic interactions
(calculated using APBS[6])

Results and Methods

This project involves extensive molecular dynamics simulations, requiring a large amount of computational resources. Through numerical solution of Newton's equations of motion, molecular dynamics simulations permit us to follow the motion of each individual atom of the simulated system, thereby allowing us to observe the process of membrane binding with an unprecedented level of detail.

To perform our simulations, we use the open-source GROMACS molecular dynamics package [4], which offers highly efficient parallelisation and favourable scalability, making it suitable for the IT4Innovations supercomputers. Moreover, we employ the PLUMED plugin [5] for trajectory analysis and for the calculation of free energy barriers using the methods of metadynamics and umbrella sampling. In addition, we make use of the APBS software [6] for the analysis of electrostatic interactions.

The simulated system consists of a single recoverin molecule and a lipid membrane patch, both immersed in a physiological salt solution (Figure 2). The total number of atoms in a simulation typically reaches 10^5. The system is replicated in three dimensions with the use of periodic boundary conditions. The simulation lengths span up to 2 μs so as to properly sample different modes of recoverin-membrane interaction.

We have performed simulations with different membrane lipid composition in order to examine the effect of lipid composition on the orientation of recoverin. We started with a plain neutral phosphatidylcholine (PC) bilayer and then continued by adding 20% of negatively charged phosphatidylglycerol (PG) to it. Finally, we performed simulations containing membranes exclusively composed of PG. We have repeated our simulations using two different sets of force field parameters (AMBER and CHARMM) with the aim of making our results robust with respect to the choice of the force field.

The results of our simulations give us information on the overall orientation of the protein as well as molecular details of its interaction with the membrane (Figure 3). Notably, in several simulations, we have been able to observe spontaneous membrane insertion of the myristoyl anchor, to our knowledge for the first time in an all-atom simulation of a myristoylated protein. By varying the membrane composition and the initial conditions in our simulations, we have managed to identify the determinants of a successful insertion of the myristoyl group into the bilayer. The orientation of the successfully bound recoverin at a membrane with 20% PG is consistent with the information from previous nuclear magnetic resonance (NMR) measurements [7], and our simulations prove that protein residues K5, K11, K37, R43, and K84 are involved in the interaction with the negatively charged membrane containing 20% PG, as postulated on the basis of the NMR experiments [7] (Figure 4).

On-going Research / Outlook

Currently, we are looking into the effect of the chemical nature of the negatively charged lipids on the membrane binding and orientation of recoverin. To this end, we are performing simulations of recoverin with a bilayer containing phosphatidylserine (PS) lipids as a replacement of PG.

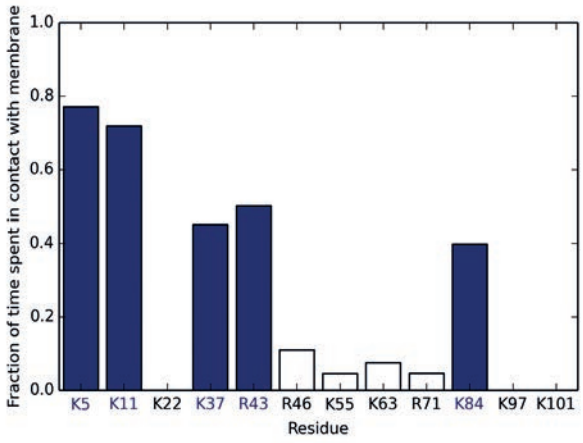


Figure 4.
Contacts of positively charged protein residues with the membrane: Results from a simulation containing 20% PG

At the same time, we are working on the computational characterisation of the conformational transitions induced by calcium binding (Figure 1), with the goal of achieving a full understanding of this remarkable calcium-myristoyl switch.

Conclusion

Thanks to the IT4Innovations infrastructure, we have been able to perform extensive molecular dynamics simulations of the neuronal calcium sensor recoverin. By giving us atomistic insight into the membrane binding of this calcium-sensing protein, these simulations have allowed us to characterise key factors affecting the physiological role of recoverin.

Despite its focus on a single member of the NCS protein family, this project provides more general insights into the function of other NCS sensors, thanks to their structural homology with recoverin. These calcium switches, ubiquitous in neurons, are involved in the regulation of many cellular processes and have been proposed as therapeutic targets for the treatment of several neuronal disorders [8].

References

[1] Burgoyne, R. D. Nature Reviews Neuroscience 8. 2007, 182-193. Doi:10.1038/nrn2093
[2] Xu, X.; Ishima, R.; Ames, J. B. Proteins-Structure Function and Bioinformatics. 2011, 79, 1910-1922. Doi: 10.1002/prot.23014
[3] Murray, D.; Ben-Tal, N.; Honig, B.; McLaughin, S. Structure. 1997, 5, 985-989. Doi: 10.1016/S0969-2126(97)00251-7
[4] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Journal of Chemical Theory and Computation. 2008, 4, 435-447. Doi: 10.1021/ct700301q
[5] Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. Computer Physics Communications. 2014, 185, 604-613. Doi: 10.1016/j.cpc.2013.09.018
[6] Baker, N. A.; Sept, D.; Joseph, S; Holst, M. J.; McCammon, J. A. Proc. Natl. Acad. Sci. USA, 2001, 98, 10037-10041. Doi: 10.1073/pnas.181342398
[7] Valentine, K. G.; Mesleh, M. F.; Opella, S. J.; Ikura, M.; Ames, J. B. Biochemistry. 2003, 42, 6333-6340. Doi: 10.1021/bi0206816
[8] Seaton, G.; Hogg, E. L.; Jo, J.; Whitcomb, D. J.; Cho, K. Seminars in Cell & Developmental Biology. 2011, 22, 530-535. Doi: 10.1016/j.semcdb.2011.07.014

MODELLING OF ELEMENTARY PROCESSES IN COLD RARE-GAS PLASMAS

Research institution:
IT4Innovations National Supercomputing Center, VŠB – Technical University Ostrava

Principal investigator:
René Kalus

Researchers:
Martin Stachoň, Aleš Vítek, Rajko Ćosić, Tomáš Luber

Project partner:
Université Toulouse III Paul Sabatier

Project IDs:
IT4I-2-4, IT4I-7-13, IT4I-9-4

Introduction

Low-temperature atmospheric-pressure plasma (LTAPP) has long been known as an efficient tool for biomedical applications. Among others, LTAPP can be used for surgical instruments sterilisation, chronic wound healing, or even malignant cells inactivation [1]. In particular, rare-gas LTAPPs have received considerable attention. A typical setup of a rare-gas LTAPP generator designed at a collaborating laboratory [2] is shown in Figure 1. Noteworthy, the LTAPP can easily be guided through flexible tubes [2] and is thus appropriate not only for surface use, but also for endoscopic treatment.

Microscopic modelling of LTAPPs is of particular importance for the optimisation of LTAPP generators for biomedical applications. Not only are necessary input data provided in this way for subsequent macroscopic modelling, but also in that a deep understanding of the processes behind the healing properties of LTAPPs is achieved. In particular the latter opens new ways for innovations. It was the basic objective of a series of computational projects solved in the IT4Innovations National Supercomputing Center to carry out such microscopic investigations on rare-gas LTAPPs as a first step towards their realistic modelling performed in close collaboration with specialists from the project partner laboratories. In particular, collisions of molecular ions (Rg_2^+) with carrier gas (Rg) have been investigated as part of a broader project of modelling of processes inside the carrier gas prior to its entering into contact with the environment (e.g. in air). While collisions of primary, monoatomic ions (Rg^+) do not require too many extensive calculations, collisions of diatomic ions, which emerge from recombination processes inside the carrier gas and survive for long periods due to their strong bonding, are much more demanding and their modelling requires considerable computer power.

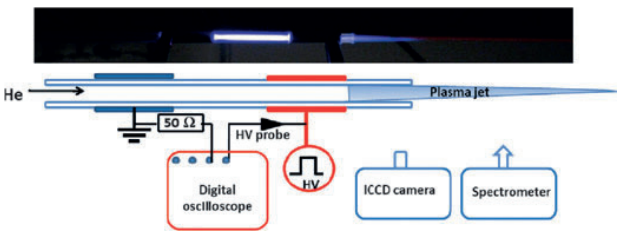


Figure 1.
Snapshot of a typical helium dielectric barrier discharge LTAPP generator (up) and its schematic sketch (bottom). Taken from [2] with kind permission of the authors.

Methods

Modelling of collision processes in LTAPP in general comprises the following steps: 1) effective cross-sections have to be calculated for all the relevant collision channels, 2) transport coefficients are then obtained from physical kinetics methods to be further used in 3) macroscopic modelling normally consisting in the numerical treatment of coupled Navier-Stokes diffusion, and Poisson equations for space charge density evolution. In the presented projects, steps (1) and (2) have been considered. First, the total collision cross-sections for non-reactive scattering (NRS),

$$Rg_2^+ + Rg \rightarrow Rg_2^+ + Rg, \qquad (1)$$

have been calculated for a broad range of collision energies within the momentum-transfer approximation [3] together with the total collision cross-section for collision induced dissociation (CID),

$$Rg_2^+ + Rg \rightarrow Rg^+ + 2Rg, \qquad (2)$$

Particular attention has been paid to non-adiabatic processes (electronic transitions) and inelastic effects in process (1) (rovibronic transitions). Second, calculated cross-sections have been used in transport properties calculations performed at the

collaborating institution using an optimised Monte Carlo code [4]. Calculations on helium, neon, and argon have been performed up to now [8-10].

The cross-sections have been calculated using a standard classical trajectory approach with heavy atomic nuclei treated classically and light electrons quantum mechanically. Butches of trajectories have been integrated numerically for a sufficiently broad range of the collision energy, E_{coll}, and impact-parameter, b. 504-512 trajectories have been integrated for a particular value of b, totally 100-200 different values of b have been considered for each collision energy, E_{coll} = 0.01 – 100 eV. Then, cross-sections have been calculated in a standard way from recorded deflection angles and/or dissociation probabilities [8-10]. A hemiquantal mean-field approach with the inclusion of quantum decoherence as implemented in [5] has been used. Classical Hamiltonian equations of motion have been solved for atomic nuclei,

$$\dot{q}_\alpha = \frac{p_\alpha}{m}, \quad \dot{p}_\alpha = \left\langle \varphi \left| -\frac{\partial \hat{\mathrm{H}}}{\partial q_\alpha} \right| \varphi \right\rangle, \qquad (3)$$

and quantum Schrödinger equation has been applied to electrons,

$$i\hbar \frac{\partial \varphi}{\partial t} = \hat{\mathrm{H}}\varphi. \qquad (4)$$

The electronic Hamiltonian, $\hat{H}$, has been calculated on the fly via semiempirical diatomics-in-molecules models [8-10]. In eqs. (3) and (4), p_α and q_α denote generalised atomic linear momenta and coordinates of nuclei, respectively, m stands for the nuclear mass, φ is current electronic wave function, dot represents the time derivative, and brackets are used to indicate integration over electronic positions. Importantly, the leading relativistic effect, the spin-orbit coupling, has also been considered [6] for heavier rare gases (neon and argon).

Results

Due to the limited extent of the present contribution, only selected examples of typical results verifying the computational methodology used are briefly mentioned. Readers are referred to references [8-10] for a more detailed discussion.

First, non-reactive scattering (eq. 1), and collision-induced dissociation (eq. 2) cross-sections are depicted in Figure 2 and compared with pseudo-experimental data obtained from measured mobilities of respective diatomic ions via an inverse method approach [8-10].

Second, diatomic ions mobilities in their carrier gas calculated over a broad range of driving electric field are depicted in Figure 3 and compared with available experimental data.

Noteworthy, a very good agreement has been achieved between the experimental and theoretical data. In addition, experimental measurements have been extended to the strong electric field region where measurements are either not reliable or impossible. Note also that the nuclear quantum effects, included here via an empirical correction [8,10], are important for light helium and to some extent also for neon, but can safely be neglected for heavier rare gases (like, e.g., argon).

On-going Research / Outlook

As mentioned in the Introduction section, the calculations presented in this contribution represent part of a broader research project aiming at realistic modelling of collision processes inside rare-gas LTAPPs as well as collisions of the rare-gas ions (both atomic and molecular) with environment molecules after the plasma jet leaves the guiding tube. Presently, calculations similar to those presented here are run for krypton and are being prepared for xenon to complete the former studies on intra-plasma collisions. In addition, preliminary studies have been launched to advance to the more challenging task of modelling collisions of rare-gas ions with the environment. Collisions of atomic ions with nitrogen and oxygen molecules are considered as a first step.

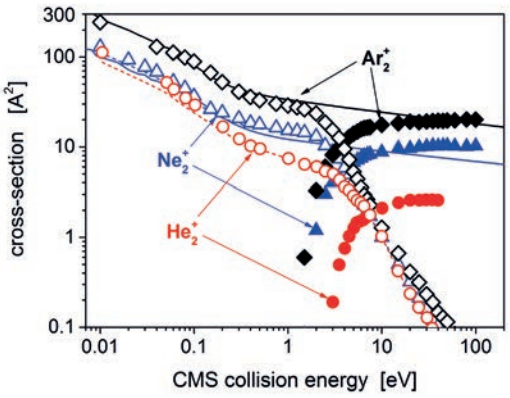


Figure 2.
Calculated momentum-transfer cross-sections of NRS (open symbols) and total cross-sections of CID (full symbols) compared with inverse-method data (full curves). Dashed curves shown for helium and neon represent corrected data with nuclear quantum effects included (see [8,10]).

Conclusion

Hemiquantal dynamics calculations have been performed for non-reactive and dissociative collisions of helium, neon and argon dimer cations with their respective carrier gas as part of a broader project aiming at a realistic modelling of collision processes in low-temperature and atmospheric pressure rare-gas plasmas. Collision cross-sections and transport properties of the ions have been calculated and compared with available experimental data. The calculations clearly show potential for further extensions to either heavier rare gases (krypton and xenon) or even to more general studies on the collision dynamics of primary rare-gas ions in the environment, e.g., after plasma dilutes in air.

References

[1] Laroussi, M. and T. Akan. Arc-free atmospheric pressure cold plasma jets: a review, Plasma Process. Polym. 4 (2007) 777. Doi 10.1002/ppap.200700066
[2] Yousfi, M., Merbahi, N., Pathak, A. and O. Eichwald. Low-temperature plasmas at atmospheric pressure: toward new pharmaceutical treatments in medicine. Fund. Clin. Pharm. 27 (2013) 1. Doi: 10.1111/fcp.12018
[3] Child, M. Molecular Collision Theory. Dover Publications, Mineloa, NY, 1996. ISBN 0486694372
[4] Yousfi, M., Hennad, A. and O. Eichwald. Improved Monte Carlo method for ion-transport in ion-molecule asymmetric systems at high electric fields. J. Appl. Phys. 84 (1998) 107. Doi: 10.1063/1.368007
[5] Janeček, I., Hrivňák, D. and R. Kalus. Theoretical modelling of postionisation fragmentation of rare-gas trimer cations. J. Chem. Phys. 125 (2006) 104315. Doi: 10.1063/1.2337632
[6] Janeček, I., Cintavá, S., Hrivňák, D., Fárník, M., Gadéa, F. X. and R. Kalus. Postionisation fragmentation of rare-gas trimers revisited with new theoretical approaches. J. Chem. Phys. 131 (2009) 114306. Doi: 10.1063/1.3224855
[7] Amarouche, J., Durand, G. and J. P. Malrieu. Structure and stability of Xe_n^+ clusters. J. Chem. Phys. 88 (1988) 1010. ISSN 0021-9606
[8] Ellis, H.W., Pai, R. Y., McDaniel, E.W., Mason, E.A. and L.A. Viehland. Transport properties of gaseous ions over a wide energy range. At. Data Nucl. Data Tables 17 (1976) 177. Doi: 10.1016/0092-640X(76)90001-2

Publications

[9] Chicheportiche, A. and M. Benhenni, M. Yousfi, B. Lepetit, R. Kalus, F. X. Gadéa, Ion collision cross sections and transport coefficients extended to intermediate energies and reduced electric fields for He_2^+ ions colliding with He, Phys. Rev. E 88 (2013) 043104. Doi: 10.1103/PhysRevE.88.043104
[10] Chicheportiche, A. and M. Stachoň, M. Benhenni, F. X. Gadéa, R. Kalus, M. Yousfi, First principles transport coefficients and reaction rates of Ar_2^+ ions in argon for cold plasma jet modelling. J. Chem. Phys. 141 (2014) 134302. Doi: 10.1063/1.4896613
[11] Benhenni, M. and M. Stachoň, F. X. Gadéa, M. Yousfi, R. Kalus. Transport and dissociation of neon dimer cations in neon gas: a joint dynamical and Monte Carlo modelling. J. Phys. B: At. Mol. Opt. Phys. 2016, 49(17), 175205. Doi: 10.1088/0953-4075/49/17/175205

Used software: http://moldyn.vsb.cz/index.php/software-menu/5-multidyn

ENZYMATIC REACTION MECHANISMS INVESTIGATED BY CAR-PARRINELLO AB INITIO DYNAMICS

Research institution:
Masaryk University

Principal investigator:
Jaroslav Koča

Researchers:
Stanislav Kozmon,
Manju Kumari,
Petr Kulhánek,
Jakub Štěpán,
Igor Tvaroška

Project ID:
OPEN-1-1

Introduction

A vast repertoire of glycans and glycoconjugates is present inside the cell. These glycoconjugates are formed by the sequential action of glycosyltransferases, augmenting the structural and functional diversity of proteins. O-GlcNAc transferase (OGT) is an important enzyme in the family of O-glycosyltransferases, catalysing the transfer of N-acetylglucosamine from UDP-N-acetylglucosamine (UDP-GlcNAc) to the hydroxyl group of Ser/Thr of various proteins. Aberrant glycosylation by OGT has also been linked to diabetic complications, cancer, and neurodegenerative diseases including Alzheimer's disease.[1] Understanding the reaction mechanism at the atomic level and determining the transition state can provide a rational basis for designing inhibitors based on transition state analogues. These compounds could then be used as potential drugs for the treatment of diseases originating in impaired functioning of OGT.[2]

Three different reaction mechanisms of OGT activity have been proposed in the literature (shown in Figure 1) differing in the proposed catalytic base. The first mechanism (Figure 1A) involved His498 as the catalytic base and was investigated in our previous study.[3] The second mechanism (Figure 1B) proposes that the α-phosphate moiety of the UDP could serve as the catalytic base. The third proposed mechanism called "proton shunting mechanism" involves a water molecule as the catalytic base and the proton is transferred via another water molecule to the Asp554 residue.

Studies of the reaction mechanism can be carried out using a variety of computational methods, differing mainly in whether they sample the potential energy surface (PES) or the free energy surface (FES). In either case, the evolution of the system from the reactant state to the product state is described by one or more parameters called reaction coordinates. The difference between PES- and FES-based methods then lies in the way other degrees of freedom (not directly involved in the reaction coordinate) are handled. In the case of PES-based methods, all other degrees of freedom are numerically optimised to find the configuration with minimum potential energy for a given value of the reaction coordinate. In contrast to that, FES-based methods allow all the other degrees of freedom to freely evolve in time, usually by a molecular dynamics simulation. The main advantage of PES-based methods is the relatively low computational cost of the geometry optimisation procedures compared to the simulations needed to calculate free energies. However, the success of these methods is highly dependent on the quality of the starting geometries, as simple geometry optimisations

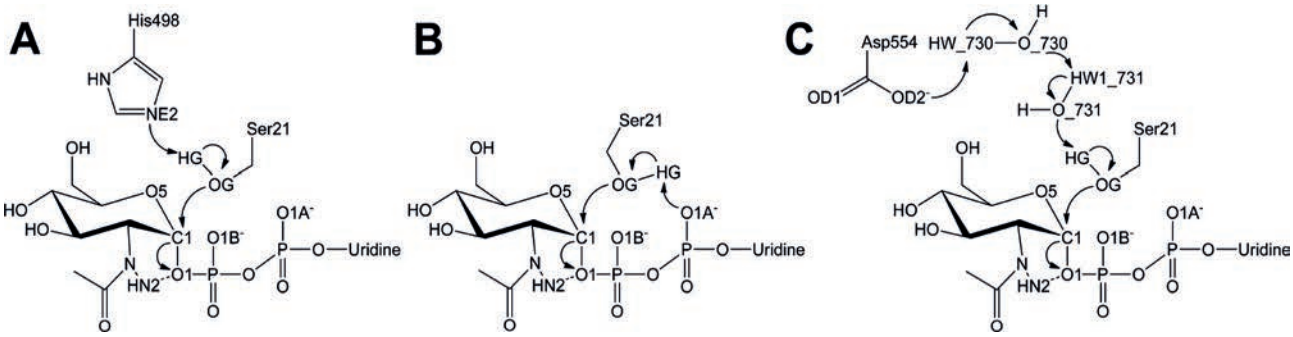


Figure 1.
Schematic representation of the proposed mechanisms:
(A) His498 (M_{His}) as a catalytic base; (B) α-phosphate (M_{PO4}) as a catalytic base; (C) the proton from Ser21 indirectly moves to Asp554 via two bridging water molecules (M_{Asp}).

are usually unable to drive the larger conformational changes needed to correct sub-optimal starting configurations. On the other hand, free energy calculations are usually more robust due to the lower dependence of MD simulations on the exact starting geometry. Additionally, only the FES-based approaches take the potentially significant effects of thermal motion (entropy) into account.

Unfortunately, the sheer computational cost of an ab initio MD simulation of an enzymatic system usually restricts reaction mechanism studies to PES-based approaches or highly approximate semi-empirical quantum methods. Only thanks to the computational power offered by the IT4Innovations National Supercomputing Center (over 1M CPU-hours), we have been able to successfully carry out a state-of-the-art simulation and calculate high-quality free energy profiles.

Results and Methods

To investigate the catalytic reaction of OGT and identify the most probable mechanism we utilised hybrid QM/MM Car-Parrinello molecular dynamics together with the string method. Using the CPMD Ab Initio method together with Kohn-Sham Density Functional Theory provides a more realistic description of reactions than widely used semi-empirical methods. The QM/MM approach is necessary to limit the computationally expensive quantum-chemical treatment only to the core of the active site, while the rest of the system is described by a simple empirical MM force field. Even with this approximation in place, the inclusion of entropic effects is still rather limited in practice due to inadequate sampling (on the picosecond time scale) caused by the high computational cost.

The string method (STM) [4] gives the minimum-free-energy path (MFEP) from reactants to products expressed a set of collective reaction coordinates. STM provides the free energy profiles along the found reaction pathways and allows us to identify the most probable mechanisms based on the height of the activation barrier. Compared with other free energy methods, the main advantages of STM are the independence of computational cost on the number of collective variables and the simplicity of running STM in parallel. STM discretises the reaction path into several dozen “beads” and then optimises the path by alternating two phases. First, forces acting on the individual beads are sampled by MD and the positions of all beads are then updated. Only the computationally inexpensive path update needs to communicate forces and positions across all beads, while the sampling stage is communication-free and thus trivially parallel. This makes STM a perfect candidate for large-scale parallel calculations.

In our OGT study, we have used two sizes of the QM region, namely 103 atoms for the M_{PO4} mechanism and 146 atoms for the M_{Asp} mechanism. Together with the MM region containing the rest of the protein and surrounding water molecules, the simulation box comprised over 90,000 atoms. The QM system was described using an inexpensive yet accurate PBE density functional.

We have used the implementation of String Method provided in PMFlib. The reaction paths were discretised into 21 beads

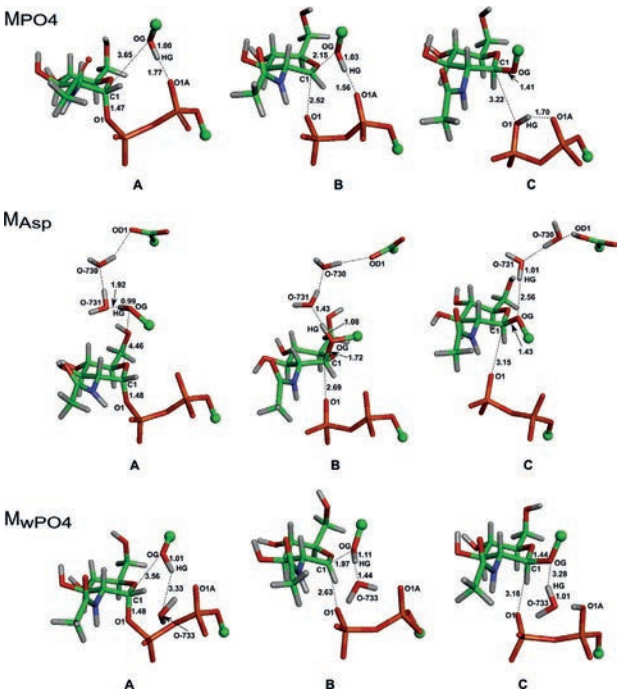


Figure 2.
Selected snapshots from the STM trajectory along the optimized reaction pathways representing (A) Michaelis (reactant) complex, (B) transition state and (C) product complex. Green spheres indicate methyl group of Ser21 and UDG present in M_{PO4}, M_{Asp}, and M_{wPO4}, and Asp554 in M_{Asp}. The average CV distances shown are in Å.

for the M_{PO4} and a newly discovered M_{wPO4} mechanism and 41 beads for the more complicated M_{Asp} mechanism. For all the reaction mechanisms, approximately 40 iterations of STM were performed before the path converged. The total simulation time amounted to ~120 and ~230 ps for the small and larger QM region, respectively.

Although we started by investigating two proposed mechanisms, we encountered a new mechanism that evolved during the preparation of M_{Asp}, and it is designated as M_{wPO4} (water-assisted phosphate mechanism). This mechanism is similar to M_{PO4}, differing only in the fact that the proton from the acceptor hydroxyl group is transferred to the leaving phosphate group through a water molecule.

Analysing the converged reaction paths, we can conclude that in all three mechanisms under investigation, the O-GlcNAc transfer, and glycosidic bond formation is a dissociative SN2 type reaction, through an oxocarbenium species. This ion is the centre for the nucleophilic attack through the protonated Ser21. All the mechanisms start the nucleophilic attack, which leads to the dissociation of the donor glycosidic bond, proton transfer, and immediate formation of the new glycosidic bond. The structures of important reaction events are shown in Figure 2.

Comparing the final free energy profiles for the three mechanisms (Figure 3) shows that the M_{PO4} mechanism is the most energetically favourable by a significant margin. The overall barrier for this mechanism is 23 kcal/mol, within the range of known experimental values (20-25 kcal/mol). In contrast to that, the large > 40 kcal/mol barriers for the remaining two mechanisms essentially rule out any reaction through these paths.

On-going Research / Outlook

Following this OGT study, we have successfully applied the methodology developed and tested here to further studies of reaction mechanisms of other important glycosyltransferases. A follow-up study on OGT using modified substrates is also currently underway.

Conclusion

In this study, we investigated the two recently proposed catalytic mechanisms for OGT together with a new mechanism found during this study. Using the String Method together with QM/MM Car-Parrinello molecular dynamics we were able to optimise the reaction path for all three mechanisms and calculate reliable free energy profiles. All three reaction mechanisms share many common features and differ almost exclusively in the fate of the proton from the acceptor hydroxyl group. The calculated free energy barriers show that the M_{PO4} mechanism is the most probable one with a barrier of 23 kcal/mol. The new insight into the catalytic reaction of OGT provided by our simulations may be used in the rational drug design of transition state analogue inhibitors with potential therapeutic uses.

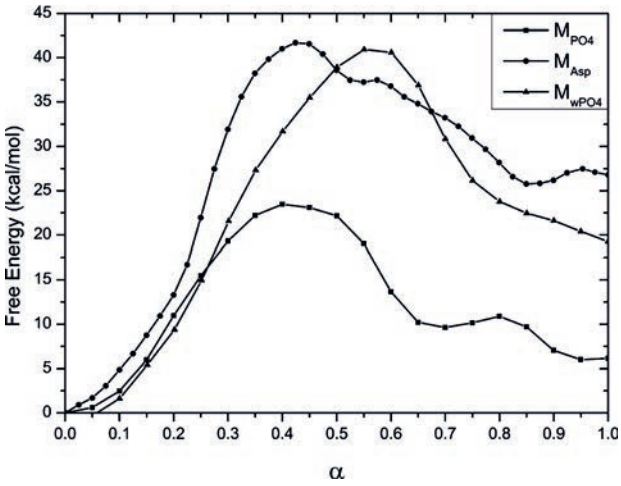


Figure 3.
Comparison of free-energy profiles of three optimised reaction pathways. The reaction coordinate α changes from the Michaelis complex (α = 0) to the product complex (α = 1).

References

[1] Hart, Gerald W.; Slawson, Chad; Ramirez-Correa, Genaro; Lagerlof, Olof. Cross talk between O-GlcNAcylation and phosphorylation: roles in signalling, transcription, and chronic disease. Annu. Rev. Biochem. 2011, 80, 825-858. Doi: 10.1146/annurev-biochem-060608-102511

[2] Compain, Philippe; Martin, Olivier R. Carbohydrate mimetics-based glycosyltransferase inhibitors. Bioorg. Med. Chem. 2001, 9, 3077-3092. Doi: 10.1016/S0968-0896(01)00176-6

[3] Tvaroška, Igor; Kozmon, Stanislav; Wimmerová, Michaela; Koča, Jaroslav. Substrate-Assisted Catalytic Mechanism of O-GlcNAc Transferase Discovered by Quantum Mechanics / Molecular Mechanics Investigation. J. Am. Chem. Soc. 2012, 134, 15563-15571. Doi: 10.1021/ja307040m

[4] Maragliano, Luca; Fischer, Alexander; Vanden-Eijnden, Eric; Ciccotti, Giovanni. String method in collective variables: Minimum free energy paths and isocommittor surface. J. Chem. Phys. 2006, 125, 024106. Doi: 10.1063/1.2212942

Publication

[1] Kumari, Manju; Kozmon, Stanislav; Kulhánek, Petr; Štěpán, Jakub; Tvaroška, Igor; Koča, Jaroslav. Exploring Reaction Pathways for O-GlcNAc Transferase Catalysis. A String Method Study. J. Phys. Chem. B. 2015, 119 (12), 4371-4381. Doi: 10.1021/jp511235f

DETECTING MATERIALS IN COAL-POLYURETHANE GEOCOMPOSITE CT-SCANS

Research institution:
IT4Innovations National Supercomputing Center

Researcher:
Marek Pecha, Martin Čermák, David Horák

Project partner:
Institute of Geonics of the Czech Academy of Sciences

Project ID:
IT4I-9-14

Introduction

In many scientific fields, image partition/segmentation methods are often used as image pre-processing techniques in complex applications, which simulate real-world processes. We can say that image partition/segmentation is methodology and proposed approaches used for the design of algorithms, by which the program/classifier determines the Regions of Interest in an image scene. In general, the Region of Interest is an important part of the image scene and is used for further processing, such as labelling, classification, mesh building, etc.

In the case of geological simulations, such as a plasticity computation on a heterogeneous solid, it is suitable to convert the real-world solid into a finite element model for authenticity of simulations; image partition/segmentation plays an important role in determination of material regions of a CT-scan image scene and setting up their binary masks. Further, the determined binary masks are used for building hull-mesh by the marching cube algorithm. For a tetrahedral volume mesh geometry generation from hull-mesh boundary representation, the NETGEN library [1] is typically used.

Further, plasticity numerical computations will be proceeding in PERMON – an abbreviation of the Parallel, Efficient, Robust, Modular, Object-oriented, Numerical software toolbox developed at the IT4Innovations National Supercomputing Center. PERMON is a set of solvers combining quadratic programming and domain decomposition methods. It makes use of and extends the PETSc framework for numerical computations. Among the main applications are contact problems of mechanics. They can be decomposed by means of the TFETI (Total Finite Element Tearing and Interconnecting) non-overlapping domain decomposition method implemented in the PermonFLLOP package. The resulting quadratic programming problems can be solved by the PermonQP package efficiently.

In this paper, we present the first stage of mesh pre-processing development, i.e. determination of the desired material binary masks.

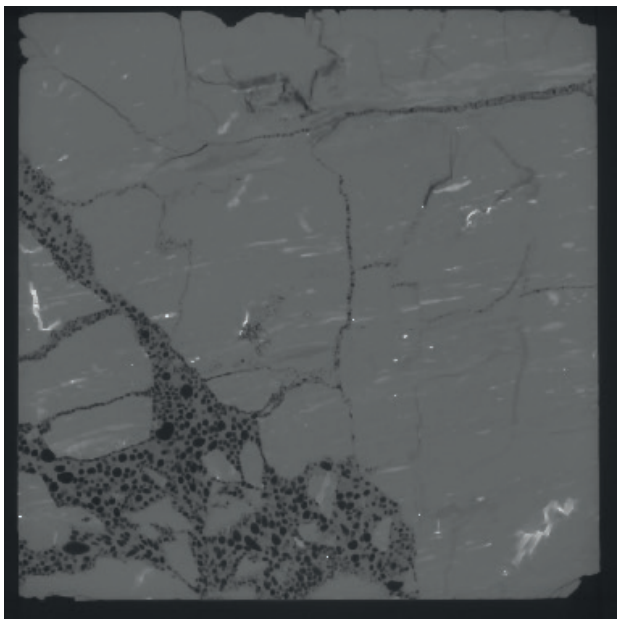


Figure 1.
Transverse CT-scan of coal-polyurethane geocomposite

Results and Methods

CT-scan material region determination

The geocomposite transverse CT-scans are typically 8-bit greyscale images, see Figure 1; it implies that appropriate intensities of desired materials are decoded into at most 256 values. In many cases the material areas are not compact in general; therefore, it is suitable to only proceed with pixel value distribution without the need to position them in a discrete image area for algorithm-design.

Thus, we present our approach, which we call the Simple Method of Reference Materials. The Simple Method of Reference

Material	Value (in greyscale)	Colour
air	43	
coal	111	
polyurethane	93	

Table 1.
Greyscale values of initial centroids for material regions

Materials (SMRM) design-fashion directly follows Steinhaus' work [2]. In general, SMRM is the semi-automatised version of the standard Lloyd algorithm [3]. Starting centroids are manually selected by a human segmentator/classifier, then the algorithm continues with the standard Lloyd iteration, i.e. it minimises the sum of squares within regions with arithmetic mean as the least-squares estimator.

For approach testing purposes, a program called PermonGeoDecompositer is being developed. It is written in C++ language and based on OpenCV framework [4]. Users can run the program on HPC systems and Linux, Windows or OS X desktops.

In our benchmark, we test our proposed approach on one transverse CT-scan of coal-polyurethane geocomposite, see Figure 1. The presented results and our approach was consulted with Dr Kolcún from the Institute of Geonics of the Czech Academy of Sciences, Ostrava, Czech Republic.

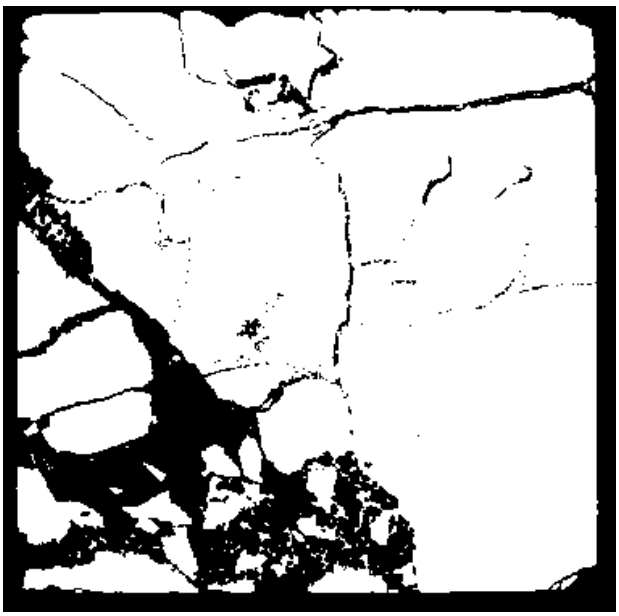


Figure 3.
Desired coal region binary mask

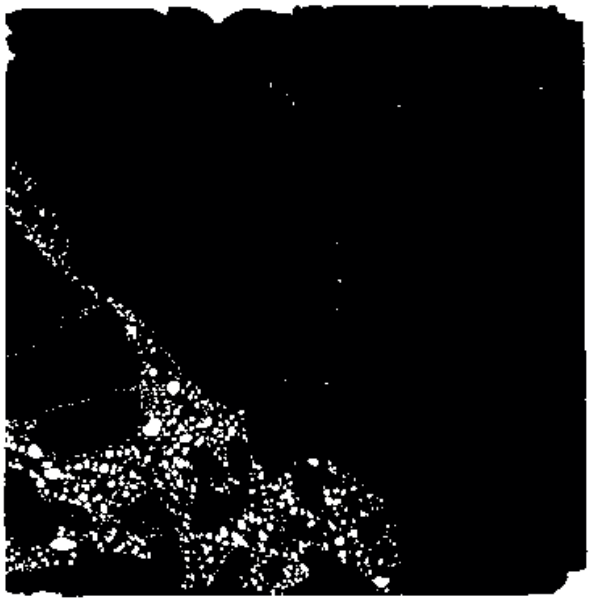


Figure 2.
Desired air region binary mask

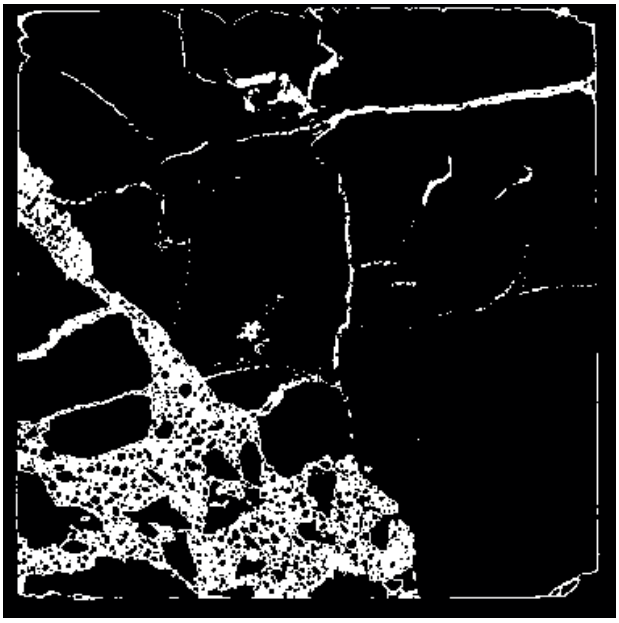


Figure 4.
Desired polyurethane region binary mask

We are now focused on the benchmark testing. At first, the human segmentator/classifier selects proper reference material values so that the program, which is mentioned above, can thereafter automatically run the k-means algorithm with the selected initial centroids when all proper reference materials values are determined by the human user.
In the following table, i.e. Table 1, manually selected greyscale values of initial centroids for each material region, i.e. an air region, a coal region, and polyurethane region, are presented.
With the initial conditions for the SMRM algorithm mentioned above, the SMRM converged to solutions which are depicted in following figures, i.e. the desired air region binary mask is depicted in Figure 2.
The desired coal region binary mask is depicted in Figure 3.
And finally the desired polyurethane region binary mask is depicted in Figure 4.

Conclusion

In our presented benchmark, we use OpenCV parallel implementation of the k-means algorithm, which is based on OpenCL technology. Inasmuch as the used implementation of the k-means algorithm for our initial conditions converges very quickly, an average computation takes 7.64 s on MacBook Pro (late 2012) in CPU-mode and 0.82 s on the Salomon Intel Xeon E5-2680v3 CPU, we therefore propose hybrid-parallel implementation for an image large dataset as follows: Run one mpi-process on each allocated processor. The MPI master process divides the input data-set among slave-processes. Each mpi-process runs OpenCV k-means on each image from the appropriate input subset. The initial conditions for the SMRM algorithm is entered by users.
However, such an approach is inefficient and user time-consuming. Therefore we are developing a new automatised approach, which we called Method of the Reference Material, in which appropriate initial centroids are selected deterministically from CT-scan value distributions.

On-going Research / Outlook

For now, we continue with extensive cooperation with the Institute of Geonics of the CAS, Ostrava, Czech Republic, especially with Dr Kolcún. We continue with developing a high-accuracy, massive-parallel solver for binary masks of material regions determination in geocomposite CT-scans, especially reinforced ferro concrete geocomposite, and an FEM mesh builder, i.e. we will continue in the development and testing of the extended automatised version of SMRM, called Method of the Reference Material (MRM mentioned above).
Afterward, we will prepare a finite element model geometry of a real-world solid, starting with extensive plasticity computation on such model.

References

[1] Netgen Mesh Generator; 2009; URL https://sourceforge.net/projects/netgen-mesher/
[2] Steinhaus; “Sur la division des corps matériels en parties” (Conference); Bulletin of the Polish Academy of Science, volume 4, pages 801-804; 1956.
[3] Lloyd S. Least squares quantisation in PCM. IEEE Transactions on Information Theory. 1982, 28(2):129-137. Doi: 10.1109/tit.1982.1056489.
[4] Bradski; “Dr. Dobb’s Journal of Software Tools”; 2000.

Project website: permon.it4i.cz

PERIODIC DFT STUDIES OF ZEOLITE BASED CATALYSTS I.

Research institution:
J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Principal investigator:
Štěpán Sklenák

Project ID:
Open-4-12

Introduction

Zeolites represent the most important group of heterogeneous catalysts widely applied in industry. These crystalline microporous aluminosilicates with 3D channel structures are made of corner-sharing TO_4 tetrahedra (T = Si, Al^-). Isomorphous framework Al/Si substitutions result in a negative charge of AlO_4^- tetrahedral, which is balanced by extra-framework cationic species representing active sites for numerous redox- or base-catalysed reactions. The cations can occupy extra-framework sites which differ in the coordination, arrangement of the near environment, and location in zeolite channels.[1] These properties are controlled by the siting of the Al atoms of the negatively charged AlO_4^- tetrahedra in the framework crystallographic T sites. Recent attention has been focused mainly on metal ions in Si-rich zeolites showing unique catalytic redox behaviour.[1] There are properties of Si-rich zeolites which strongly limit the applicability of diffraction methods to determine the siting of cations: (i) a high number of crystallographically distinguishable framework T sites of which only some are partly occupied by Al atoms, (ii) low number of Al atoms in the framework (Si/Al > 8), and (iii) large unit cells.[2] Therefore the knowledge regarding the siting of cations and the local structure of cationic sites in Si-rich zeolites is very limited.[1,3] However, this knowledge is essential to evaluate the catalytic and sorption properties of cation exchanged zeolites. X-ray experiments only provide the positions of cations in the framework, but not the local structure of the cationic sites, as the coordinates of the cations are combined with the averaged coordinates of the framework reflecting mainly empty cationic sites and also the corresponding siliceous structures (i.e., without the framework Al/Si substitutions). Conversely, ^{6,7}Li MAS NMR spectroscopy represents a powerful tool to identify Li^+ siting. The issue of Li^+ ions incorporation into various matrixes has attracted particular attention due to its crucial importance in various fields such as, for example, energy storage and CO_2 capture.

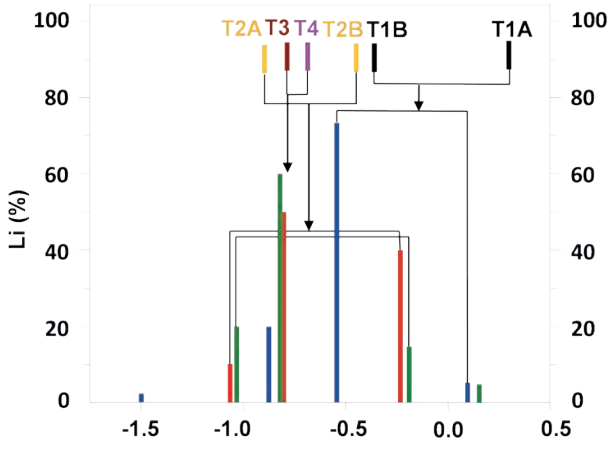


Figure 1.
Experimental ^{7}Li chemical shifts and their intensities in the spectra of the dehydrated FER/20 – FER/30 samples, ^{7}Li chemical shifts calculated for Li^+ ions balancing Al atoms in the T1-T4 sites, and their assignments to the experimental data. FER/20 (▮), FER/27 (▮), and FER/30 (▮); Li^+ balancing Al in the T1 (▮), T2 (▮), T3 (▮), and T4(▮) sites.

Results and Methods

Electronic Structure Calculations.

The CP2K suite of software [4] was employed using the BLYP functional, GTH pseudopotentials, and the TZV2P-GTH basis set. The energy cutoff of 1120 Ry was used.

Molecular Dynamics.

Molecular dynamics (MD) simulations were performed as implemented in the QUICKSTEP program,[5] a part of the CP2K suite of software.[4] Born-Oppenheimer MD simulations of 3,000–8,000 fs durations were performed in the canonical ensemble, with a time step of 1.0 fs and a mean temperature of 400 K regulated using a chain of Nose-Hoover thermostats. The duration of the MD simulation depended on the number of steps needed to equilibrate the calculated system. When the calculated

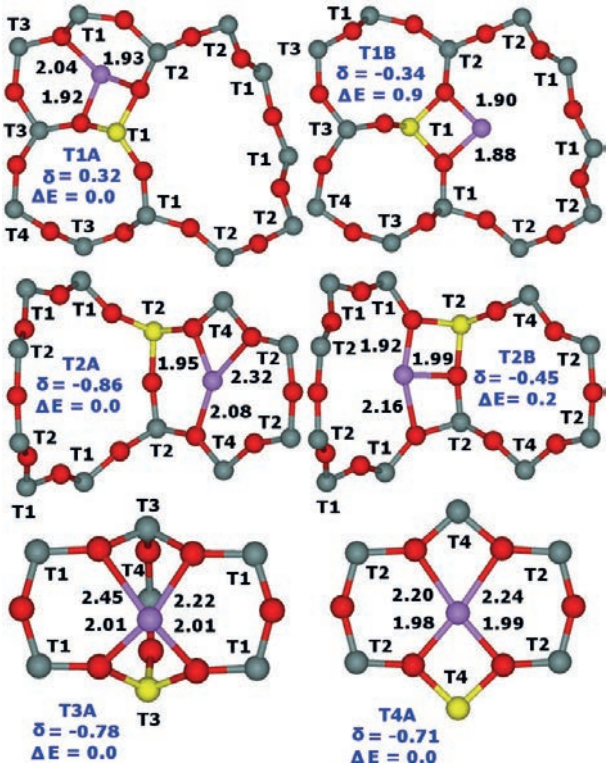


Figure 2.
Optimised structures (Li-O distances in Å) with the designations of the T sites of the low energy Li^+ sites (T1A and T1B for Al(T1), T2A and T2B for Al(T2), T3A for Al(T3), and T4A for Al(T4)), the relative energies in kcal/mol, and the corresponding ^7Li chemical shifts in ppm converted using the ^7Li NMR shielding of 90.16 ppm for $Li^+(H_2O)_4$ with the ^7Li chemical shift at 0.00 ppm. Silicon atoms are in grey, oxygen atoms in red, aluminium atoms in yellow, and lithium in violet.

systems were equilibrated, the structures of ten distinct "snapshots" were collected from each molecular dynamics simulation and optimised. The most stable structures of all distinct Li^+ sites for all five models (i.e., Al in all the five [6] framework T sites) were used for subsequent NMR computations.

Geometry Optimisations

The lattice parameters and the atomic positions were optimised employing conjugate-gradient algorithm minimisation of energies and forces as implemented in the QUICKSTEP program,[5] a part of the CP2K suite of software.[4]

Calculations of ^7Li NMR Shielding

Clusters of seven coordination shells around the Al atom (Al-O-Si-O-Si-O-Si-O-H_{link}) and the Li^+ ion were extracted from the optimised structures. Due to the presence of silicate rings in the framework of ferrierite, the created seven-shell clusters contained pairs of very close H_{link} atoms. Since the close H_{link} atoms represented the same Si atom, they were replaced by the corresponding $Si(OH_{link})_2$ moiety. This was repeated until the cluster contained no such pairs. The ^7Li NMR shielding values were calculated by the gauge independent atomic orbital method (GIAO) using Gaussian09, the B3LYP functional, and the pcS basis sets of Jensen; pcS-4 for the Li and Al atoms and pcS-1 for all the other atoms. The calculated ^7Li NMR shielding values were converted to ^7Li chemical shifts using the calculated shielding of 90.16 ppm for $Li^+(H_2O)_4$ (^7Li chemical shift is 0.00 ppm by definition). The geometry of $Li^+(H_2O)_4$ was optimised at B3LYP/cc-pVQZ and subsequently the ^7Li NMR shielding of 90.16 ppm was obtained at the same level of theory.

Discussion

The ^7Li NMR resonances identified for FER/20 – FER/30 using the ^7Li-^7Li correlation experiments are shown in Figure 1, which also depicts the relative concentrations of the Li^+ ions corresponding to the individual ^7Li NMR resonances obtained from the simulation of ^7Li MAS NMR spectra using the ^7Li chemical shifts from the ^7Li-^7Li correlation experiments.

Periodic DFT calculations including molecular dynamics conformational sampling of all possible Li^+ sites for Al(T1a), [6] Al(T1b), [6] Al(T2), Al(T3), and Al(T4) yield two low energy (i.e. their relative energy ≤ 2.0 kcal/mol with respect to the most stable Li^+ site for Al in a particular T site) Li^+ sites for T1 and T2, while only one for T3 and T4. The Li^+ sites for Al(T1a) [6] and Al(T1b) [6] are the same. Subsequently the ^7Li NMR shieldings are calculated using seven shell clusters cut out from the optimised structures and converted into ^7Li chemical shifts (Figure 2). The calculated relative energies of the other Li^+ sites are significantly higher and therefore they are not populated. The optimised structures of the low energy Li^+ sites together with the relative energies and the corresponding ^7Li chemical shifts are shown in Figure 2.

There are no experimental data based on diffraction methods regarding the siting of Li^+ in ferrierites. However, the knowledge of the Al siting in the three ferrierite samples used [6] permits a verification of the siting of Li^+ ions obtained in this study.[7] The Al atoms occupy the T sites in the samples (for details see Figure 6 of Ref.[6]) as follows: T2, T3, and T4 in FER/20; T1, T2, T3, and T4 in FER/27; and T1, T3, and T4 in FER/30. Figure 3 compares the relative concentration of Al atoms (in %) corresponding to the T sites obtained from (i) the corresponding Li^+ siting analysed using ^7Li MAS NMR and (ii) ^{27}Al MAS NMR experiments.[6]

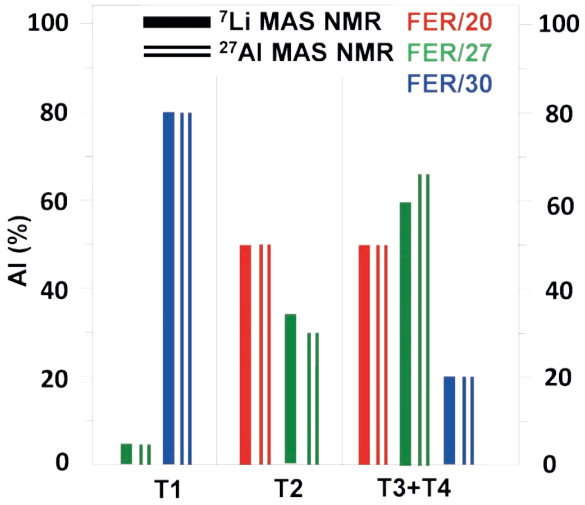


Figure 3.
Relative concentration of Al atoms (in %) corresponding to the distinguishable framework T sites obtained from (i) the corresponding Li^+ siting analysed using ^7Li MAS NMR and (ii) ^{27}Al MAS NMR.[6]

On-going Research / Outlook

In the near future, we will further investigate Li-FER samples with Al-O-(Si-O)$_2$-Al sequences.

Conclusion

We present a newly developed method to the determination of the siting of Li^+ and the local structure of Li^+ sites in crystalline aluminosilicate matrixes. This approach combines high resolution ^7Li-^7Li correlation MAS NMR spectroscopy with periodic DFT calculations of the structure of Li^+ sites and subsequent DFT cluster computations of the ^7Li NMR shielding. Comparison of the experimental ^7Li chemical shifts measured for the three ferrierite samples possessing high Si/Al ratios (i.e. low concentrations of Li^+) with those calculated at DFT results in assignment of the observed ^7Li NMR resonances to the Li^+ sites. The results show that the siting of Li^+ is controlled by the siting of Al atoms in the zeolite framework and that there are more than one low energy Li^+ sites for Al in some framework T sites.

References

[1] Dedecek, J., Sobalik, Z. and Wichterlova, B. Siting and Distribution of Framework Aluminium Atoms in Silicon-Rich Zeolites and Impact on Catalysis. Catalysis Reviews. 2012, 54, 135-223. Doi: 10.1080/01614940.2012.632662
[2] Database of Zeolite Structures. Available at: http://www.iza-structure.org/databases/
[3] Mortier, W.J. Extraframework cationic positions in zeolites, Elsewier, Amsterdam, 1982.
[4] The CP2K developers group. Available at: http://cp2k.org/.
[5] VandeVondele, J., Krack, M., Mohamed, F., Parrinello, M., Chassaing, T. and Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. Computer Physics Communications. 2005, 167, 103-128. Doi: 10.1016/j.cpc.2004.12.014
[6] Dedecek, J., Lucero, M.J., Li, C., Gao, F., Klein, P., Urbanova, M., Tvaruzkova, Z., Sazama, P. and Sklenak, S. Complex Analysis of the Aluminum Siting in the Framework of Silicon-Rich Zeolites. A Case Study on Ferrierites. Journal of Physical Chemistry C. 2011, 115, 11056-11064. Doi: 10.1021/jp200310b

Publication

[7] P. Klein, J. Dedecek, H.M. Thomas, S.R. Whittleton, V. Pashkova, J. Brus, L. Kobera, S. Sklenak. NMR crystallography of monovalent cations in inorganic matrixes: Li^+ siting and the local structure of Li^+ sites in ferrierites. Chemical Communications. 2015, 51, 8962-8965. ISSN 1359-7345

PERIODIC DFT STUDIES OF ZEOLITE BASED CATALYSTS II.

Research institution:
J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Principal investigator:
Štěpán Sklenák

Project ID:
Open-4-12

Introduction

The negative charge of the framework of zeolites originated from isomorphous substitution of SiO_4 by AlO_4^- tetrahedra, enables stabilisation of protons and/or cations and their oxo-complexes in the void volume of crystalline matrix. The concentration of Al in the zeolite framework not only determines the concentration of these counter-ion species, but also the acid strength of protons, their distances and redox behaviour of metal-ion species. High acid strength of protons is found with dealuminated faujasites (FAU) and mordenites of Si/Al (molar) > 5, and MFI and BEA* zeolites of Si/Al > 12. H-FAU, H-MOR, H-MFI and H-BEA* represent most important acid-zeolite structures exploited as catalysts in fluid catalytic cracking and hydrocracking units converting high-boiling oil fractions to valuable gasoline, diesel and other products as well as in base petrochemicals and fine chemicals production.

Results and Methods

Computational details

Structural models of the beta zeolite. Two models of the beta zeolite with the P1 symmetry were used to investigate the effect of Si/Al on the deprotonation energies. The first model (hereafter the 15 model) features four Al atoms in one unit cell (Si/Al 15) and the second model (hereafter the 5 model) contains ten Al atoms in one unit cell (Si/Al 5.4) (Figure 1).

All AlO_4^- entities are compensated by protons in AlOHSi groups. The Al_1, Al_2, and Al_3 atoms are substituted in both the models in the same framework positions to allow comparison of the deprotonation energies. According to our previous experimental studies on the distribution of Al atoms in beta zeolites with Si/Al > 11, two different distributions of the framework Al atoms in two double 6 MR units were chosen in the 15 model. One double 6 MR cage features two Al atoms (Al_1 and Al_2) forming an $Al_1SiSiAl_2$ sequence in one 6 MR (Figure 1), while the other double 6 MR unit contains two single Al atoms (one of them is Al_3) located in two 6 MRs (Figure 2).

Two variants of the distributions of Al in the 5 model were employed. Both the variants feature the same location of the Al_1Si-

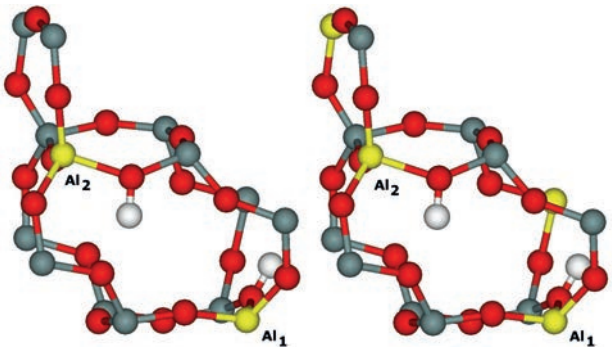


Figure 1.
Positions of the Al_1 and Al_2 atoms forming an $Al_1SiSiAl_2$ sequence in a double 6 MR unit. The 15 model (left) and the 5 model (right). In the 5 model, the Al_1 and Al_2 atoms also form Al_1SiAl and Al_2SiAl sequences besides creating $Al_1SiSiAl_2$. Silicon atoms are in grey, aluminium atoms in yellow, oxygen atoms in red, and hydrogen atoms in white.

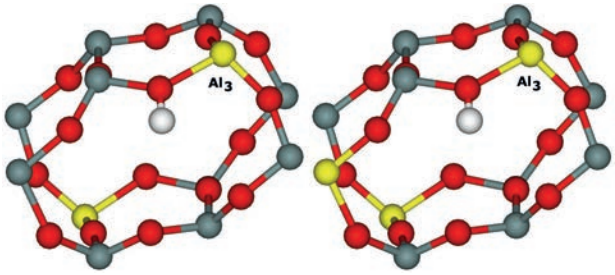


Figure 2.
Positions of single Al atoms in a double 6 MR unit. The 15 model and the 5 model, variant 2 with a single Al_3 atom (left) and the 5 model, variant 1 with Al_3SiAl (right). Silicon atoms are in grey, aluminium atoms in yellow, oxygen atoms in red, and hydrogen atom in white.

$SiAl_2$ sequence in one 6 MR as in the 15 model, however, the Al_1 and Al_2 atoms also form Al_1SiAl and Al_2SiAl sequences besides creating $Al_1SiSiAl_2$ as in the 15 model (Figure 1). The Al_3 atom forms an Al_3SiAl sequence in the variant 1 (Figure 2), while the Al_3 atom is a single Al atom in the variant 2 (Figure 2). AlSiAl sequences were employed in the 5 model since Si(2Si,2Al) atoms were observed by ^{29}Si MAS NMR in BEA-5. The starting geometry of both the models was generated from the experimental orthorhombic structure of polymorph A of the beta zeolite (http://www.iza-structure.org/databases: space group, P4122; cell parameters, a = 12.632, b = 12.632, c = 26.186 Å).

Periodic DFT calculations were carried out employing the VASP code.[1,2] The Kohn-Sham equations were solved variationally in a plane-wave basis set using the projector-augmented wave (PAW) method of Blöchl, as adapted by Kresse and Joubert. The exchange-correlation energy was described by the PW91 generalised gradient approximation (GGA) functional. Brillouin zone sampling was restricted to the Γ-point. A plane-wave cutoff of 400 eV was employed for geometry optimisations, while a smaller cutoff of 300 eV was used for the molecular dynamics simulations. The atomic positions were geometrically optimised by employing a conjugate-gradient algorithm minimisation of energies and forces, while the lattice parameters were fixed (constant volume) at their experimental values. The molecular dynamics (MD) simulations used the exact Hellmann-Feynman forces acting on atoms and applied the statistics of the canonical ensemble to the motion of the atomic nuclei using the Verlet velocity algorithm to integrate Newton's equations of motion. The time step for the integration of the equations of motion was 1 fs. The simulation was run for 5 ps at 300 K for both L and H models. MD simulations were carried out for all four possible locations of the proton on AlO_4^- tetrahedra (i.e. Al_1-O_1H-Si, Al_1-O_2H-Si, Al_1-O_3H-Si, Al_1-O_4H-Si, Al_2-O_1H-Si, …., Al_3-O_3H-Si, and Al_3-O_4H-Si) for the Al_1, Al_2, and Al_3 atoms, which are common for both the 15 and 5 models. The most stable position of the proton for each of the Al atoms was then used for subsequent calculations. The structures of ten distinct "snapshots" collected from the last 2,500 fs of the molecular dynamics simulations were optimised. The most stable conformer for both the models was then used for subsequent calculations.

Deprotonation energy

The deprotonation energies of the three protons compensating the three Al atoms which are common for both the 15 and 5 models were calculated using the equation $Z\text{-}H \rightarrow Z^- + H^+$. The calculated deprotonation energies are a measure of the acidity of the corresponding Brønsted sites. Details regarding calculations of the deprotonation energies employing periodic methods are discussed elsewhere. The differences between two deprotonation energies (i.e. relative deprotonation energies) are used to compare the acidity of the corresponding two Brønsted sites. The deprotonation energies of the Brønsted sites Al_1OHSi, Al_2OHSi, and Al_3OHSi were calculated for both the 15 and 5 models to estimate how the acid strength of the individual OH groups depends on Si/Al and the distribution of Al atoms in the framework.[3] Our calculations show for the 15 model for Al_1 and Al_2 forming $Al_1SiSiAl_2$ that the OH group belonging to Al_1 is free, while the other OH group of Al_2OHSi forms a hydrogen bond with the oxygen atom of a SiOSi chain in the 6 MR. Our results reveal higher deprotonation energies for Al_1OHSi and Al_2OHSi in the 5 model of BEA-5 (ΔE_{DP} 6.6 and 3.7 kcal/mol^{-1}, respectively) compared to the same acidic sites in the 15 model of BEA-11 since both Al_1OHSi and Al_2OHSi yield hydrogen bonds with the oxygen atom of SiOSi chains in the 5 MR and 6 MR, respectively. The higher deprotonation energy of the OH group belonging to Al_2 is in agreement with the fact that the corresponding hydrogen bond is stronger in the 5 model (the OH…O distance is 1.71 Å) than in the 15 model (the OH…O distance is 1.79 Å). It should be noted that both the Al_1 and Al_2 atoms also form Al_1SiAl and Al_2SiAl sequences due to a higher number of Al atoms in the 5 model. Two variants of the distribution of framework Al atoms were employed in the 5 model, one with an Al_3 atom forming an Al_3SiAl chain (variant 1) and the other one with a single Al_3 atom (variant 2). Al_3 is a single Al atom in the 15 model and the corresponding Brønsted OH group gives a hydrogen bond with an oxygen atom of a SiOSi chain in the 6 MR, while there is no such hydrogen bond in both the variants of the 5 model. Despite the absence of the hydrogen bond in both the variants of the 5 model, the deprotonation energy is slightly higher by 0.3 (single Al_3) and 1.1 (Al_3 in Al_3SiAl) kcal/mol than that of the 15 model.

The calculations show that the strength of the protonic acid sites in the 5 and 15 models does not significantly differ. The increase of the deprotonation energies for the 5 model relative to the 15 model (up to 6.6 kcal/mol) is smaller than the difference between the lowest and highest deprotonation energies calculated for the individual Brønsted sites AlOHSi for the 15 model (i.e., Δ = 8.7 kcal/mol because E_{DP} = 81.6 (Al_1) and 90.3 (Al_2) kcal/mol). This finding is also consistent with the small shift in the OH stretching frequency in the FTIR spectra.

On-going Research / Outlook

In the future, we will further investigate the acidity of H forms of zeolites.

Conclusion

Although a much higher concentration of protonic sites is present in the Al-rich beta zeolite associated with AlO_4^-, these Brønsted sites represent strongly acidic sites of strength comparable to that of the protons of Si-rich beta zeolites.

References

[1] Kresse, G. and Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B - Condensed Matter and Materials Physics. 1996, 54, 11169-11186. Doi: 10.1103/PhysRevB.54.11169

[2] Kresse, G. and Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational Materials Science. 1996, 6, 15-50. Doi: 10.1016/0927-0256(96)00008-0.

Publication

[3] P. Sazama, B. Wichterlova, S. Sklenak, V.I. Parvulescu, N. Candu, G. Sadovska, J. Dedecek, P. Klein, V. Pashkova, P. Stastny. Acid and redox activity of template-free Al-rich H-BEA* and Fe-BEA* zeolites. Journal of Catalysis. 2014, 318, 22-33. Doi: 10.1016/j.jcat.2014.06.024

PERIODIC DFT STUDIES OF ZEOLITE BASED CATALYSTS III.

Research institution:
J. Heyrovsky Institute of Physical Chemistry of the Czech Academy of Sciences

Principal investigator:
Štěpán Sklenák

Project ID:
Open-4-12

Introduction

Beta zeolites (the BEA* topology, polymorphs A and B) Si/Al > 11 belong to the most widely used aluminosilicates applied as catalysts in petrochemical technologies, syntheses of chemical commodities as well as fine specialties. The framework of the beta zeolite possesses three-dimensional intersecting channels of 6.6 x 6.7 Å size providing advantageously easy diffusion even of voluminous organic molecules. The recent discovery of the synthesis of new Al-rich beta zeolites with Si/Al molar ratios of 4-5 (i.e. a markedly increased concentration of framework Al atoms and thus the total negative charge of the framework) represents a substantial step ahead in future advanced applications of zeolites of the BEA* topology. The high concentration of framework Al atoms enables the accommodation of a much higher amount of protonic and/or metal-ion species compared to Si-rich BEA*. In addition, since Al-rich beta zeolites are prepared by employing a template-free hydrothermal synthesis route using seeding of BEA* crystals, their production in a large industrial scale could lead to great cost-benefits compared to conventional Si-rich beta zeolites synthesised exclusively in the presence of templates.

The benefit of a high concentration of Al in Al-rich BEA* and thus an increased concentration of various metal-ion and metal-oxo species has been manifested in the high activity of Fe- and Cu-Al-rich BEA* in selective reduction of nitrogen oxides and N_2O decomposition. Since the structure and location of acidic or metal-ion sites, decisive for their catalytic function, are directly related to the location and distribution of Al atoms in the Framework, we attempt to analyse the arrangements and properties of Al atoms in Al-rich BEA* and relate them to the structure of counter-ion species. AlSiAl sequences are absent in the framework of Si-rich pentasil-ring zeolites (Si/Al > 8.5) of the MFI, FER, MOR and BEA* topology.

Results and Methods

Periodic DFT calculations were employed to investigate siting of Co(II) ion probes in the dehydrated Co-BEA* zeolites. Two models possessing the P1 symmetry, designated as $ß_{T6T6}$ and $ß_{T4T4}$ were employed to investigate two distinct possible arrangements of Co(II)-ß cationic sites formed by the deformed 6MR. They feature one unit cell of the beta zeolite with two differently oriented AlSiSiAl sequences in the 6MR accommodating one Co(II) ion. The starting orthorhombic structure was downloaded from the zeolite structural database. In addition, four models featuring the Co(II) in α-type 6MRs of cationic sites of the FER, MFI, MOR, and BEA* zeolite topologies were used to compare the structures of the bare Co(II)-α ions in various pentasil-ring zeolites. The starting structures were obtained from the zeolite structural database. Periodic DFT calculations were carried out employing the VASP code [1-4]. The high-spin electron configuration Co $d^5\uparrow d^2\downarrow$ was employed for the Co(II) ion accommodated in the zeolite.[5] The Kohn-Sham equations were solved variationally in a plane-wave basis set using the projector-augmented wave (PAW) method of Blöchl, as adapted by Kresse and Joubert. The exchange-correlation energy was described by the PW91 generalised gradient approximation (GGA) functional. Brillouin zone sampling was restricted to the Γ-point. The plane-wave cutoff of 400 eV was utilised for geometry optimisations while a smaller cutoff of 300 eV was used for the molecular dynamics simulations. The atomic positions were optimised at constant volume employing a conjugate-gradient algorithm minimisation of energies and forces while the lattice parameters were fixed at their experimental values. The molecular dynamics (MD) simulations used the exact Hellmann-Feynman forces acting on atoms and applied the statistics of the canonical ensemble to the motion of the atomic nuclei using the Verlet velocity algorithm to integrate Newton's equations of motion. The time step for the integration of the equations of motion was 1 fs. The simulations were run for 10 ps at 400 K. The structures of ten "snapshots" collected from the molecular dynamics simulations were optimised for each model.

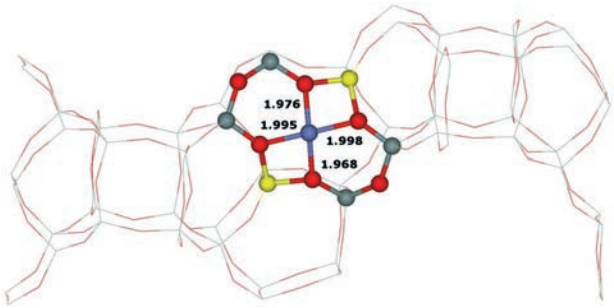


Figure 1.
Optimised structures of the $ß_{T6T6}$ site after molecular dynamics simulations. The distances are in angstroms. Silicon atoms are in grey, oxygen atoms in red, aluminium atoms in yellow, and cobalt atoms in blue.

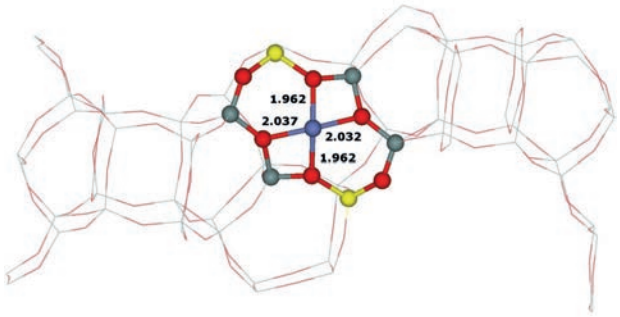


Figure 2.
Optimised structures of the $ß_{T4T4}$ site after molecular dynamics simulations. The distances are in angstroms. Silicon atoms are in grey, oxygen atoms in red, aluminium atoms in yellow, and cobalt atoms in blue.

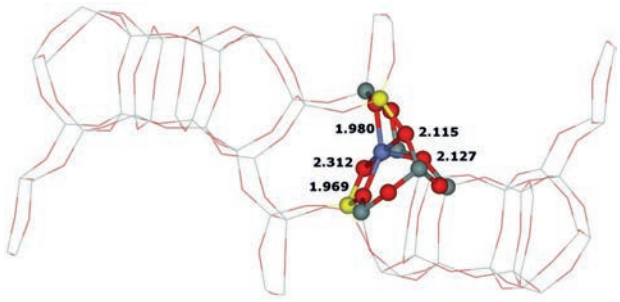


Figure 3.
Optimised structures of the α site after molecular dynamics simulations. The distances are in angstroms. Silicon atoms are in grey, oxygen atoms in red, aluminium atoms in yellow, and cobalt atoms in blue.

Computations on the structure and stability of the Co(II) ions in cationic sites

Our calculations of the two models of the ß type 6MRs with different location of AlSiSiAl sequences ($ß_{T6T6}$ and $ß_{T4T4}$) yielded the structures of the exchanged Co(II) ions in $ß_{T6T6}$ and $ß_{T4T4}$ sites (Figures 1 and 2, respectively).

The Co-O bond length ranged from 1.97 (Co-O_{Al}) to 2.00 (Co-O_{Al}) Å and 1.96 (Co-O_{Al}) to 2.04 (Co-O_{Si}) Å for the $ß_{T6T6}$ and $ß_{T4T4}$ models of Co-BEA*, respectively. The bond lengths of the first Co-O shell determined by EXAFS in Co(II)-MFI and Co(II)-FER zeolites with prevailing population of the Co(II)-ß type ions ranged from 1.99 to 2.00 Å with four or three oxygen ligand coordination. This Co-O distance close to 2.00 Å satisfies requirements of Co(II) bonding to oxygen atoms in both zeolite structures. This is achieved by the deformation of the framework T-O bonds induced under the Co(II) ion coordination to framework rings, as previously observed by IR spectra of T-O framework vibrations. The Co(II) ion accommodated in the $ß_{T6T6}$ site is ligated to four oxygen atoms of two AlO_4^- tetrahedra, while the Co(II) ion in $ß_{T4T4}$ coordinates to two O atoms of two AlO4- tetrahedra and two O atoms of two SiO_4 tetrahedra. This presumes that the ligand field of the Co(II) ions in $ß_{T6T6}$ and $ß_{T4T4}$ differs and thus, Co(II) ions in these sites should differ in their d-d spectra. Our prior computational studies of Fe(II)-FER, Co(II)-FER, and Cu(II)-FER showed significant rearrangements of the α and $ß_{T4T4}$ rings of cationic sites upon binding the divalent cations.[5-7] The reason why the $ß_{T4T4}$ site accommodating Co(II) ion does not similarly rearrange to bind to four O atoms of two AlO4- tetrahedra is most likely that the 6MR forming the ß site creates a rigid double 6MR (hexagonal prism) with another 6MR. The stabilisation energy obtained from the preferential binding of Co(II) to four O atoms of two AlO_4^- tetrahedra is most probably smaller than the energy needed to deform the structure of the double 6MR to permit the preferential binding of Co(II). Similarities in the stability, coordination and Co-O bond distances of Co(II) ions in cationic sites were also indicated by MD simulations for the Co(II)-α ions bonding in BEA* (Figure 3). The calculations of Co(II) ion coordination and Co-O bond length support structure stability of the bare Co(II) ions coordinated in the α and ß type 6MRs of dehydrated BEA* and balancing the negative AlO_4^- charges of their AlSiSiAl sequences and the ability of the given rings to perturb their framework T-O bonds thus satisfying requirements of Co(II) ions to be bound to framework oxygen atoms. These results imply only small differences in the geometry of α and ß type framework rings coordinated Co(II) ions of the respective pentasil-ring zeolites and thus in the ligand field of Co(II) ions in the corresponding sites.

On-going Research / Outlook

In the near future, we will further investigate other metal and metal-oxo species.

Conclusion

The arrangement of Si-Al atoms in the BEA* framework and that occurring in the inner surface of the channels were investigated for the synthesised Al-rich beta zeolites and compared to those of Si-rich BEA*. While Si-rich BEA* (Si/Al > 12) do not exhibit AlSiAl sequences, in Al-rich BEA* (Si/Al ~ 4-5) the concentration of Al atoms in AlSiAl range from 40 to 100% of the total Al depending on their actual arrangement (in rings or as infinite sequence) in the framework.

References

[1] Kresse, G. and Hafner, J. Ab initio molecular dynamics for open-shell transition metals. Physical Review B. 1993, 48, 13115-13118. Doi: 10.1103/PhysRevB.48.13115

[2] Kresse, G. and Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. Physical Review B. 1994, 49, 14251-14269. Doi: 10.1103/PhysRevB.49.14251

[3] Kresse, G. and Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B. 1996, 54, 11169-11186. Doi: 10.1103/PhysRevB.54.11169

[4] Kresse, G. and Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational Materials Science. 1996, 6, 15-50. Doi: 10.1016/0927-0256(96)00008-0

[5] Sklenak, S., Andrikopoulos, P.C., Whittleton, S.R., Jirglova, H., Sazama, P., Benco, L., Bucko, T., Hafner, J. and Sobalik, Z. Effect of the Al Siting on the Structure of Co(II) and Cu(II) Cationic Sites in Ferrierite. A Periodic DFT Molecular Dynamics and FTIR Study. Journal of Physical Chemistry C. 2013, 117, 3958-3968. Doi: 10.1021/jp310236d

[6] Sklenak, S., Andrikopoulos, P.C., Boekfa, B., Jansang, B., Novakova, J., Benco, L., Bucko, T., Hafner, J., Dedecek, J. and Sobalik, Z. N_2O decomposition over Fe-zeolites: Structure of the active sites and the origin of the distinct reactivity of Fe-ferrierite, Fe-ZSM-5, and Fe-beta. A combined periodic DFT and multispectral study. Journal of Catalysis. 2010, 272, 262-274. Doi: 10.1016/j.jcat.2010.04.008

Publication

[7] P. Sazama, E. Tabor, P. Klein, B. Wichterlova, S. Sklenak, L. Mokrzycki, V. Pashkkova, M. Ogura, J. Dedecek. Al-rich beta zeolites. Distribution of Al atoms in the framework and related protonic and metal-ion species. Journal of Catalysis. 2016, 333, 102-114. Doi: 10.1016/j.jcat.2015.10.010

QUANTUM-MECHANICAL SIMULATIONS OF INTERFACES IN ADVANCED MATERIALS

Research institutions:
Central European Institute of Technology, Masaryk University, Institute of Physics of Materials of the Czech Academy of Sciences

Principal investigator:
Mojmír Šob

Researchers:
Petr Šesták, Miroslav Černý, Petr Řehák, Monika Všianská, Martin Zelený, Jana Pavlů, Martin Friák, Tomáš Komárek, Ondřej Svoboda, David Nezval

Project partner:
Montanuniversität Leoben

Project ID:
OPEN-5-10, OPEN-6-10

Introduction

Our highly industrialised European society requires completely new types of materials with unprecedented properties. For example, materials that can operate at higher temperatures in energy conversion units (e.g. engine parts, turbine blades) would allow for their higher efficiency and, consequently, lower green-house-gas emissions and cleaner technologies. In order to meet a number of strict criteria and provide different properties, new materials will consist of a number of different phases coexisting in complex polycrystalline aggregates. When searching for and designing these new metallic alloys the key issue is their thermodynamic stability, which must be guaranteed even in the case of thermal and mechanical cycling loading. In order to properly address these conditions, the influence of internal interfaces between different grains as well as stresses and strains must be combined with traditional thermodynamic phenomenological modelling.

Internal interfaces in solids, in particular grain boundaries (GBs), represent one of the most important types of extended defects and their instability-related failure may have catastrophic consequences. GBs determine a number of materials properties. For example, macroscopic strength of polycrystalline materials depends strongly on GB cohesion. GBs are challenging for theoretical studies because of their distorted atomic structure. Using quantum-mechanical methods, we can reliably determine their properties.

Properties of GBs and associated interface states are very sensitive to various compositional changes (e.g. impurity segregation). It was found that the impurities even in very low concentrations (ppm) segregated at GBs can drastically change material properties [1,2]. For instance, they can even completely suppress local magnetic moments at GBs in magnetic materials as shown by quantum-mechanical calculations in elemental Ni [3,4]. Considering the current interest in fabrication, characterisation and use of ultra-fine grained (UFG) or nano-granular materials that can exhibit unprecedented properties not available in the bulk, GBs can become even more important than the grain interior. When reaching truly nano-scale grain sizes, even elastic properties (that are typically unaffected in case of larger grains) become dominated by GBs.

The aim of our project was to obtain new fundamental information on the thermodynamic stability of modern advanced materials and to determine the structure and properties of their interfaces including the mechanical and electronic characteristics. Our work significantly contributes to the knowledge-based design of new materials for energy applications. In particular, we shed new light on tensorial elastic properties of GBs as a critically-important complement to a number of previous GB-related studies that were focused exclusively on scalar characteristics (energies, strength, changes in inter-atomic bonds) or electronic and atomic structure. The fact that tensorial properties often require much more demanding approaches [5] is compensated by the wealth of insight that they provide. The knowledge of elastic properties allows for assessing the mechanical stability of the studied system. Elastic constants also play a crucial role in high-temperature thermodynamic stability as they determine the behaviour of long-wave phonons.

As a case study, we selected one of the most important intermetallic compounds for industrial applications, Ni_3Al (appearing, e.g., as the γ' phase in Ni-based superalloys), and the Σ5(210) GB (Figure 1), which is the smallest one that exhibits an extra volume compared to the bulk.

This particular research was designed to take full advantage of the IT4Innovations Salomon supercomputer, which allows us to perform computationally extremely demanding quantum-mechanical calculations of grain boundaries in advanced materials employing supercells with a much higher number of atoms than what is possible in other computer centres.

Methods and Results

Our ab initio calculations were performed within the framework of the density functional theory using the Vienna Ab initio Simulation Package (VASP) [6,7]. The exchange and correlation energy was treated in the generalised gradient approximation as parametrised by Perdew, Burke, and Ernzerhof [8] and implemented in projector augmented wave pseudopotentials [9]. A plane-wave cut-off energy of 500 eV with a 5×17×3 Monkhorst-Pack k-point mesh was used. Elastic constants were computed by the stress-strain method [10].

Our calculations show that the $\Sigma5(210)^{Al,Ni}$ GB with both types of atoms (Al and Ni) at the interface (Figure 1a) has a lower GB energy than the $\Sigma5(210)^{Ni,Ni}$ GB variant (Figure 1b). Calculated elastic constants of the $\Sigma5(210)^{Al,Ni}$ GB are given in Table 1 together with bulk elastic constants in the same coordination system spanned by the vectors $\langle 1\bar{2}0\rangle$, $\langle 001\rangle$, $\langle 210\rangle$. Analysing the computed elastic constants listed in Table 1, it is important to note that the computed shear elastic constant C_{55} of the $\Sigma5(210)^{Al,Ni}$ GB is considerably lower (29 GPa) than that in the bulk (76 GPa) and clearly represents the weakest link for the mechanical stability.

In order to visualise how the $\Sigma5(210)^{Al,Ni}$ GB responds to uniaxial loading along different directions, we exhibit directional dependences of the Young's modulus in Figure 2. The stiffest directions with the Young's modulus of about 250 GPa (colour-coded cyan in Figure 2) are within the GB interface, while the softest directions with the Young's modulus approaching 50 GPa (reddish colours in Figure 2) are inclined to the GB interface.

Conclusion

First-principles electronic structure calculations were applied to study tensorial elastic properties of the Ni_3Al $\Sigma5(210)$ GBs. The elasto-chemical aspects were illustrated by the behaviour of elastic constants computed for GBs with different atoms at the interface and by their strong sensitivity to the interface chemical composition. A few-fold reduction of the elastic constants C_{55} is identified as the crucial weakest link for the mechanical stability.

Outlook

In the upcoming months, an extension of our study is planned to address segregation-driven compositional changes at the $\Sigma5(210)$ GBs in Ni_3Al. These phenomena will be exemplified by comparing properties with and without segregated sp impurities as in Refs. [3,4].

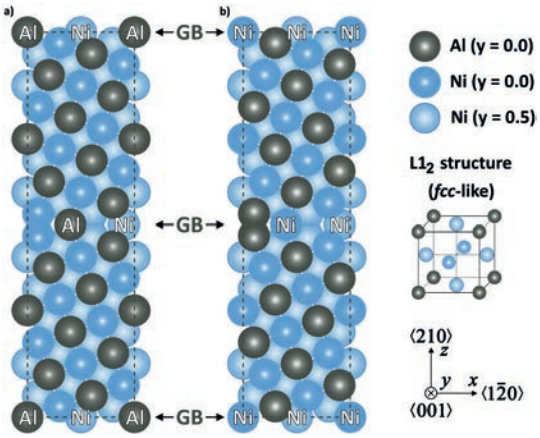


Figure 1.
Schematic visualisation of two types of 72-atom supercells used in our electronic structure calculations of the $\Sigma5(210)$ grain boundaries in Ni_3Al with different interface stoichiometries, in particular, $\Sigma5(210)^{Al,Ni}$ with both Al and Ni atoms at the GB (a) and $\Sigma5(210)^{Ni,Ni}$ with only Ni atoms in the GB plane (b).

C_{11} = 248 (298)	C_{12} = 152 (152)	C_{13} = 115 (100)
C_{22} = 218 (246)	C_{23} = 148 (152)	C_{33} = 246 (298)
C_{44} = 61 (128)	C_{55} = 29 (76)	C_{66} = 116 (128)

Table 1.
The ab initio computed elastic constants (GPa) of the $\Sigma5(210)^{Al,Ni}$ GB variant compared with the values computed for the Ni_3Al bulk (in brackets) in the coordination system of the studied GBs (shown in Figure 1).

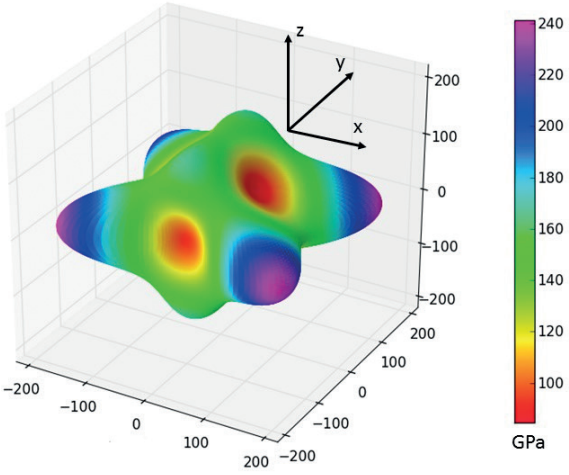


Figure 2.
Directional dependence of the Young's modulus of the studied $\Sigma5(210)^{Al,Ni}$ GB when the GB interface is the x-y plane. Displayed dependences were computed and visualised by the SC EMA library (scema.mpie.de) [11-13].

Acknowledgement

We gratefully acknowledge the financial support from the Ministry of Education, Youth and Sports of the Czech Republic under the Project CEITEC 2020 (Project No. LQ1601), the Czech Science Foundation (Projects No. GA14-22490S and GA 16-24711S), the Academy of Sciences of the Czech Republic (Institutional Project No. RVO:68081723) and the Fellowship of Jan Evangelista Purkyně (awarded to Dr Martin Friák). Computational resources were provided by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects IT4Innovations (Project No. LM2015070), CESNET (Project No. LM2015042) and CERIT-SC (Project No. LM2015085) within the programme "Projects of Large Research, Development and Innovations Infrastructures."

References

[1] Sutton, A.P. and Balluffi, R.W. Interfaces in crystalline materials. Oxford: Clarendon Press, 1995. ISBN 0-19-850061-0.
[2] Lejček, P. Grain boundary segregation in metals. Heidelberg: Springer-Verlag, 2010. ISBN 9783642125058.
[3] Všianská, M. and Šob, M. The effect of segregated sp-impurities on grain-boundary and surface structure, magnetism and embrittlement in nickel. Progress in Materials Science. 2011, 56(6), 817-840. Doi: 10.1016/j.pmatsci.2011.01.008.
[4] Všianská, M. and Šob, M. Magnetically dead layers at sp-impurity-decorated grain boundaries and surfaces in nickel. Physical Review B. 2011, 84, -. Doi: 10.1103/physrevb.84.014418.
[5] van de Walle, A. A complete representation of structure-property relationships in crystals. Nature Materials. 2008 June, 7(6), 455-8. Doi: 10.1038/nmat2200.
[6] Kresse, G. and Hafner, J. Ab initio molecular dynamics for liquid metals. Physical Review B. 1993, 47(1), 558-561. Doi: 10.1103/PhysRevB.47.558.
[7] Kresse, G. and Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B. 1996, 54, 11169-11186. Doi: 10.1103/PhysRevB.54.11169.
[8] Perdew, J.P., Burke, K. and Ernzerhof, M. Generalised gradient approximation made simple. Physical Review Letters. 1996, 77(18), 3865-3868. Doi: 10.1103/PhysRevLett.
[9] Blöchl, P.E. Projector augmented-wave method. Physical Review B. 1994, 50(24), 17953-17979.
[10] Zhou, L., Holec, D. and Mayrhofer, P.H. First-principles study of elastic properties of cubic $Cr_{1-x}Al_xN$ alloys. Journal of Applied Phys. 113, 043511. 2013, 113(7), 079901-. Doi: 10.1063/1.4793084.
[11] Titrian, H. et al. Self-consistent scale-bridging approach to compute the elasticity of multi-phase polycrystalline materials. Materials Research Society Proceedings. 2013, 1524. Doi:10.1557/opl.201.41.
[12] Friák, M. et al. Theory-guided materials design of multi-phase Ti-Nb alloys with bone-matching elastic properties. Materials. 2012, 5(10), 1853-1872. Doi: 10.3390/ma5101853.
[13] Zhu, L.-F. et al. Ab initio study of single-crystalline and polycrystalline elastic properties of Mg-substituted calcite crystals. Journal of the Mechanical Behaviour of Biomedical Materials. 2013, 20, 296-304. Doi: 10.1016/j.jmbbm.2013.01.030.

Publications

[14] Friák, M., Všianská, M., Holec, D. and Šob, M. Tensorial elastic properties and stability of interface states associated with Σ5(210) grain boundaries in $Ni_3(Al,Si)$. Science and Technology of Advanced Materials (2016), submitted.
[15] Friák, M., Všianská, M., Holec, D. and Šob, M. Tensorial elastic properties and mechanical stability of Σ5(210) grain-boundary complexions in Ni_3Al from ab initio calculations. 8th Int. Conf. on Materials Structure & Micromechanics of Fracture (MSMF8), Brno, 27-29 June 2016.

EXPLORING THE EFFECT OF LOADING CONDITIONS ON THE IDEAL TENSILE STRENGTH

Research institutions:
Central European Institute of Technology, Masaryk University, Institute of Physics of Materials of the Czech Academy of Sciences

Principal investigator:
Mojmír Šob

Researchers:
Miroslav Černý, Petr Šesták, Petr Řehák, Monika Všianská

Project partner:
Central European Institute of Technology, Brno University of Technology

Project ID:
Open-6-10

Introduction

During the last two decades, theoretical calculations regarding atomic configuration and properties of extended defects in materials became possible using ab initio electronic structure (ES) calculations, based on the fundamental quantum theory (Schrödinger equation). Here the atomic numbers of constituent atoms and some structural information are employed as the only input data. Such calculations are routinely performed within the framework of the density functional theory (DFT). Thus, for a given material, we can compute the total energies, equilibrium lattice parameters, elastic moduli, relative stabilities of competing crystal structures, energies associated with point and planar defects, alloy heats of formations, etc. In addition, we also obtain information about electronic densities of states and charge densities (see Figure 1) that enables us to attain a deeper insight into structure and properties of solids.

Real materials do not have a perfect crystal structure. They usually contain many defects. Planar defects like grain boundaries (GB) predetermine, to a large extent, mechanical properties of polycrystalline materials and, in particular, their strength. Therefore, a lot of effort has been devoted to study these phenomena not only at macroscopic level, but also at nano- and atomistic scales using simulations based on the ES calculations [1-4]. Another very important aspect of GBs in solids is segregation of impurities to the GBs. Segregated impurities may have drastic effects on the strength and brittleness of GBs. Due to impurity segregation, the GB cohesion may be considerably reduced and the material may become brittle [1]. Further, the GBs may be preferentially attacked by corrosive agents. It turns out that the magnetic effects are also very important in the segregation – some impurities reduce magnetic moments of atoms at the grain boundary [2-3].

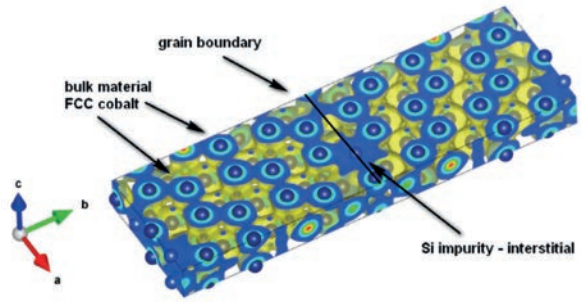


Figure 1.
3D representation of the charge density for grain boundary in cobalt

First principles tensile tests are often employed to explore the effect of segregated impurity atoms on the GB cohesion. Quantities such as the tensile strength or cleavage energy represent the typical output information of these tests and serve as a platform for predicting changes in mechanical response of impurity decorated GBs. The loading axis is typically oriented perpendicularly to the GB plane.

In our research we simulated the uniaxial tensile test of fcc nickel and cobalt containing the Σ5(210) tilt GB with 12 segregated sp-impurities. Segregation preference of the impurity atoms at GB was determined with the help of calculations of segregation energy [2-3]. It turned out that Al, Ga, In, Sn, Sb and Te preferred the substitutional positions and Si, P, S, Ge, As and Se the interstitial positions. The computational tensile tests considered two distinct approaches. One of them conformed to a separation of rigid grains simulating thus brittle cleavage along a defined cleavage plane. This approach was employed in the very first computational tensile tests [1] and has been routinely employed in many theoretical studies. We will refer to this model as to the rigid grain shift (RGS). The other approach was computationally more demanding since it also includes the Poisson contraction and full relaxation of the lattice. Thus, this model corresponds to the optimised uniaxial loading (OUL).

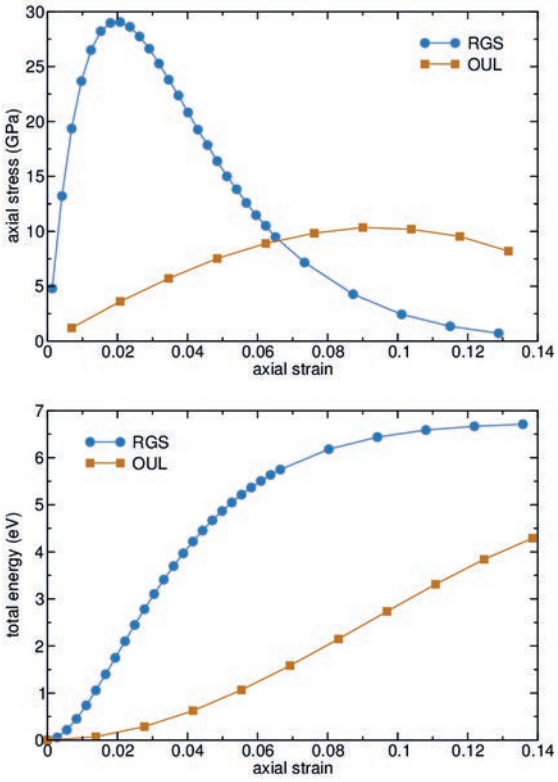


Figure 2.
The axial stress and the total energy as functions of the axial strain computed for Σ5(210) GB in fcc Ni with substitutionally segregated phosphorus

Methods

The calculations were performed using the ab-initio total-energy and molecular-dynamics program VASP (Vienna ab-initio simulation program) developed at the Fakultät für Physik of the Universität Wien (http://www.vasp.at/). This code is very well parallelised and enables running one particular simulation on multiple threads nodes [5-7]. This allows performing the ab initio simulations in reasonable time. Optimisation of atomic positions within the supercell (OUL model) was performed using the internal VASP procedure while for optimisation of the cell shape we used our own external procedure that cooperated with the VASP code via reading its output files and writing new structure files.

The VASP program was exclusively compiled and tested using the Intel Fortran Compiler for the supercomputers available at the IT4Innovations National Supercomputing Center. This enabled us to perform the simulations with a high efficiency.

Results

Comparison of the calculated values of the tensile strength for GBs in fcc Ni and Co with and without impurities shows that the RGS model predicts almost three times higher values than the OUL model. This is illustrated in Figure 2 for the particular example of phosphorus atoms segregated in substitutional positions at GB in Ni.

This reveals that the Poisson contraction included in the OUL model plays an important role, and this difference in strength values can be explained by the fact that omitting the Poisson contraction during the tensile tests induces tensile transverse stresses that increase the tensile strength.

Another important finding is that predictions of strengthening or weakening the GB by segregation of impurity atoms based on the two models differ for most of elements. This qualitative difference deserves an explanation. For this reason, we computed the tensile strengths of a clean GB and impurity decorated GBs as functions of the transverse stresses. The computed results reveal that the tensile strength of the clean GB almost linearly increases with the increasing transverse stresses. On the other hand, the response of GB to transverse stresses dramatically changes with the presence of impurities. Due to a lower sensitivity of the tensile strength of impurity decorated GBs to the transverse stresses, the prediction of strength enhancement or reduction depends on the level of transverse stresses. This, of course, results in opposite predictions supplied by the two investigated models.

Conclusion

The main outcomes can be summarised as follows:

- The difference between computed strength values for two deformation models demonstrates that the choice of deformation model can play a crucial role in predicting the effect of impurities at GBs.
- The tensile strength of GBs increases with increasing lateral stresses.
- The different effect of transverse stresses on the tensile strength of GBs with and without segregated impurities explains the qualitative difference in results of the two distinct models considered in our study.

Most of these results have been summarised in our scientific paper [8] and also presented at international conferences [9–12].

References

[1] M. Yamaguchi, M. Shiga, Kaburaki, H. Grain boundary decohesion by impurity segregation in a nickel-sulfur system. Science. 2005, 307(5708), 393-397. Doi: 10.1126/science.1104624

[2] Všianská, M. and Šob, M. Magnetically dead layers at sp-impurity-decorated grain boundaries and surfaces in nickel. Physical Review B. 2011, 84, -. Doi: 10.1103/physrevb.84.014418.

[3] Všianská, M. and Šob, M. The effect of segregated sp-impurities on grain-boundary and surface structure, magnetism and embrittlement in nickel. Progress in Materials Science. 2011, 56(6), 817-840. Doi: 10.1016/j.pmatsci.2011.01.008.

[4] Liu, W., Ren, C., Han, H., Tan, J., Zou, Y., Zhou, X., Huai, P., Xu. H. First-principles study of the effect of phosphorus on nickel grain boundary. J. Appl. Phys. 2014, 115(4), 043706-7. Doi: 10.1063/1.4863181.

[5] Kresse, G. and Hafner, J. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. J. Phys.: Condens. Matter. 1994, 6(40), 8245-8257. Doi: 10.1088/0953-8984/6/40/015.

[6] Kresse, G. and Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B. 1996, 54, 11169-11186. Doi: 10.1103/PhysRevB.54.11169.

[7] Kresse, G. and Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. Phys. Rev. B. 1999, 59(3), 1758-1775. Doi: 0.1103/PhysRevB.59.1758.

Publications

[8] Černý, M., Šesták, P., Řehák, P., Všianská, M., and Šob, M. Ab initio tensile tests of grain boundaries in the fcc crystals of Ni and Co with segregated sp-impurities. Mater. Sci. Eng. A. 2016, 669, 218-225. Doi: 10.1016/j.msea.2016.05.083.

[9] Šesták, P., Černý, M., Řehák, P., Všianská, M., Šob, M.: Theoretical strength of Σ5(210) GB in FCC cobalt with segregated interstitial and substitutional sp-impurities, 2015 Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Berlin, Germany 15-20 March 2015 (oral presentation).

[10] Černý, M., Šesták, P., Řehák, P., Všianská, M., Šob, M.: Tensile strength of Ni grain-boundary with segregated sp-impurities. 2015, Frühjahrstagung der Deutschen Physikalischen Gesellschaft, Berlin, Germany 15-20 March 2015 (oral presentation).

[11] Černý, M., Šesták, P., Řehák, P., Všianská, M., Šob, M.: First principles tensile test of Ni grain-boundary with segregated sp impurities. 17th International Conference on Strength of Materials (ICSMA 17), Brno, 9-14 August 2015 (oral presentation).

[12] Šesták, P., Černý, M., Řehák, P., Všianská, M., and Šob, M.: Tensile strength of Σ5 (210) grain boundary in cobalt with segregated sp-impurities, 17th International Conference on Strength of Materials (ICSMA 17), Brno, 9-14 August 2015 (oral presentation).

HYDROPHOBIC IMPURITIES IN WATER CLUSTERS

Research institution:
IT4Innovations National Supercomputing Center, VŠB – Technical University of Ostrava

Principal investigator:
Aleš Vítek

Researchers:
René Kalus, Daniel J. Arismendi-Arrieta, Rita Prosmiti

Project partner:
Instituto de Fisica Fundamental, CSIC, Madrid

Project ID:
OPEN-7-48

Introduction

Crystalline hydrates are formed by many hydrophobic gases that do not exhibit strong attractive interactions with water, and in most cases the gas molecules or atoms occupy cages formed by a network of hydrogen bonds. Gas molecules or atoms usually reside inside one of a few cage structures, and such structures are called clathrate hydrates. Clathrate hydrates have been extensively studied because they occur widely in nature and are of great environmental importance. Methane hydrate consists of methane molecules trapped in the cavities of hydrogen bonds network can be found at the bottom of the sea or in permafrost of Siberia. It is very delicate structure that may be a very promising resource of energy for the coming future, or on the other hand if the conditions and climate of the world change and temperatures rise, escaped methane gas from this structure may pose a big threat as a strong greenhouse gas.

A number of spectroscopic and diffraction studies have been conducted to explore under which conditions bulk rare-gas hydrates may be formed or decomposed and what types of these hydrates may be obtained under such conditions [1]. Theoretical investigation of these systems is useful for interpreting properly different experimental observations. Since the formation of bulk clathrate hydrates must start from microscopic seeds (clusters), it seems reasonable to start such investigations at this level. Numerous theoretical studies investigating the properties of water clusters have been published in the past, but almost all of them investigated water clusters under zero pressure, and only a few studies have performed computations for non-zero pressure and describe the structural changes induced by non-zero pressure [2]. Rare gas atom admixture is the first step of our research because of an effort to minimise computation demands, but if we replace it with a spherically symmetric methane molecule it can bring completely new theoretical insight into the stability and properties of methane hydrate.

Our research has two goals; firstly to theoretically describe the behaviour of guest hydrophobic particles in the water environment. The second goal is from the basic research, to develop and test realistic methodologies for simulation of a finite system under high pressure using a realistic barostat in the Monte Carlo method based on extended Lagrangian $H = E^{kin} – E^{int} – PV$, where the representative volume V will be calculated using convex hull algorithms. This extended Lagrangian method will be used in the recently improved 2D multiple histogram method [3].

Results and Methods

Using the isobaric-isothermal parallel tempering Monte Carlo method together with a multiple histogram method, we have performed numerical simulations of two finite systems, $(H_2O)_{20}$ and $Kr(H_2O)_{20}$ to investigate in which thermodynamic conditions a water molecules create a stable cage encapsulating the hydrophobic argon atom [6] (see Figure 1). If the NPT Monte Carlo method is applied to the bulk system, there is no problem with the definition of system volume, which is identified with the volume of the simulation cell. However, in the case of finite systems the situation is much more involved and several model approaches for the evaluation of the system volume are available. In this study, we compare results obtained from simulations of the two systems considered enclosed in a spherical hard-wall container with variable radius with alternative methods described by Baltazar [5]. Full temperature-pressure phase diagrams of those systems have been obtained by the multiple histogram method (see Figure 2). The water-water interactions are described empirically by the TIP4P ice pair potential, while for the argon-water interaction we have used empirical and ab initio potentials to see how different water-argon models affect the results of simulation.

In the isothermal-isobaric ensemble, the residual part of the mean value of dynamical parameter, F, is expressed at particular temperature T and pressure P as an ensemble average,

$$\langle F\rangle_{T,p} = \frac{\int_V \int_q F(q,V)\exp\left(-\frac{E^{\text{int}}(q)+PV}{k_B T}\right) \mathrm{d}q\mathrm{d}V}{\int_V \int_q \exp\left(-\frac{E^{\text{int}}(q)+PV}{k_B T}\right) \mathrm{d}q\mathrm{d}V}, \quad (1)$$

where q and V are system configuration and volume. Respectively, E^{int} denotes the interaction energy of the system and k_B is the Boltzmann constant. If the dynamical parameter depends on system configuration through the interaction energy only,

$$F = F\left(E^{\text{int}}(q), V\right), \quad (2)$$

the complicated high-dimensional integral on the r.h.s. of Eq. (1) can be replaced using the two-dimensional multiple histograms method [5] by computationally much cheaper two-dimensional (2D) integration,

$$\langle F\rangle_{T,p} = \frac{\int_V \int_{E^{\text{int}}} F(E^{\text{int}},V)\exp\left(-\frac{E^{\text{int}}+PV}{k_B T}\right) \Omega(E^{\text{int}},V)\,\mathrm{d}E^{\text{int}}\mathrm{d}V}{Z_{T,p}}, \quad (3)$$

where $\Omega(E^{int}, V)$ is the classical density of states and

$$Z_{T,p} = \int_V \int_{E^{\text{int}}} \exp\left(-\frac{E^{\text{int}}+PV}{k_B T}\right) \Omega(E^{\text{int}},V)\,\mathrm{d}E^{\text{int}}\mathrm{d}V \quad (4)$$

is the system configuration integral.

The parallel tempering (or replica exchange) method for isobaric-isothermal ensemble [6] means simulating hundreds of systems in parallel (about 30 different temperatures between 20 K and 1.000 K and 30 different pressures between 1 KPa and 10 GPa). This method is used for the computation of complicated high-dimensional integral on the r.h.s. of Eq. (1) and evaluation of two-dimensional energy-volume histograms. Such computed histograms are then used for computation of two-dimensional density of states $\Omega(E^{int},V)$. For the computation of dynamical parameters (in particular constant-pressure heat capacity C_p and Pearson coefficient $\rho(E^{int},V)$, which are used for creating a phase diagram of the simulated system). In an ideal case, each system is simulated on one core. Very low communications between cores (exchanges of system configurations) brings almost ideal scalability up to hundreds of cores (the upper limit is the number of simulated systems, in our case 900 computer cores). One Monte Carlo step of each system includes gradually translation, rotation and deformation movement of the each molecule. The average size of all random movement is set in order to about 50% of the proposed new configurations being accepted. A typical single simulation of a system containing 20 water molecules needs 10^7 Monte Carlo steps for each system; the first half of the Monte Carlo steps is discarded to let the simulated systems come to equilibrium. Such simulation

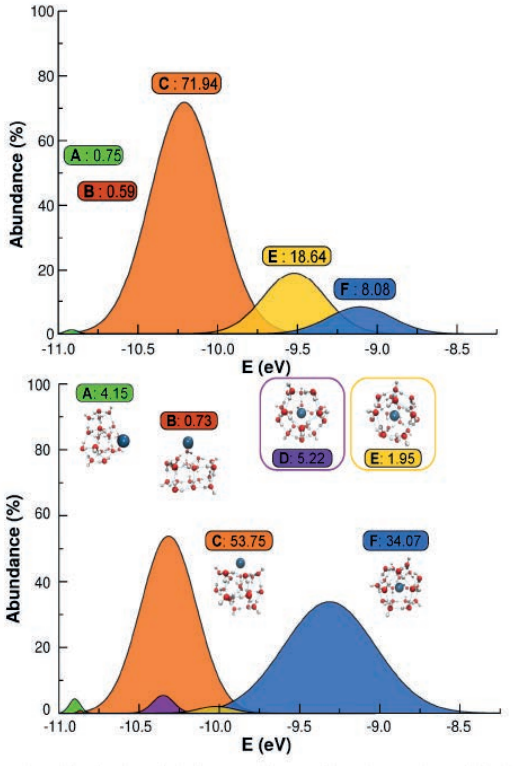


Figure 1.
3D representation of the charge density for grain boundary in cobalt

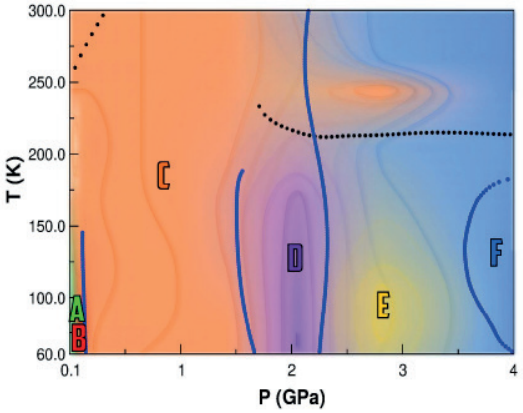


Figure 2.
Contour plot of the abundance of each isomeric structure for the $Kr(H_2O)_{20}$ cluster at PT -plane

means 40,000 core-hours (two days on 900 cores). The computation demands increase with the size of system with a power of 2.5. All simulations have been performed by the MCSIMUL code developed by the MOLDYN Group (see the http://moldyn.vsb.cz/web pages). Iterative solving of a set of equations to calculate the density of states $\Omega(E^{int},V)$ and numerical computation of a two-dimensional integral of r.h.s Eq. (4) is computationally cheap; it is paralleled via an Open MP approach and consumed a negligible amount of core-hours in comparison with the Monte Carlo simulation.

Conclusion

Further work in this direction employing specialised treatments for the weak host-guest interactions under pressure should be introduced, in order to estimate systematic errors and evaluate the accuracy of the interaction potential from the gas phase to extended state systems. Unfortunately, no experimental data are as yet available for Kr-water clusters of any size for direct comparison. Our future efforts should focus on extending the present simulations to larger size clusters, although theoretically still tractable. It is easy to calculate from scaling of the Monte Carlo method with the size given in the section Results and methods that simulation of, e.g., 36 water molecules consumed about 170,000 core-hours. Clusters of size 20, 24, 28 and 36 water molecules are interesting thanks to the fact that they may represent cavities in macroscopic gas hydrate systems. Such clusters correspond to the cavities of the sI, sII, sH, sO, and sT type structures of the Kr clathrate hydrates aiming, as a first step, to an indirect comparison with experimental observations available for these systems, and thus to a better quantitative description of the underlying interactions.

References

[1] Loveday, J.S., et al. Phys. Chem. Chem. Phys. 2008, 10, 937. Doi: 10.1039/B704740A
[2] Hartke, B. J. Chem. Phys. 2009, 130 art. no. 024905. Doi: 10.1063/1.3058479
[3] Vítek, A. and Kalus, R. Comp. Phys.Comm. 2014, 185, 1595. Doi: 10.1016/j.cpc.2014.02.024
[4] Martoňák R. et al. Comp. Mat. Science. 2006, 37, 526. Doi: 10.1016/j.commatsci.2005.12.028
[5] Okabe T., Kawata, M., Okamoto Y., and Mikami, M. Chem. Phys. Lett. 2001, 335, 435. Doi: 10.1016/S0009-2614(01)00055-0

Publications

[6] Vítek, A. et al. Computational investigations of the thermodynamic properties of size-selected water and Ar-water clusters: high-pressure transitions. Phys. Chem. Chem. Phys. 2015, 17(14), 8792-8801. Doi: 10.1039/c4cp04862h
[7] Arismendi-Arrieta, D. J., Vítek, Aleš and Prosmiti, Rita. High Pressure Structural Transitions in Kr Clathrate-Like Clusters. The Journal of Physical Chemistry C. 2016, 120(45), 26093-26102. Doi: 10.1021/acs.jpcc.6b07584
[8] Prosmiti, R., Vítek, A., Arismendi-Arrieta, D. J., Rodriguez-Cantano, R., Villarreal, P., Kalus, R., and Delgado-Barrio, G. Structural changes in water and Ar-water clusters under high pressure. Journal of Physics: Conference Series [online]. 2015, 635(3), 032008-. Doi: 10.1088/1742-6596/635/3/032008

06 | LIFE SCIENCES AND BIOLOGY

ESTIMATION OF EMPIRICAL DISTRIBUTION FUNCTION FOR TESTING OF MODALITY OF 3D PARTICLE SIZES BASED ON 2D OBSERVATIONS

Research institution:
Institute for Research and Applications of Fuzzy Modelling, University of Ostrava

Principal investigator:
Michal Burda

Project ID:
IT4I-1-9

Introduction

In medical research, adipose tissue particle size distribution is of great interest [1-3]. Concretely, the adipose tissue of morbid obese persons is subjected to chronic pressure for lipid storage. Adipose tissue mass can expand: (1) hypertrophy of existing mature fat cells, or (2) hyperplasia of adipocyte. Studies reported that large subcutaneous fat cells are associated with metabolic changes, whereas the adipose cell number does not concern cardio-metabolic health if these are of normal size [1].

In the other words, the presence of large particles (along smaller ones) may indicate cardio-metabolic health complications. Therefore, an analysis of whether the adipose tissue particles are all of the same size, or whether there are two different dominant sizes of them, is of great importance. This answer can be provided by using a statistical test for bimodality. Generally, a random variable is bimodal if its distribution shows two modes, i.e. two peaks in the probability density.

Unfortunately, the fat tissue is opaque, and as such the fat tissue particles cannot be directly observed. Therefore, a cross-section of the tissue is taken that only allows the view on 2D circular profiles of the original 3D spherical particles. Of course, the observed 2D particle sizes differ from the real 3D sphere sizes, because the cross-section does not usually lead through the centre of the particles. Hence, the problem of testing for particle size bimodality resides in estimating the 3D sphere radii from observed 2D section radii, and accepting or rejecting the unimodality of 3D radii distribution.

Results and Methods

The classic problem of stereology is to estimate information about a three-dimensional (3D) structure of materials from two-dimensional (2D) sections. Many different disciplines, such as medicine, biology, geology, metallurgy, and chemistry, need to estimate size distribution of spherical particles contained in a material. Because of opaqueness, only cross-sections can be observed and hence we have a problem of unfolding the sphere size distribution from a given sample of cross-section radii, which is known as Wicksell's corpuscle problem [4].

The best known solution to the unfolding problem is probably Saltykov's [5]: the raw 2D data are discretised into size classes along a set of intervals, from which a histogram can be plotted. Saltykov's method assumes the maximum observed radius to be equal to the true 3D radius. From that, a probabilistic approach is applied to deduce the true frequencies of size classes. An improvement of the basic idea by the use of kernel estimators can be found in [6], [7]. On the other hand, [8] extends the method to non-spherical shapes.

Testing for modality means determining whether a data sample comes from a single population or from a mixture of two different populations. A testable hypothesis is whether a mixture of two distributions describes data significantly better than a single distribution. A well-known test of unimodality is Hartigan's Dip test [9] that measures multimodality by maximum difference between the empirical distribution function and the uniform distribution (being the least favourable unimodal distribution). Frankland and Zumba, [10], [11], present a framework for fitting unimodal and bimodal distribution and testing their significance.

As mentioned, the particle size distribution can only be observed from cross-sections. Hence, the problem of testing for their bimodality resides in two tasks:

- estimating the 3D sphere radii from 2D section radii;
- accepting or rejecting unimodality of 3D radii distribution.

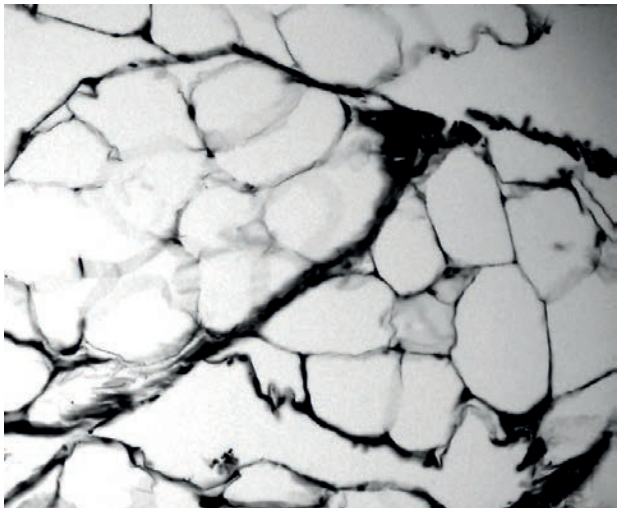


Figure 1.
The adipose tissue cross-section image as observed by microscope

Unfortunately, the existing methods do not enable us to perform these tasks separately, with results from the first steps being processed in the second step.
A novel method was recently developed to unfold 2D radii to 3D radii simultaneously with a test for modality by fitting both unimodal and bimodal distributions and performing a likelihood ratio test.
Although asymptotically chi-squared, the real likelihood ratio distribution is unknown for finite samples. A simulation is often performed to obtain the empirical distribution function instead. The aim of this project was to perform thorough simulations of the proposed method performance on randomly generated data. In the simulation, a cross-section over a randomly generated set of particles with known sizes is performed and the ability of the test to accept or reject the unimodality is evaluated.
The necessity to use high performance computing primarily originates from the fact that the proposed method is quite time-demanding and, to assess the proposed statistical test properties, a lot of simulations have to be performed with different initial settings.

On-going Research / Outlook

Thanks to the simulations on the Anselm high performance computer, a number of simulations were performed that enabled us to justify the correctness of the proposed method. Co-operating with Dr Vojtěch Kamarád from the Medical faculty of the University of Ostrava, the proposed method was used to analyse real data of patients with morbid obesity. The results of the analyses were used as inputs for future medical research.

Conclusion

A novel method for modality testing of 3D spherical particle sizes was developed so that only data about 2D cross-section circular radii are needed. The method was tested on randomly sampled data and some corrections were proposed in order to provide more accurate results. Also the power of the test was estimated for distribution parameters close to the values of practical medical application. The usefulness of the method was proved by a real-world medical application on adipose tissue cross-section data. The results were published in [12].

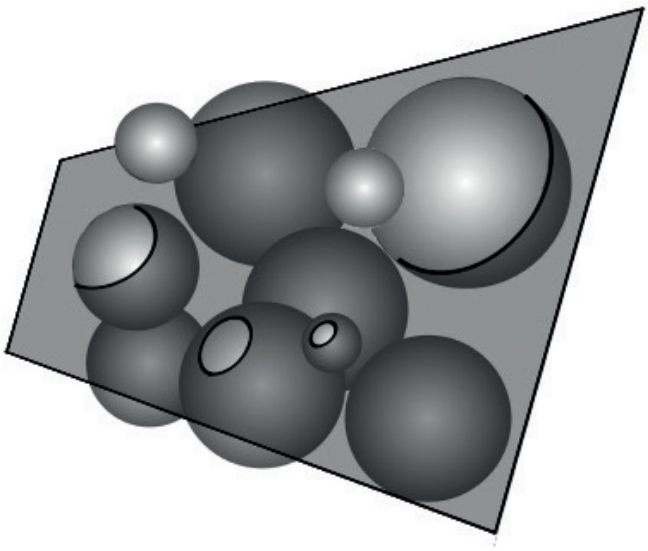


Figure 2.
A schematic view on a cross-section of spheres. As can be seen, it is usual that the cross-section does not usually lead through the particle centres. Therefore, the observed radii of circular profiles are smaller by a random amount than the radii of real particles.

References

[1] Després, J.-P. Abdominal obesity and cardiovascular disease: is inflammation the missing link? The Canadian Journal of Cardiology. 2012, 28(6), 642-652. Doi: 10.1016/j.cjca.2012.06.004
[2] Roullier, V., Cavaro-Mnard, C., Calmon, G. and C. Aub. Fuzzy algorithms: Application to adipose tissue quantification on MR images. Biomedical Signal Processing and Control. 2007, 2(3), 239-247. Doi: 10.1016/j.bspc.2007.07.005
[3] Ducluzeau, P.-H., Manchec-Poilblanc, P., Roullier, V., Cesbron, E., Lebigot, J., Bertrais, S. and C. Aub. Distribution of abdominal adipose tissue as a predictor of hepatic steatosis assessed by MRI. Clinical Radiology. 2010, 65(9), 695-700. Doi: 10.1016/j.crad.2010.03.013
[4] Wicksell, S. D. On the corpuscle problem. Biometrika. 1925, 17, 84-89
[5] Saltykov, S. A. Stereometric metallography. Moscow: Metallurgizdat, 1958.
[6] Weese, J., Korat, E., Maier, D. and J. Honerkamp. Unfolding sphere size distributions with a density estimator based on tikhonov regularization. J. Comput. Phys. Dec. 1997, 138(2), 331-353. Doi: 10.1006/jcph.1997.5822
[7] Jeppsson, J., Mannesson, K., Borgenstam, A. and J. Agren. Inverse Saltykov analysis for particle-size distributions and their time evolution. Acta Materialia. 2011, 59(3), 874-882. Doi: 10.1016/j.actamat.2010.09.046
[8] Sahagian, D. 3D particle size distributions from 2D observations: stereology for natural applications. Journal of Volcanology and Geothermal Research. Sept. 1998, 84(3-4), 173-196. Doi: 10.1016/S0377-0273(98)00043-2
[9] Hartigan, J. A. and P. M. Hartigan. The dip test of unimodality. The Annals of Statistics. 1985, 13(1), 70-84. Doi: 10.1214/aos/1176346577
[10] Frankland, B. W. and B. D. Zumbo. Quantifying bimodality part 1: An easily implemented method using SPSS. Journal of Modern Applied Statistical Methods. 2002, 1(1), 157-166. Doi: 10.22237/jmasm/1020255780
[11] Frankland, B. W. and B. D. Zumbo. Quantifying bimodality part 2: A likelihood ratio test for the comparison of a unimodal normal distribution and a bimodal mixture of two normal distributions. Journal of Modern Applied Statistical Methods. 2009, 8(1), 51-67.

Publication

[12] Burda, M., Kupka, J. and V. Kamarád. Testing for modality of 3D particle size distribution based on 2D observations. International conference on medical imaging using bio-inspired and soft computing (MIBISOC 2013). Asturias, Spain: European Centre for Soft Computing. 2013, 69-75. ISBN 978-84-695-7710-3

A NON-NEWTONIAN MODEL OF BLOOD CAPTURING SEGREGATION OF ERYTHROCYTES – THE FIRST STEP TO PROPER MODELLING OF THE BLOOD COAGULATION PROCESS

Research institution:
Charles University in Prague

Principal investigator:
Marek Čapek

Project ID:
OPEN-6-22

Introduction

The motivation for the project is the question of whether mathematical modelling can be of direct help in medical decisions when encountering certain situations in the cardiovascular system.

Among these situations of paramount importance are thrombus development because of atherosclerosis and unwanted coagulation on the surface of prosthetic devices such as artificial implants.

Understanding the blood coagulation process could therefore help either the pharmaceutical industry with the design of new anticoagulants or companies producing prosthetic devices with the design of these devices. The proper design of the implants will hinder blood from coagulation on the artificial surfaces of these implants. The profound insight into the process of blood coagulation could also help neurosurgeons to decide whether the operation of an aneurysm is necessary and safe, as the blood coagulation process can set off even in aneurysms where specific blood flow conditions are present.

Our aim is to develop a comprehensive but feasible model for coupled blood flow and blood coagulation. The model should be able to describe some of the important rheological and biochemical phenomena.

The model development and numerical implementation is based on the current work of M. Čapek in his PhD thesis. The computations based on the model and it's numerical implementation require significant computational resources as it leads to large systems of nonlinear equations.

Results and Methods

The most promising approaches treat blood as a mixture of species which are amenable to continuous treatment, i.e. blood plasma and red blood cells. This approach is reasonable as distribution of biochemical species, which occurs in blood

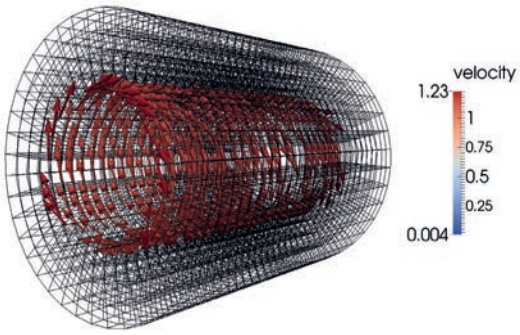


Figure 1.
Initial velocity Profile for the Couette viscometer

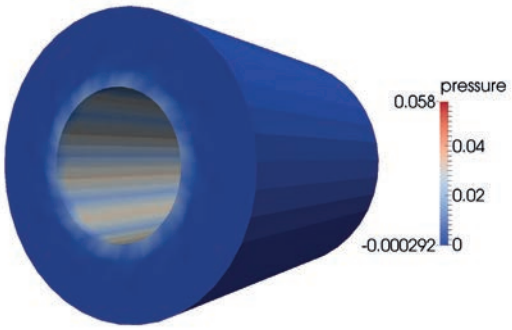


Figure 2.
Pressure field for the Couette Viscometer experiment

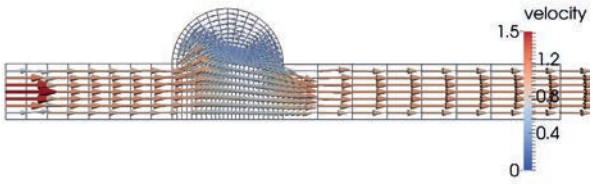


Figure 3.
Initial velocity profile in aneurysm

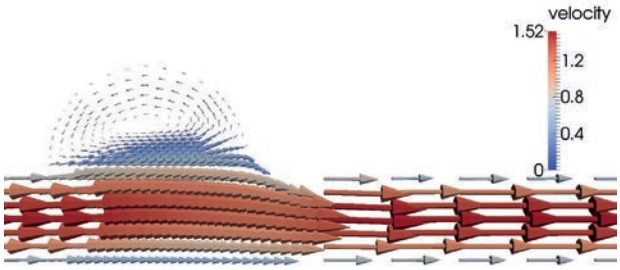


Figure 4.
Developed velocity profile in aneurysms

coagulation, is fundamentally determined by the concentration of red blood cells, i.e. haematocrit [5]. The work of Jung [3] has already laid the basis of this approach; however, it does not take account of the thermodynamic admissibility of proposed models, which was taken into account in the thermodynamical framework developed by Rajagopal [1]. This framework was already used mainly for single constituent modelling of blood flow [9], where red blood cells and blood plasma was only treated as a single continuum. The framework was already used for the mixture of two Newtonian fluids.

An alternative approach developed by Owens [2] tries to infer constitutive equations from micro-structure based observations. This approach has one advantage with respect to the previous models – the material parameters are easily derivable from properties of micro-structure. On the other hand this approach has the disadvantage that it leads to much greater computationally demanding discrete problems. In our computations we focus on this type of model.

Blood clotting is a result of delicate interplay between biochemical reactions and blood flow. We are mostly interested in models that couple the comprehensive model of coagulation biochemistry to models of platelet deposition and flow [8, 10]. Thermodynamic admissibility is taken into account in [9], however this model is not so complete as it does not account for binding sites on platelets, which is an important ingredient of the whole process as the binding sites are a very important place for biochemical reactions.

We have already implemented a limited part of blood coagulation in blood flow, wherein we have applied an approach of one version of the mixture theory based on the phase-field ideas. The blood clot, which composes mainly of fibrin network, is treated as a polymeric biofilm, like in the works on the polymeric theory of biofilms [11,12]. The phase-field has then a 'marker' function, which enables us to prescribe different constitutive equations for the blood and clot. The growth of a biofilm-like clot will be the result of tens of diffusion-convection-reaction equations coupled with blood flow. However, these equations are convectively dominated, which requires stabilisation. Discontinuous Galerkin appears as the best treatment of the equations, although this approach comes with the penalty of growing computational complexity. We started our experiments with phase-field modelling by rephrasing the successful simplified model of the blood coagulation process of Weller [13] into phase-field settings.

The deal.ii library [6] enables rapid code development retaining large flexibility for application developers. Deal.ii interfaces with common numerical packages, of which we will primarily be using PETSc. We used to assemble the whole Jacobian Newton iteration using the automatic differentiation tool Sacado from Trilinos. However, in the end we have chosen to perform the Picard semi-implicit iteration. Scalability of the deal.ii code is also ensured with the use of library P4est, which enables the distribution of large meshes across multiple processors. We will mainly use its support of scalable parallel applications using Open MPI. Deal.ii also enables hybrid parallel programming, i.e. message passing in connection with multithreading (using threading building blocks implementation of threads). MPI is accessed by so called wrappers around PETSc functionality, while access to MPI through Trilinos is done in a similar way. As our systems are large we implement matrix-free algorithms in deal.ii, which enable parallelisation on three levels – distributed, shared and vectorisation [4]. We therefore limit the memory requirements and thus avoid the possibility of memory bottleneck. Another branch of our research follows the work of Timo Heister [14], wherein a scalable Navier-Stokes solver is set up using a block-triangular preconditioner for the FGMRES solver. We have already rewritten our code from the distributed direct solver of SuperLU_Dist from Trilinos to distributed FGMRES (implemented in deal.ii) with the preconditioner from [14], however, finer tuning of the solver is needed in order to achieve reasonable parallel efficiency. We experiment with projection methods for solution of the Cahn-Hilliard-Navier-Stokes system [15]. We will use either ILU decomposition or algebraic multigrid as the preconditioner for the convection-diffusion-reaction equations of our system.

On-going Research / Outlook

We are about to integrate our rheological model of blood flow inspired by [2] with a proper model of the 'chemical' part of the blood coagulation process. We want to try the new framework for solving Cahn-Hilliard-Navier-Stokes presented in [16] and compare it with the splitting method of [15].

We perform simulations based on [13] in 3D, and we are currently recoding the implementation of our modification to Owens' model [2] into 3D as well. For fitting the experimental data we use parallel Matlab provided at IT4Innovations, which is a reasonable trade-off between the speed of coding and the time of computations.

Conclusion

We are just at the beginning of our endeavours to simulate such a complicated process as blood coagulation. We cooperate actively with specialists from the field of haematology. The biggest challenge lies in the necessity to bridge the gaps between various scientific disciplines.

The possibility of performing demanding computations at IT4Innovations speeds up iterations of proposed models, i.e. to frequently modify the constitutive equations and parameters and experiment with them via computer simulations.

We hope that the achieved results will not only be useful from the theoretical point of view. We have mentioned the possibilities for the application of our model in medicine in the Introduction.

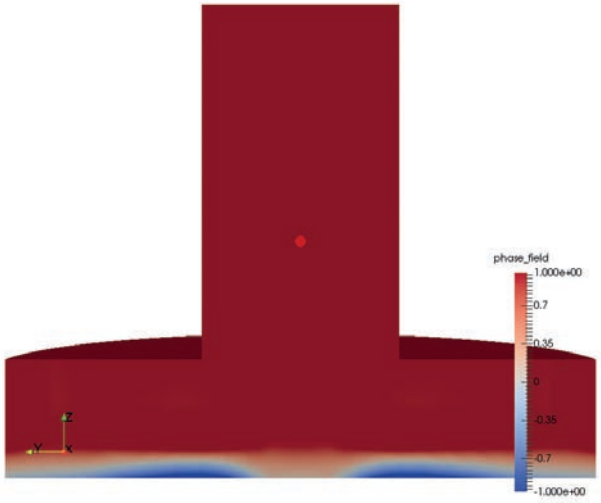


Figure 5.
3D simulation of Weller's model in the phase-field setting

References

[1] Rajagopal, K.R. and A.R. Srinivasa. A thermodynamic frame work for rate type fluid models. J. Non-Newtonian Fluid Mechanics. 2000, vol. 88. Doi: 10.1016/S0377-0257(99)00023-3

[2] Owens, R. G. A new microstructure-based constitutive model for human blood. Journal of Non-Newtonian Fluid Mech. 2006, 140, 57-70. Doi: 10.1016/j.jnnfm.2006.01.015

[3] Jung, J., Hassanein, A. and R.W. Lyczkowski. Hemodynamic computation using multiphase flow dynamics in a right coronary artery. Ann Biomed Eng. 2006, 34(3), 393-407. Doi: 10.1007/s10439-005-9017-0

[4] Kronblicher M., Kormann K. A generic interface for parallel cell-based finite element operator application. Computers and Fluids. 2012, 63, 135-147. Doi: 10.1016/j.compfluid.2012.04.012

[5] Aarts P., van den Broek S. A., Prins G., Kuiken G., Sixm J. Heethaar. Blood platelets are concentrated near the wall and red blood cells, in the center in flowing blood. Arteriosclerosis. Nov-Dec 1988, 8(6), 819-24. Doi: 10.1161/01.ATV.8.6.819

[6] Bangerth, W., Hartmann, R. and G. Kanschat. Deal.II – a general-purpose object-oriented finite element library. ACM Transactions on Mathematical Software. 2007, 33(4), article no. 24. Doi 10.1145/1268776.1268779

[7] Burstedde, C., Wilcox, Lucas C. and Omar Ghattas. p4est: Scalable Algorithms for Parallel Adaptive Mesh Refinement on Forests of Octrees. Published in SIAM Journal on Scientific Computing. 2011, 33(3), 1103-1133. Doi: 10.1137/100791634

[8] Kuharsky, A. L., Fogelson, A. L. Surface-mediated control of blood coagulation: the role of binding site densities and platelet deposition. Biophys J. Mar 2001, 80(3), 1050-1074. Doi: 10.1016/S0006-3495(01)76085-7

[9] Mohan, A. Modeling the growth and dissolution of clots in flowing blood. Dissertation thesis, Texas A&M University, 2005.

[10] Gregg, K.L. A mathematical model of blood coagulation and platelet deposition under flow. Dissertation thesis, University of Utah, 2010.

[11] Zhang, T., Cogan, N. Phase-Field Models for Biofilms II. 2-D Numerical Simulations of Biofilm-Flow Interaction. Communications in Computational Physics. 2008, 4, 72-101.

[12] Lindley, B., Wang, Q., Zhang, T. Multicomponent hydrodynamic model for heterogeneous biofilms: Two-dimensional numerical simulations of growth and interaction with flows. Phys. Rev. E. 2012, 85, 031908. Doi: 10.1103/PhysRevE.85.031908

[13] Weller, F. A free boundary problem modeling thrombus growth: Model development and numerical simulation using the level set method. Math. Biol. 2010, 61, 805-818. Doi: 10.1007/s00285-009-0324-1

[14] Heister, T. A Massively Parallel Finite Element Framework with Application to Incompressible Flows. Doctoral thesis. University of Gottingen, 2011.

[15] Guillén-Gonzáles, F., Tierra, G. Splitting schemes for a Navier-Stokes-Cahn-Hilliard model for two fluids with different densities. Journal of Computational Mathematics. 2014, 32(6), 643-664. Doi: 10.4208/jcm.1405-m4410

[16] Bosch, J., Kahle, Ch., Stoll, M. Preconditioning of a coupled Cahn-Hilliard-Navier-Stokes system. arXiv:1610.03991 math.NA

FEM MODEL OF THE SOFT TISSUES – COMPARISON WITH ULTRASOUND ELASTOGRAPHY

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Jan Grepl

Researchers:
Petr Horyl,
Marek Penhaker

Project ID:
IT4I-9-19

Introduction

In the entire history of medicine, we can observe how objective measurement of various parameters of the human body is important. Over the course of time and with advances in technology, various types of instruments and complex diagnostic machines were invented. However, a great number of parameters, where correct measurement would significantly lead to effecting the better prediction and subsequent evaluation of diseases, still exists [1].

One of the areas where an objective approach to measurement has not been clearly established is in the measurement of soft tissues like muscles, skin, organs, etc. Meanwhile, a relative stiffness comparison of different soft tissue areas is widely used every day by doctors as one of the basic diagnostic techniques. At present, a number of diseases exist where earlier diagnosis leads to significantly better and faster recovery. Nevertheless, the majority of patients get the detailed diagnostic examination with advance techniques in more advanced stages of disease. Our work addresses the soft tissue elastic modulus and tissue stiffness primarily due to the indication of serious diseases such as breast cancer and compartment syndrome.

Results and Methods

One of the methods used to measure the elastic properties of various soft tissues is ultrasound elastography. Elastography is a non-invasive technique based on a diagnostic ultrasound or magnetic resonance imaging device [2]. The principle is in representation of the elastic properties of the soft tissues. By this method it's possible to obtain additional information in the form of elastic modulus.

One of the possibilities to evaluate the soft tissue is by the elastic modulus (Young modulus). This chapter deals with numerical determination using the finite element method (FEM) of the soft tissue elastic modulus from results of already performed static experimental measurement [3]. The data needed as input for the FEM model can be obtained in many ways, for example, by using (nano) indentation. It is of most importance to obtain information about the depth of object indentation (distance from surface), the corresponding amount of force to achieve that depth, and lastly the maximum value of tissue depth (distance from surface to bone borderline). For evaluation of this model, the data from a 24-point non-invasive indentation measurement of the upper limb has been used.

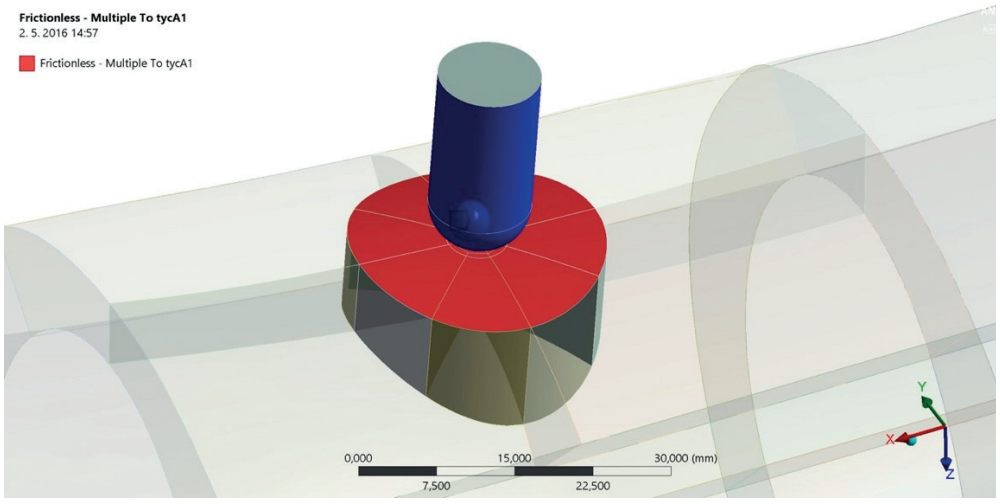


Figure 1. The accurate 3D geometrical model for FEM simulations

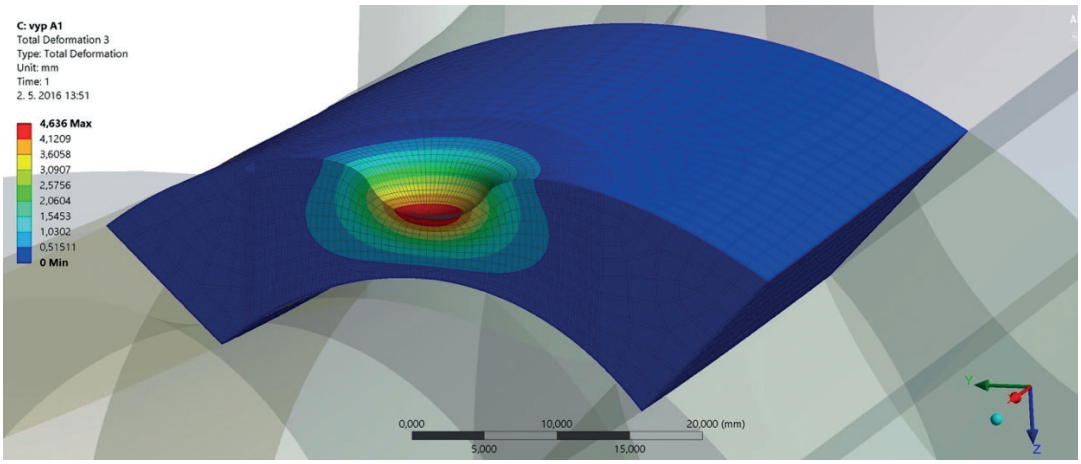


Figure 2.
The Deformed shape after one FEM simulation.

In briefly, the principle of FEM simulations is in replication of conditions of the static experimental indentation measurement while the known result data is used to tune the computational model. Knowing the fact that the elastic modulus is directly dependent on the numerical values obtained by measurement, a model created as such can be used for reverse calculation of elastic modulus [4]. All FEM simulations were performed in ANSYS Mechanical.

For the initial simulation an accurate 3D geometrical model of the upper limb has been created. After numerous simulations it was proven that only smaller parts of the geometry for the simulation needed consideration – simplified general geometry is possible to use with sufficient accuracy (Figure 1).

For the evaluation of the value of elastic modulus, the principle of parametric simulations for reverse calculation has been used. The value of elastic modulus in material properties for each sub region has been set as a variable parameter. The "direct optimisation adaptive single objective" condition has then been applied. The goal for this condition was set to find the elastic

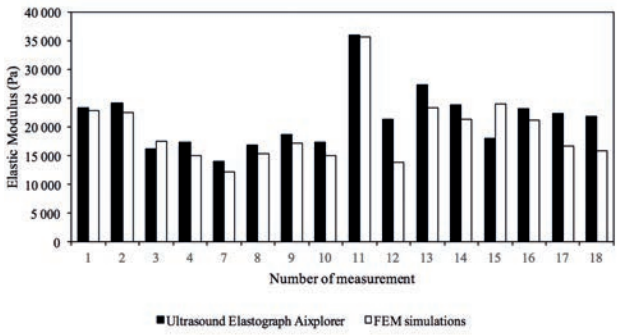


Figure 4.
Comparison of the obtained elastic modulus values from Ultrasound Elastograph Aixplorer with FEM simulations.

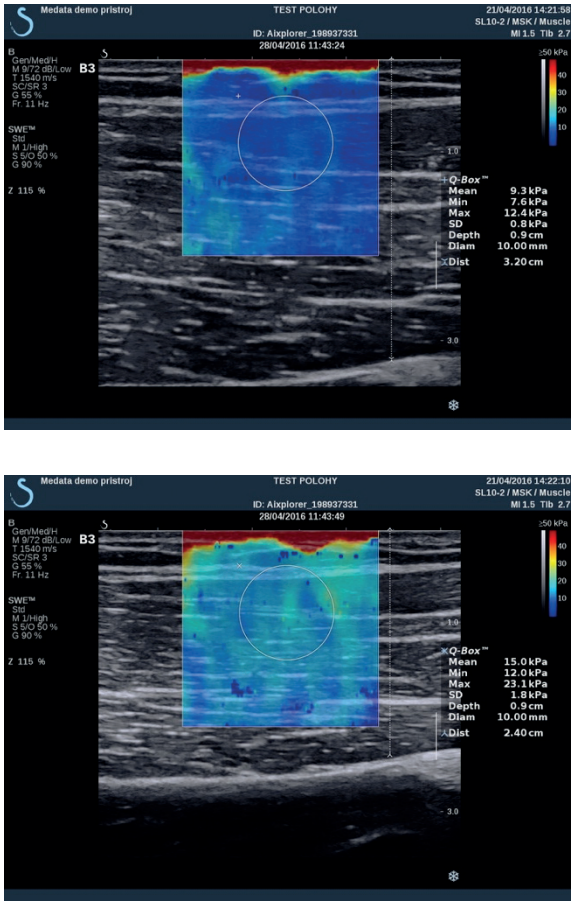


Figure 3.
The Comparison of the SWE images before (top) and after compression of the tissue (bottom).

modulus value such that the resulting displacement in the sub region corresponds to the value obtained from experimental measurement. The dependent value of force has been used as a load boundary condition.
Considering the evaluation of the 24-point experimental measurement data, hundreds of repetitive simulations were necessary to run. For this reason, the HPC Salomon of IT4Innovations Infrastructure has been used.
For evaluation of FEM simulations results, another method of elastic modulus measurement was needed. The SuperSonic Imagine from the latest generation of ultrasound elastograph Aixplorer, kindly provided by MEDATA, Ltd. has been used for this. This device uses SherWave Elastography technology (SWE) to generate shear wave pulses in the body with transitional pulses. The elasticity of the tissue is then directly expressed by measuring the velocity of wave propagation according to the following equation:

$$E = 3 \cdot \rho \cdot c^2 \qquad (1)$$

Where E is elastic modulus, ρ is tissue density and c is propagation velocity of shear waves.
For the possibility to compare the results, it was necessary to perform the elastographic examination under the same conditions as the initial experiment. Firstly, the reference point for depth measurement in the standard ultrasonic B-image has been chosen. The clearly visible bone borderline has been considered as this. The B-image was then overlaid by a coloured SWE map, showing the distribution of tissue elasticity by elastic modulus calculation and visualisation in real time.
The image of subsurface tissue in both a relaxed and compressed state was then obtained (Figure 2). The amount of compression was set in such a way to obtain the same depth as in the initial experimental indentation measurement.
To more precisely verify whether the result values by FEM simulations are comparable, the larger experimental verification measurement was realised. That consists of measurement by both methods on a select group of 8 volunteers in two selected sub-areas of the upper limb. That results in 16 values for comparison.
From the results comparison in Figure 3 we can state, that the results of both methods for obtaining elastic modulus are comparable. Although the largest difference in measurement no. 12 with different value of 35% can be recognised from Figure 4, the average value of difference reached the value of 14%.

On-going Research / Outlook

The next steps in the research will be in performing a greater number of measurements. A deeper focus on the optimisation and refinement of the FEM model and approximation techniques is also required. Another way might be in analysis of the Poisson number effect to the resulting values. All future changes to the structure itself, and in optimising the simulations can lead to a more accurate determination of the evaluation parameter, but the proposed principle of the evaluation process described in this work will be retained.

Conclusion

The described method for soft tissue evaluation by elastic modulus is based on two steps. The first step is to perform (or obtain results) of non-invasive indentation measurement. Indentation must be performed from the surface of the tissue to the measured depth by subsequent force value recording. The second step is in determining of the predicted value of elastic modulus by a number of repeated FEM simulations. Using HPC to run the simulations and based on a large number of runs, it is also possible to determine an approximate function to directly estimate the value of elastic modulus only from the measurement data.
In the comparison of FEM simulation results with only the widely used non-invasive method – ultrasound elastography, the 14% difference of elastic modulus mean value in the range from 14 kPa to 36 kPa was found. Considering the simplicity of data gathering, the non-invasivity of the proposed method, considerable financial savings of measurement and current stage of the project, these results can be considered as acceptable.

References

[1] Baura, G. D. System theory and practical applications of biomedical signals. Hoboken, NJ: Wiley-Interscience: USA, 2002, 431 p. ISBN 0-471-23653-5
[2] Choi, J.S.; Han, B.-K.; Ko, E.Y.; Ko, E.S.; Shin, J.H.; Kim, G.R. Additional diagnostic value of shear-wave elastography and color Doppler US for evaluation of breast non-mass lesions detected at B-mode US. European Radiology. 2016, 26(10), 3542-3549. ISSN 0938-7994
[3] Han, L.; Burcher, M.; Noble, J.A. Non-invasive measurement of biomechanical properties of in vivo soft tissues. Lecture Notes in Computer Science. 2002, 2488, 208-215. ISBN 978-3-540-44224-0
[4] Liu, Yi; Amy E. Kerdok, Robert D. Howe.: A Nonlinear Finite Element Model of Soft Tissue Indentation (2004) Medical Simulation, International Symposium, ISMS 2004, Cambridge, MA, USA. June 17-18, 2004, 67-76. ISBN 978-3-540-22186-9

Publication

[1] Grepl, J.; Penhaker, M.; Horyl, P. The Use of High-Performance Computing for Reverse Calculation of Soft Tissue Elastic Modulus by Noninvasive Measurement Data and Finite Element Method. In Springer series: Lecture Notes in Artificial Intelligence LNCS/LNAI 2017. In preparation.

CONVERGENCE TESTING OF A K-SPACE PSEUDOSPECTRAL SCHEME FOR TRANSCRANIAL TIME-REVERSAL FOCUSING

Research institution:
Brno University of Technology

Principal investigators:
Jiří Jaroš, Bradley E. Treeby

Researcher:
James Robertson

Project partner:
University College London

Project ID:
OPEN-6-11, OPEN-7-15

Introduction

Ultrasound can be applied directly to the brain through the skull for non-invasive therapeutic applications. High intensity focus ultrasound was recently FDA approved for the treatment of essential tremor by ablation of the thalamus [1]. Meanwhile, novel low intensity applications such as neurostimulation and opening the blood brain barrier (BBB) with ultrasound are under active research [2],[3]. However, the skull forms a major obstacle to the effective transcranial transmission of ultrasound. The high sound speed, density and acoustic absorption of bone relative to soft tissue leads to aberration and attenuation of incident waves. This prevents effective focusing and can lead to heating of the skull and damage to surrounding tissue [4].

Time-reversal focusing is a non-invasive method of focusing ultrasound through the skull. It uses simulations of ultrasound propagation from a target in the brain to a virtual ultrasound array. The signals recorded at the transducer surface are then time-reversed, and used to derive drive signals for transcranial ultrasonic therapy [4]. The effectiveness of simulated time-reversal is directly related to the accuracy of the numerical scheme used. This can be affected by a number of factors, the most fundamental being the numerical accuracy of the scheme used to solve the acoustic wave equation.

Clinical software often makes use of ray tracing algorithms to calculate phase shifts. However, more advanced full-wave models have been used and shown to give improved performance [5]. These schemes are numerically convergent, meaning that desired accuracy can be achieved with sufficiently fine spatial and/or temporal sampling, dependent on the scheme.

In the present work, convergence testing was carried out on a model of simulated time-reversal using a model of the skull and a hemispherical transducer. The spatial and temporal sampling criteria required for different errors in refocusing quality to converge were measured. These results will help ensure simulation accuracy in future studies of numerical ultrasound propagation.

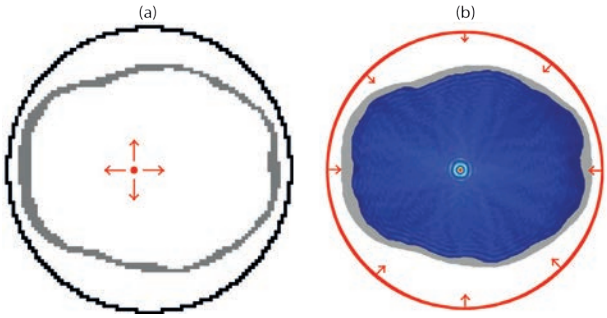


Figure 1.
Convergence testing setup. (a) Forward simulation at low spatial sampling (b) Reversal simulation at high spatial sampling. Source terms are shown in red.

Results and Methods

Numerical convergence testing was carried out for a model of transcranial propagation from a hemispherical transducer, through the skull to a deep brain target. Simulations modelling the propagation of ultrasound from the deep brain to a hemispherical virtual transducer were carried out using variable spatial and temporal discretisation, measured in points per wavelength (PPW). The simulation layout is shown in Figure 1.

Simulations were carried out using the OpenMP version of the open source k-Wave toolbox [6]. The toolbox includes a pseudospectral time-domain (PSTD) scheme with optional correction for the time-domain simulation of acoustic fields. This PSTD method approximates the spatially varying acoustic variables via discrete Fourier transform. This allows the efficient and accurate computation of spatial derivatives.

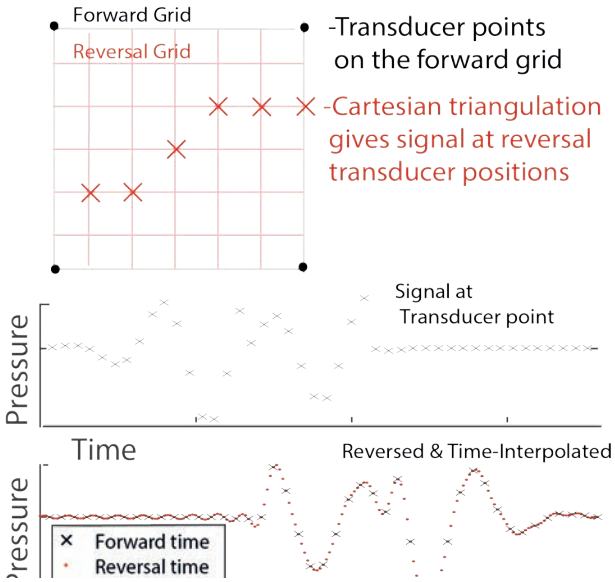


Figure 2.
Spatial and temporal interpolation of time-varying signals as part of the time-reversal process.

The field is then updated in a time stepping manner. However, as with most numerical methods, this approach can cause discretisation errors related to numerical dispersion, misrepresentation of medium discontinuities, and staircasing of medium geometries.

In each simulation the skull was mapped as a homogeneous layer based on a parametric mesh derived from a segmented MRI image. A parametric mesh was used to ensure consistency between schemes with different spatial sampling. The skull was assigned a density of 1,990 kg/m^3 and a sound speed of 3,200 m/s. The rest of the simulated domain was assigned the acoustic properties of water.

The signals recorded at the transducer surface were spatially and temporally interpolated, time reversed, and then used as the source term for high resolution reversal simulations. The interpolation step is outlined in Figure 2. Due to the large size of the simulation time-series data (~100 GB) and the memory required to interpolate this data in space and time, the interpolation step was carried out on the UCL Legion computing cluster, on a Dell 2U R820 32 core system with 1.5 TB of RAM.

The reversal simulations were then carried out, and the peak positive pressure field was recorded across the brain volume to evaluate focusing effectiveness. The metrics assessed were the peak spatial pressure amplitude and position, and the -3 dB focal volume. Simulations were carried out on the IT4Innovations Salomon supercomputing cluster. Each simulation was carried out on one node equipped with an 8-core Intel Xeon E5-4627v2@ 3.3GHz CPU, and 256GB of RAM. The largest simulations, used in the reversal step, had a domain size of 1,024^3 grid points and 22,718 time steps, with a total runtime of 112.3 hours. Checkpoint-restart was used to enable the running of sufficiently long simulations. Forward simulations that could not be carried out on a workstation were also run with these computation resources. In total, approximately 105,000 core-hours were consumed.

The resulting dependence of refocusing quality on the spatial sampling of the forward simulations is shown in Figure 3. It demonstrates that > 10 PPW are required to ensure convergence across all metrics, especially for peak pressure reconstruction.

However both -3 dB focal volume, and the position of the peak pressure are less dependent on spatial sampling, with positioning having no dependence. This means that for applications where fine control over the pressure amplitude is less important, coarser forward simulations may suffice.

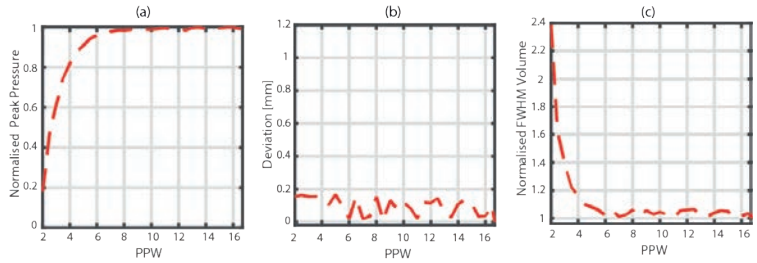


Figure 3.
Results of convergence testing in 3D.
(a) Convergence rate of peak pressure across the brain
(b) Convergence of the spatial position of the peak pressure
(c) Convergence of the -3 DB volume of the focus.

On-going Research / Outlook

This work forms part of a larger body of work relating to accurate simulation of transcranial ultrasound propagation for both time-reversal focusing and the prediction of intracranial fields. The developing applications of ultrasonic neuro-stimulation and opening the BBB are likely to require fine control of ultrasound at the target. Now that the spatial and temporal sampling criteria required to ensure numerical accuracy have been established, further simulation studies can be carried out to examine other sources of error. The next potential source of error arises from the mapping of medium properties. Future work will focus on determining the relative impact of different errors in medium geometry and acoustic properties, and the impact of modelling internal skull heterogeneities in more detail.

Conclusion

The results indicated that 10 spatial points per wavelength are required to ensure numerical convergence when using the k-space corrected PSTD scheme. The field parameter most affected by the numerical accuracy of the forward simulation is reconstruction of the peak pressure at the target, followed by the -3 dB focal volume. Spatial targeting was not significantly affected, indicating that coarse spatial sampling may suffice for certain applications. These results are of immediate relevance for anyone attempting to simulate the transcranial propagation of ultrasound. Confidence in the numerical accuracy of simulations will allow methods for the effective transcranial focusing of ultrasound to advance. Accurate prediction of intracranial fields will also be achievable with sufficient numerical accuracy, which will allow poorly understood phenomena such as neurostimulation to develop.

References

[1] Piper R. J.; M. A. Hughes, C. M. Moran, and J. Kandasamy, “Focused ultrasound as a non-invasive intervention for neurological disease: a review,” Br. J. Neurosurg., vol. 8697, no. April, pp. 1-8, Apr. 2016. Doi: 10.3109/02688697.2016.1173189

[2] Alonso A.; “Ultrasound-induced blood-brain barrier opening for drug delivery,” Front. Neurol. Neurosci., vol. 36, pp. 106-115, 2015. Doi: 10.1159/000366242

[3] Ye P. P.; J. R. Brown, and K. B. Pauly, “Frequency Dependence of Ultrasound Neurostimulation in the Mouse Brain,” Ultrasound Med. Biol., pp. 14-16, Apr. 2016. Doi: 10.1016/j.ultrasmedbio.2016.02.012

[4] Marquet F.; M. Pernot, J.-F. Aubry, G. Montaldo, L. Marsac, M. Tanter, and M. Fink, “Non-invasive transcranial ultrasound therapy based on a 3D CT scan: protocol validation and in vitro results,” Phys. Med. Biol., vol. 54, no. 9, pp. 2597-2613, May 2009. Doi: 10.1088/0031-9155/54/9/001

[5] Jones R. M. and K. Hynynen, “Comparison of analytical and numerical approaches for CT-based aberration correction in transcranial passive acoustic imaging,” Phys. Med. Biol., vol. 61, no. 1, pp. 23-36, 2016. Doi: 10.1088/0031-9155/61/1/23

[6] Treeby B. E.; J. Jaros, A. P. Rendell, and B. T. Cox, “Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method," J. Acoust. Soc. Am., vol. 131, no. 6, pp. 4324-4336, 2012. Doi: 10.1121/1.4712021

Publications

[1] Robertson, J. L., Cox, B. T. and B. E. Treeby. Quantifying Numerical errors in the simulation of transcranial ultrasound using pseudospectral methods. In: IEEE International Ultrasonics Symposium. IEEE, 2014, pp. 2000-2003. Doi: 10.1109/ULTSYM.2014.0498

[2] Robertson, J. L., Cox, B. T., Jaros, J. and B. E. Treeby. Simulation of transcranial ultrasound propagation for ultrasonic neuromodulation and stimulation. Submitted, 2016.

STRUCTURAL INVESTIGATIONS OF THE HUMAN MITOCHONDRIAL LON PROTEASE

Research institution:
Charles University in Prague

Principal investigator:
Lubomír Kováčik

Researcher:
Sami Kereïche

Project ID:
OPEN-5-15

Introduction

Mitochondria are important semiautonomous organelles of eukaryotic cells whose main task is to provide cells with energy. In mitochondria, there are more than 1,000 proteins, which participate in a control system that maintains mitochondrial homeostasis and harmonises levels of other proteins responsible for their proper function. The human Lon protease (hLon) participates in this system, as it primarily degrades misfolded, oxidatively modified and regulatory proteins [1]. Apart from this, it also binds mitochondrial DNA (mtDNA), by which it regulates its expression [2], and it is also known to act as a chaperone assisting the proper assembly of protein complexes [3]. Its over- or under-expression is linked to serious diseases like epilepsy, myopathy, hereditary spastic paraplegia, and cancer [4], and it has been recently recognised as a potential target for development of anti-cancer drugs [5]. With our IT4Innovations computations, we set out to determine the three-dimensional structure (3D) of this important enzyme in its full length from datasets acquired by cryo-electron microscopy (cryo-EM). The IT4Innovations computations involve Bayesian two- and three-dimensional image processing of a few hundred thousand captured projections of the hLon protein, which results in the determination of their spatial orientations and computation of the three-dimensional structure of the hLon protein.

In short, the Lon protease family (MEROPS [6] clan SJ, family S16) is a widespread family of ATP dependent proteases [7] present in living species of all phylogenetic kingdoms. Lon proteases belong to the AAA^+ protein superfamily (ATPases associated with diverse cellular activities [8]), which is a large and functionally diverse family of proteins united by common motifs in their amino-acid sequences. Many of the resolved structures of AAA^+ proteins have revealed that a specific protein activity is assigned to a structurally distinguishable compartment, e.g. [9-13]. In Lons however, the three N-, AAA^+-, and proteolytic domains are mapped to a single polypeptide chain, which distinguishes them to other AAA^+ proteases with the exception of bacterial FtsH and m-AAA proteases [14-16]. This arrangement implies that the Lon protease is a highly flexible protein. In addition, enzymatic and substrate binding studies have revealed that Lon functions are very likely based on the changes of its allosteric conformation [17-20]. These hints indicate that it had likely been the intrinsic flexibility of Lon that had prevented the determination of its structure and that still hinders the acquisition of a high-resolution structure (i.e. resolution < 9 Å).

The difficulty of Lon's structural determination has been demonstrated in previous electron microscopic studies of Lon proteases [21-23], which failed to determine at least its oligomeric state, as well as by the inability to crystallise the Lon protease of any species in full length. Only partial x-ray crystallographic structures of LonA type (the human mitochondrial Lon type) are known, of which the two of B. subtilis (PDB ID [24]: 3M6A, 3M65 [25] and the one of the N-terminal domain of E. coli (PDB ID: 3LJC [26]) provide an almost complete structural information of bacterial Lon, leaving only an ~40 amino acids long gap. However, the oligomeric status and the three-dimensional quaternary arrangement of the Lon protease remained unknown.

Therefore, we decided to use cryo-electron microscopy joined by advanced image processing methods to study the structure of Lon. In cryo-EM, specimens do not need to be crystallised. Instead, suspension containing the protein of interest is placed on a specimen grid and quickly frozen [27], so that it cannot undergo structural changes. Therefore, images acquired with cryo-electron microscopes render the native structure of the protein, which can be "caught" frozen in different functional states. The current available image processing methods are capable of recognition of the structural changes among functional states.

With the help of IT4Innovations computations, we have managed to determine the 3D structure of the human mitochondrial Lon protease in its full length. More specifically, we used its proteolytically inactive form and incubated it with cofactors ADP and AMP-PNP, and acquired its 3D structure with resolutions of 21 and 15Å, respectively [28]. In addition, we attempted to determine the structure of a hLon form missing the first 156 amino acids (hLonΔ270) in order to explain the observed dramatic reduction of its enzymatic activities.

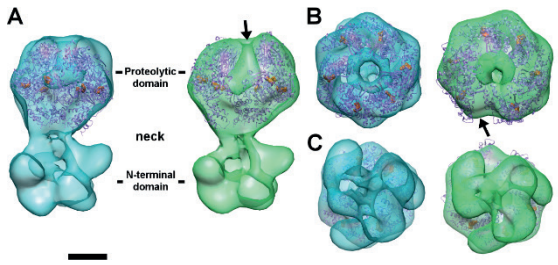


Figure 1.
Reconstructed cryo-electron microscopic structures of the proteolytically inactive human mitochondrial Lon protease, fitted with known crystal structure of the proteolytic and ATP-binding domains of B. subtilis (violet ribbons) whose amino acid sequences are highly identical to the corresponding portion of the amino acid sequence of the human Lon. Blue: Lon incubated with the ATP-analogue AMP-PNP, green: Lon incubated with ADP. ADP molecules, which are present in the crystal structure of B. Subtilis, are highlighted in red. A: side views, B: top views, C: bottom views. Arrows point at the opening in the ADP-incubated Lon structure. Scale bar: 5 nm.

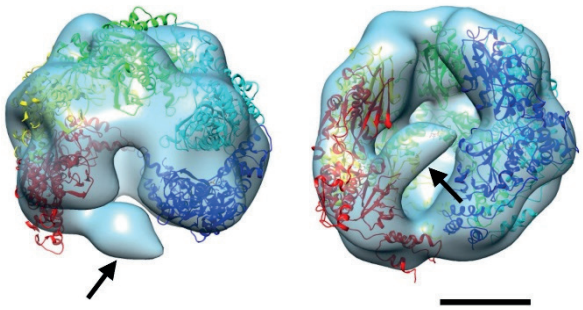


Figure 2.
Reconstructed cryo-electron microscopic structure of the mutant form of human mitochondrial Lon protease lacking the first 156 amino acids, fitted with five copies of the known crystal structure of the proteolytic and ATP-binding domains of Bacillus Subtilis (rainbow ribbons). Left: top view, B: side view. Arrows point at an extra mass reconstructed at the position of the expected sixth subunit. Its presence suggests that this subunit is actually present but that its position is unstable. Scale bar: 5 nm.

Results and Methods

Electron microscopy

In order to achieve the goal of the 3D reconstruction of the Lon protease, we acquired several large datasets of projections of the described hLon mutant forms with the state-of-the-art transmission electron microscopes (TEM) equipped with sensitive direct-detection cameras at Ceitec, Brno, and IGBMC in Strasbourg, France. Briefly, we used magnifications of ~50,000× yielding a pixel size of ~2Å and defocus ranging from -1.5 µm to -4 µ, the cameras were operated in the movie mode allowing frame alignments of fractionated exposure, and the TEMs were operated automatically.

Image processing

We pre-processed the datasets with the computers available in our institute, which involved GPU determination of frame shifts (frame alignment) and CTF (contrast-transfer function) of the aligned images, as well as semi-automated particle selection. Having prepared the particles, we transferred them to the IT4Innovations clusters and ran 2D and 3D Bayesian image analysis in the Relion software [29], which utilises parallel CPU processing. We typically used 20 nodes with the full number of CPUs and 1 thread because the amount of RAM per CPU was sufficient (~5 GB/CPU on the Salomon cluster).

Acquired structures

We realised that Lon is a hexamer with a unique structure, in which the proteolytic domain is hexameric, whereas its N-terminal domain is arranged as a trimer of dimers (Figure 1). Interestingly, the globular parts of the N-terminal domains were stacked in two three-tiered rings, which connected to the hexameric proteolytic chamber by a tight trimeric neck very likely formed by a coiled-coil pattern (i.e. by two adjacent and interacting helices). However, due to insufficient resolution of our reconstructed structures, we were unable to evaluate this assumption. In the presence of the AMP-PNP cofactor, the proteolytic chamber appeared as an almost perfectly six-fold closed-ring that was closed in the mouth of the neck, whereas in the presence of ADP, the ring and a channel in the neck opened up. This ATP-dependent structural change was accompanied by movements of the rest of the structure as well. We were able to fit six B. subtilis 3M6A crystallographic structures into the reconstructed maps without clashes, as well as six E. coli 3LJC N-terminal domain structures into the N-terminal part of our reconstructed maps. Unfortunately, due to insufficient resolution, we were unable to track the six monomers through the tight neck.

The reconstructed structure of the shortened Lon mutant hLonΔ270 (Figure 2) is difficult to interpret. We could fit only five 3M6A monomers into it, while the position of the sixth one was only indicated by a central reconstructed density. Moreover, we could not identify any density that could be assigned to the trimeric neck or to the remainder of the N-terminal domain. Therefore, we concluded that the loss of the enzymatic activities of this mutant is correlated to its structural disorder, i.e. that Lon's N-terminal domain is essential for all its activities.

On-going Research / Outlook

At the moment, we are assembling datasets from different measurements into one large dataset, with the aim to achieve higher resolution of the AMP-PNP incubated structure. This would allow us to track individual subunits of the proteolytic chamber down to the N-terminal domain. The uttermost goal of our future high-resolution studies is to correlate hLon's structure with the occupancy of its ATP-binding sites, which requires acquisition of several structures with resolution < 6Å. Furthermore, we are interested in structural changes of Lon caused by the binding of protein substrates and mtDNA. In order to achieve these goals, we wish to acquire many more cryo-EM datasets and process them on the IT4Innovations clusters.

Conclusion

The parallel processing on the IT4Innovations clusters allowed us to determine the so far unknown structures of the human mitochondrial Lon protease in a relatively short time. In contrast to other computational approaches that do not require computational infrastructure as large as the IT4Innovations clusters, the Bayesian processing implemented in the Relion software brought the desired results. Therefore, we would be happy to continue using the IT4Innovations resources for the structural studies of the Lon protease.

References

[1] T.V. Rotanova, I. Botos, E.E. Melnikov, F. Rasulova, A. Gustchina, M.R. Maurizi, A. Wlodawer. Slicing a protease: Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains, Protein Sci, 15 (2006) 1815-1828.

[2] I. Lee, C.K. Suzuki. Functional mechanics of the ATP-dependent Lon protease-lessons from endogenous protein and synthetic peptide substrates, Bba-Proteins Proteom, 1784 (2008) 727-735.

[3] R. Fukuda, H. Zhang, J.W. Kim, L. Shimoda, C.V. Dang, G.L. Semenza. HIF-1 regulates cytochrome oxidase subunits to optimise efficiency of respiration in hypoxic cells, Cell, 129 (2007) 111-122.

[4] S. Venkatesh, J. Lee, K. Singh, I. Lee, C.K. Suzuki. Multitasking in the mitochondrion by the ATP-dependent Lon protease, Bba-Mol Cell Res, 1823 (2012) 56-66.

[5] M. Pinti, L. Gibellini, Y. Liu, S. Xu, B. Lu, A. Cossarizza. Mitochondrial Lon protease at the crossroads of oxidative stress, ageing and cancer, Cellular and molecular life sciences: CMLS, 72 (2015) 4807-4824.

[6] N.D. Rawlings, F.R. Morton, A.J. Barrett. MEROPS: the peptidase database, Nucleic acids research, 34 (2006) D270-272.

[7] J. Garcia-Nafria, G. Ondrovicova, E. Blagova, V.M. Levdikov, J.A. Bauer, C.K. Suzuki, E. Kutejova, A.J. Wilkinson, K.S. Wilson. Structure of the catalytic domain of the human mitochondrial Lon protease: Proposed relation of oligomer formation and activity, Protein Sci, 19 (2010) 987-999.

[8] A.F. Neuwald, L. Aravind, J.L. Spouge, E.V. Koonin. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes, Genome research, 9 (1999) 27-43.

[9] T. Ogura, A.J. Wilkinson. AAA+ superfamily ATPases: common structure – diverse function, Genes Cells, 6 (2001) 575-597.

[10] T. Ishikawa, M.R. Maurizi, A.C. Steven. The N-terminal substrate-binding domain of ClpA unfoldase is highly mobile and extends axially from the distal surface of ClpAP protease, J Struct Biol, 146 (2004) 180-188.

[11] S. Lee, J.M. Choi, F.T. Tsai. Visualising the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB, Mol Cell, 25 (2007) 261-271.

[12] M.J. Landsberg, P.R. Vajjhala, R. Rothnagel, A.L. Munn, B. Hankamer. Three-Dimensional Structure of AAA ATPase Vps4: Advancing Structural Insights into the Mechanisms of Endosomal Sorting and Enveloped Virus Budding, Structure, 17 (2009) 427-437.

[13] L. Ambro, V. Pevala, J. Bauer, E. Kutejova. The influence of ATP-dependent proteases on a variety of nucleoid-associated processes, J Struct Biol, 179 (2012) 181-192.

[14] S. Langklotz, U. Baumann, F. Narberhaus. Structure and function of the bacterial AAA protease FtsH, Biochim Biophys Acta, 1823 (2012) 40-48.

[15] S. Lee, S. Augustin, T. Tatsuta, F. Gerdes, T. Langer, F.T.F. Tsai. Electron Cryomicroscopy Structure of a Membrane-anchored Mitochondrial AAA Protease, Journal of Biological Chemistry, 286 (2011) 4404-4411.

[16] M. Vostrukhina, A. Popov, E. Brunstein, M.A. Lanz, R. Baumgartner, C. Bieniossek, M. Schacherl, U. Baumann. The structure of Aquifex aeolicus FtsH in the ADP-bound state reveals a C2-symmetric hexamer, Acta Crystallogr D Biol Crystallogr, 71 (2015) 1307-1318.

[17] E. Gur, R.T. Sauer. Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine, Proc Natl Acad Sci U S A, 106 (2009) 18503-18508.

[18] A.S. Menon, A.L. Goldberg. Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP, Journal of Biological Chemistry, 262 (1987) 14929-14934.

[19] L. Waxman, A.L. Goldberg. Selectivity of intracellular proteolysis: protein substrates activate the ATP-dependent protease (La), Science, 232 (1986) 500-503.

[20] M.L. Wohlever, T.A. Baker, R.T. Sauer. Roles of the N domain of the AAA plus Lon protease in substrate recognition, allosteric regulation and chaperone activity, Molecular Microbiology, 91 (2014) 66-78.

[21] J.R. Hoskins, M. Pak, M.R. Maurizi, S. Wickner. The role of the ClpA chaperone in proteolysis by ClpAP, Proc Natl Acad Sci U S A, 95 (1998) 12135-12140.

[22] S.C. Park, B. Jia, J.K. Yang, D.L. Van, Y.G. Shao, S.W. Han, Y.J. Jeon, C.H. Chung, G.W. Cheong. Oligomeric structure of the ATP-dependent protease La (Lon) of Escherichia coli, Mol Cells, 21 (2006) 129-134.

[23] H. Stahlberg, E. Kutejova, K. Suda, B. Wolpensinger, A. Lustig, G. Schatz, A. Engel, C.K. Suzuki. Mitochondrial Lon of Saccharomyces cerevisiae is a ring-shaped protease with seven flexible subunits, Proc Natl Acad Sci U S A, 96 (1999) 6787-6790.

[24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank, Nucleic Acids Res, 28 (2000) 235-242.

[25] R.E. Duman, J. Lowe. Crystal Structures of Bacillus subtilis Lon Protease, Journal of Molecular Biology, 401 (2010) 653-670.

[26] M. Li, F. Rasulova, E.E. Melnikov, T.V. Rotanova, A. Gustchina, M.R. Maurizi, A. Wlodawer. Crystal structure of the N-terminal domain of E. coli Lon protease, Protein Sci, 14 (2005) 2895-2900.

[27] J. Dubochet, M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, A.W. McDowall, P. Schultz. Cryo-electron microscopy of vitrified specimens, Q Rev Biophys, 21 (1988) 129-228.

[28] S. Kereiche, L. Kovacik, J. Bednar, V. Pevala, N. Kunova, G. Ondrovicova, J. Bauer, L. Ambro, J. Bellova, E. Kutejova, I. Raska. The N-terminal domain plays a crucial role in the structure of a full-length human mitochondrial Lon protease, Sci Rep, 6 (2016) 33631.

[29] S.H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination, Journal of Structural Biology, 180 (2012) 519-530.

Publication

[1] Kereiche, S.; L. Kovacik, J. Bednar, V. Pevala, N. Kunova, G. Ondrovicova, J. Bauer, L. Ambro, J. Bellova, E. Kutejova and I. Raska. The N-terminal domain plays a crucial role in the structure of a full-length human mitochondrial Lon protease, Sci Rep 6. 2016, 33631. Doi: 10.1038/srep33631

ENHANCEMENT OF THE IT4IBLENDER TOOL FOR PARALLEL IMAGE SEGMENTATION OF COMPUTER TOMOGRAPHY AND MAGNETIC RESONANCE IMAGES

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Tomáš Karásek

Researchers:
Pavel Maršálek, Petr Ferfecki, Petr Horyl, Tomáš Brzobohatý

Project ID:
OPEN-7-24

Introduction

Areas utilising powerful computer technologies are getting larger with their availability. One of the areas where such technologies are widely used is medicine. Especially, devices for medical imaging use computer technologies extensively. Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) scanners can be mentioned as typical examples of such. These devices can provide image information about the patient's inner body structure. With this information higher standards in diagnostic medicine can be reached. To move the capabilities even further, post-processing of acquired data is done. It brings extended possibilities such as creation of virtual models of human organs or other tissues.

To achieve the best possible results not only powerful hardware, but also capable software has to be used. Typically, the software for 3D model reconstructions from medical images has to proceed through several steps that can be considered as general for any software of such kind. The scope of this project was enhancement of the IT4IBlender tool for parallel image segmentation. Our software is used as a plug-in to the 3D graphics and animation software Blender [1]. Within the tool great emphasis is placed on utilisation of parallel computing and HPC resources. The more demanding tasks are implemented in parallel using OpenMP and MPI in C++ language. Our implementation can exploit the computational power of both non-accelerated nodes as well as Intel Xeon Phi accelerated compute nodes of a cluster.

Results and Methods

The first thing to do in post-processing software is loading of the data. Concerning medical image data, it is stored in the DICOM format. To load this data into Blender, the Grassroots DICOM (GDCM) C++ library is integrated in our plug-in.

After loading the medical image data, its visualisation is provided to users in a typical 3-axial view (axial, sagittal, coronal). It might be beneficial to restrict such image data to only some regions of user interest and thus reduce the computational time of the further applied image processing techniques. In our plug-in, we can restrict the computations only to the volume area of our interest by using the box or sphere cutting tool.

Another step within the 3D model reconstruction process is image filtering. The imaging techniques like CT or MRI suffer from the presence of residual image noise. This usually has significant impact on subsequent operations performed with an image. The image noise is characterised by high frequency content within the image, therefore usually simple low pass filters are used for noise reduction. Such filters have a blurring or smoothing effect on an image because they average out abrupt changes in pixel intensities. The drawback of such filtering is that it also smoothes out the edges, thus the object boundaries inside the image deteriorate. To preserve the edges and filter out the noise, more sophisticated filtering techniques have to be used. In our plug-in we have implemented several filtering techniques, such as simple Gaussian smoothing [2], a more sophisticated anisotropic diffusion filter [3], and a very effective and state-of-the-art Sparse 3D Transform-Domain Collaborative filter (BM3D) [4].

After filtering the images, image segmentation is performed. It serves as image simplification in order to localise the objects and their boundaries. Two methods are currently implemented in our plug-in. These are image thresholding [2] and k-means clustering [2]. The parallel implementation of the k-means method by using Intel Xeon Phi was introduced in the paper [5]. Both methods can be used separately or combined together.

If the segmentation of the selected body tissue is finished, a reconstruction technique that provides iso-surfaces as a result can be applied. In our plug-in we use Screened Poisson Surface Reconstruction and Metaballs. Both methods use information about identified boundaries as an input parameter. To create a polygonal mesh from iso-surfaces, the Marching Cubes method is used. This method passes through the input scalar field and from the eight neighbouring points simultaneously (imaginary cube) computes the polygons, which represent a corresponding part of the iso-surface. This method is time demanding, and as such efficient parallel implementation was necessary. For visualisation, the Blender Cycles engine is used. To accelerate the computations, the architecture of the Graphic Processing Unit (GPU) (the implementation is already included in Cycles) and Intel's Many Integrated Core (MIC) architecture can be employed. Our implementation for MIC architecture contains all features of the engine with improved performance [6].

We have incorporated an additional set of tools within the plug-in to better suit the diagnostic and surgery planning purposes of medical doctors. In the current version of the plug-in, painting and erasing for image segmentation, 3D volume cutting and 2D/3D measuring tools are implemented.

The painting and erasing tool is meant to correct the results of the semi-automatic segmentation. It is called 2D Painter/Eraser and it can also be used to perform manual segmentation. The tool uses a typical brush and erase buttons, which open the segmentation layer for editing, see Figure 1.

A similar tool that allows the user to mark the surface area of a selected 3D object, but serves another purpose than image segmentation, is called 3D Painter/Eraser. This tool works as a 3D brush, which paints on the surface. The surface area marked by the 3D Painter/Eraser is subsequently processed to obtain the total area as depicted in Figure 1.

Since surgeons are interested in using this software for planning surgeries where, e.g., precise measurement of organ volume or resection of a selected part of a tissue is necessary for successful treatment, we have developed specific tools for such manipulations. The first one is the Cut Volume tool for creating arbitrary cuts of a segmented object volume, and the second one called Volume Measurement calculates the volume of an object, see Figure 2.

Because surgeons and radiologists usually work with 2D images, the 3D models resulting from the reconstructions can sometimes be confusing for them. Due to this reason and to be able to check the correctness of the provided reconstructions, we have developed the Backward Projection tool. This tool can show the shape of the reconstructed 3D model as a projection on the original CT data. This allows users to see all important features of the data and at the same time compare the accuracy of the provided reconstruction. An example of the application of the tool is demonstrated on the model of a liver in Figure 3.

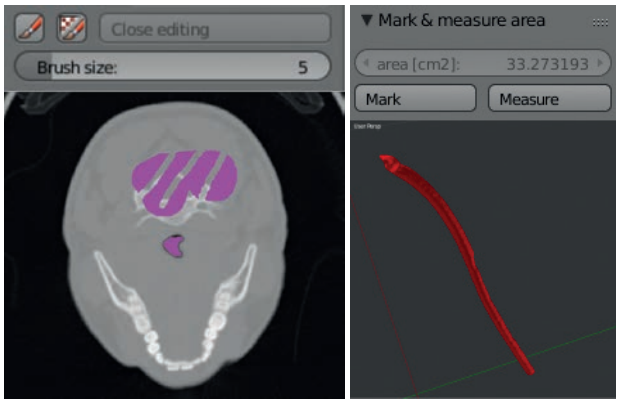


Figure 1.
left – 2D Painter/Eraser,
right – measured area by 3D Painter/Eraser

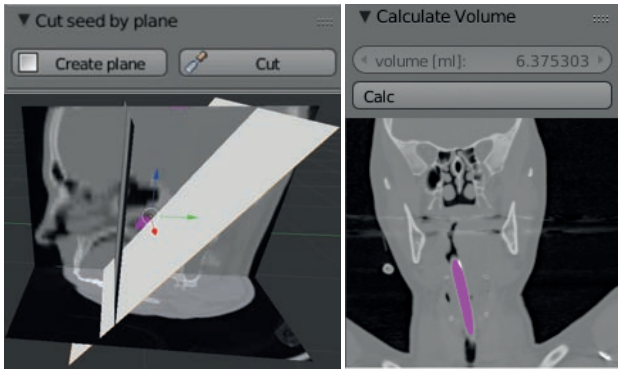


Figure 2.
left - Cut Volume tool,
right – Volume Measurement

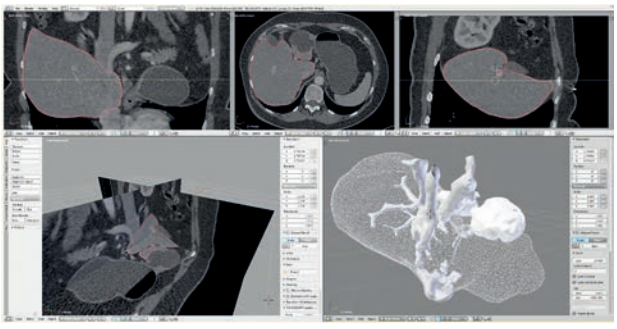


Figure 3.
left – 2D Painter/Eraser,
right – measured area by 3D Painter/Eraser

On-going Research/Outlook

Our plan is to further develop the plug-in for Blender, especially to include a broader variety of image segmentation and filtering methods and thus offer the best possible quality of the output 3D reconstructions.

Conclusion

We have created a software extension in the form of a plug-in for Blender that offers a complete post-processing tool for medical image data. Valuable 3D model reconstructions utilising either standard PC or the HPC resources can be produced with the help of the tool within the Blender environment. The tool is meant for doctors for diagnostic purposes or for pre-planning surgeries.

References

[1] blender.org – Home of the Blender project – Free and Open 3D Creation Software. Available from: <https://www.blender.org>
[2] Sonka M.; V. Hlavac and R. Boyle. Image Processing, Analysis and Machine Vision, Thomson, 2006. ISBN: 978-0-412-45570-4
[3] Perona P. and J. Malik. Scale-space and Edge Detection Using Anisotropic Diffusion. In: IEEE Trans. on Pattern Analysis and Machine Intelligence. 1990, 12, 7, 629-639. Doi: 10.1109/34.56205
[4] Dabov K.; A. Foi, V. Katkovnik, K. Egiazarian. Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering. In: IEEE Trans. Image Process. 2007, 16, 8, 2080-2095.
[5] Strakos, P.; M. Jaros, T. Karasek, L. Riha, M. Jarosova, T. Kozubek, P. Vavra, T. Jonszta. Parallelisation of the Image Segmentation Algorithm for Intel Xeon Phi with Application in Medical Imaging. In P. Ivnyi, B.H.V. Topping, (Editors), Proceedings of the Fourth International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering, Civil-Comp Press, Stirlingshire, UK, Paper 7, 2015. Doi 10.4203/ccp.107.7
[6] Jaros, M.; L. Riha, T. Karasek, P. Strakos, A. Vasatova, M. Jarosova, T. Kozubek. Acceleration of Blender Cycles Path-Tracing Engine using Intel Many Integrated Core Architecture. Proceedings of the 14th International Conference on Computer Information Systems and Industrial Management Applications, Warsaw, Poland, 2015. Doi: 10.1007/978-3-319-24369-6_7

Publications

[1] Jaros M.; M. Jarosova, P. Strakos, and T. Karasek. Parallel plug-in for 3D reconstruction of medical images. SCCG2016, 2016
[2] Strakos P.; M. Jaros, T. Karasek, M. Jarosova, D. Krpelik, A. Vasatova, T. Buresova, J. Timkovic, J. Stembirek, and P. Vavra. Medical image processing tools for blender with HPC support. WSCG 2016 – 24th WSCG Conference on Computer Graphics, Visualisation and Computer Vision, 2016
[3] Krpelik D.; M. Jaros, M. Jarosova, P. Strakos, T. Buresova, A. Vasatova and T. Karasek. Detection of orbital floor fractures by principal component analysis. In: Saeed K., Homenda W. (eds) Computer Information Systems and Industrial Management. CISIM 2016. Lecture Notes in Computer Science. Springer, 2016, 9842, 129-138. Doi 10.1007/978-3-319-45378-1_12
[4] Strakos P.; M. Jaros, T. Karasek, and T. Kozubek. Parallel implementation of collaborative filtering technique for denoising of CT images. In: Kozubek T., Blaheta R., Šístek J., Rozložník M., Čermák M. (eds) High Performance Computing in Science and Engineering. HPCSE 2015. Lecture Notes in Computer Science, vol. 9611. Springer, 2015, 9611, 126-. Doi: 10.1007/978-3-319-40361-8_9

MODELLING BEAM DISTORTION DURING FOCUSED ULTRASOUND SURGERY IN THE PROSTATE USING K-WAVE

Research institution:
Brno University of Technology

Principal investigator:
Jiří Jaroš

Researchers:
Bradley Treeby, Panayiotis Georgiou

Project partner:
University College London

Project IDs:
OPEN-5-13, OPEN-6-11, OPEN-7-15

Introduction

Prostate cancer is one of the most commonly occurring cancers for men in Europe and the US, and a leading cause of cancer-related death [1]. For patients with early-stage localised disease, the cancer is often treated using external beam radiation therapy (EBRT) [2]. This procedure usually involves implanting a small number of gold fiducial markers into the prostate to verify the position of the prostate gland between treatments [3]. For many of these patients, their cancer will recur. In those cases, further treatment using a salvage therapy is necessary, which does not expose them to additional levels of radiation. HIFU is currently offered in hospitals as a minimally invasive salvage therapy for treating prostate cancer in patients whose cancer recurred after failed EBRT [2].

The efficacy and safety of salvage HIFU treatment after failed EBRT has been investigated in clinical studies. These studies reported good local cancer control but with some cases having high complication rates, comparable to other salvage therapies [2]. However, none of these studies consider the impact of implanted fiducial markers on the delivery of the HIFU treatment.

The objective of this work was to systematically investigate, using computational simulations, how the fiducial markers affect the delivery of HIFU treatment. The impact of the marker was studied through a series of large-scale simulations modelling the propagation of ultrasound pressure waves in the prostate with a single gold marker obstructing the beam's path. In each simulation a single spherical or cylindrical marker was included at different positions and orientations. For each marker configuration, a set of metrics (spatial-peak temporal-average intensity, focus shift, focal volume) was evaluated to quantify the distortion introduced at the focus in comparison to the corresponding metrics of a homogeneous simulation without a marker.

Results and Methods

The simulations were performed using the open-source k-Wave acoustic simulation toolbox developed by our group [4]. The toolbox solves a generalised version of the Westervelt equation which accounts for the combined effects of nonlinearity, heterogeneous material properties and acoustic absorption following a frequency power law. K-Wave is designed and optimised to fully utilise the computational resources offered by the IT4Innovations' Salomon and Anselm supercomputers on which the simulations were executed.

The simulated domain corresponded to a physical volume with dimensions 44.7x29.4x60.0 mm^3. The background medium was assigned the material properties (density and sound speed) of prostate tissue. For the heterogeneous simulations a spherical or cylindrical volume corresponding to the marker was also included and assigned the material properties of gold. All simulations were non-linear and accounted for absorption using a power law. The simulations were performed using a regular Cartesian mesh with a 1536x1024x2048 pt^3 grid-size. The large grid-size used, determined after a series of convergence tests, is necessary so that higher harmonics are supported in order to increase the accuracy with which acoustic non-linearity is captured. Such simulations have extreme computational and memory requirements. For example, a single heterogeneous simulation (with a marker) using 128 physical cores on Salomon requires approximately 332 GB of RAM and 8.5 days to complete, has a 226 GB input file and generates a 446 GB output. In total this study has consumed approximately 5 million core-hours.

The transducer model used in the simulations was derived from the transrectal probe of the Sonablate 500 (SonaCare Medical) clinical HIFU system used for treating prostate cancer. More specifically, the transducer geometry was assumed to be a single-element spherical cap with 22 mm width, 35 mm length and a fixed 40 mm focal length. A cross-section of the transducer model can be seen in Figure 1. The driving parameters of the transducer model were adjusted so that the spatial-peak temporal-average intensity at the focus of a homogeneous simulation (without a marker) is similar (1.1 kW/cm^2) to the values reported for the Sonablate 500 [5].

The distortion introduced by the fiducial markers to the HIFU beam was investigated by including a single spherical or

cylindrical gold marker positioned at different coordinates in each simulation. The position of the transducer was kept fixed across all simulations. The spherical marker had a 3 mm diameter, whereas the cylindrical had a 3 mm height and 1 mm diameter. In total, 143 marker positions were simulated: 113 with a spherical marker and 10 with a cylindrical marker at 3 orientations. All the simulated positions are shown in Figure 1.

To quantify the effect of a single marker on the focusing of the HIFU beam, four metrics were evaluated using the simulation results for each marker position. These metrics were compared to the corresponding quantities obtained from a homogeneous simulation without a marker. In Figure 2 the effect of a single marker is shown in comparison to a homogeneous simulation.

The first quantity evaluated was the focal shift, calculated from the coordinates of the transducer's geometric focus and the coordinates of the maximum pressure point. The next two metrics were the spatial-peak temporal-average intensity evaluated at the geometric focus (I_{focus}) of the transducer and at the coordinates of the maximum-pressure point (I_{max}). These two quantities provide an indication of how much energy is redistributed due to the marker. The fourth metric evaluated the -6 dB focal volume for each simulation, which provides an indication of how the size of the focal region changes in comparison to the homogeneous simulation due to the presence of the marker.

The four metrics evaluated for each position of the spherical marker are shown in Figure 3. Each metric is plotted with respect to the marker distance from the transducer's focus with the distance increasing from the focus when the marker is positioned closer to the transducer. The metrics demonstrate that the marker acts as a strong reflector of the HIFU beam. When the marker is positioned in the pre-focal region, it causes reflections which induce a decrease in the focal intensity and focal volume, and a shift of the maximum pressure point away from the transducer's focus. These effects become more pronounced as its distance from the transducer's focus decreases, with the distortion introduced by the marker greatly increasing when placed within approximately 5 mm of the focus. The analogous simulations performed for the cylindrical marker also demonstrate that these effects depend on the shape and orientation of the marker.

Conclusion

A series of large-scale simulations was performed in order to study and quantify the impact of a single spherical or cylindrical gold fiducial marker on the HIFU beam. The four metrics evaluated for each marker position have shown that the distortion introduced by the marker increases as its distance from the transducer's focus decreases and depends on the marker's shape and orientation. The distortion when the marker is positioned within 5 mm of the focus significantly increases. This may result in an undertreated region beyond the marker due to less energy arriving at the focus, and an over-treated region due to reflections. Both effects may be undesirable depending on the location of the marker relative to the targeted cancerous region or other organs at risk.

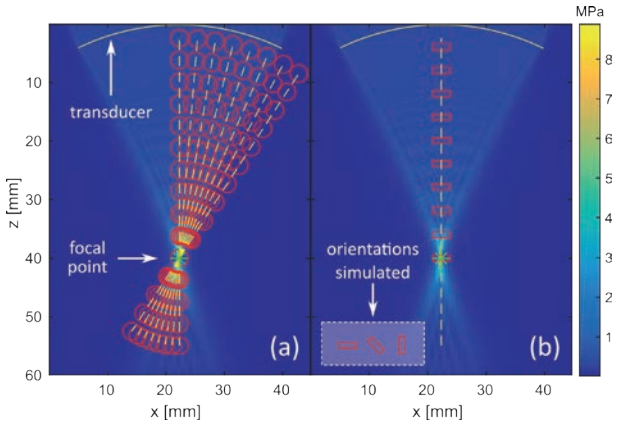


Figure 1.
The positions and orientations simulated for (a) the spherical and (b) the cylindrical markers. The background is the maximum pressure field from a homogeneous simulation.

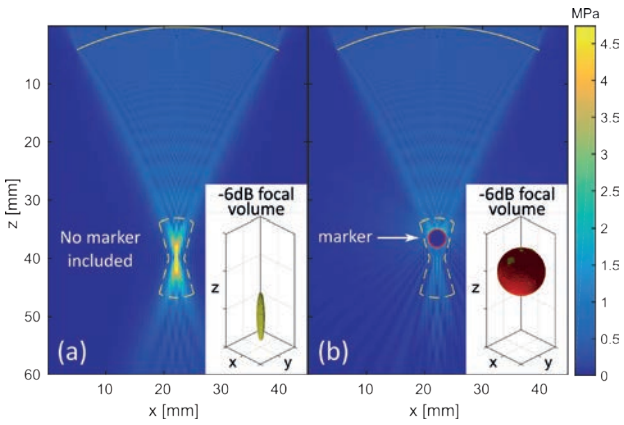


Figure 2.
Change in the maximum pressure field (b) due to the presence of a single spherical marker close to the focus compared to (a) the homogeneous simulation without a marker. The insets show the change in the -6 dB focal volume due to the marker.

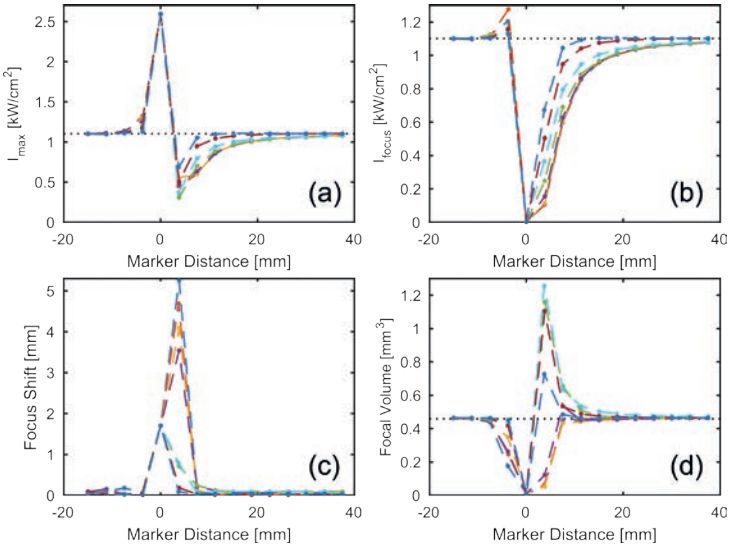


Figure 3.
The four metrics used to quantify the effect a fiducial is shown here for the spherical marker case.

On-going Research / Outlook

While investigating the impact of the markers, some additional factors have been omitted which may affect the distortion introduced by the marker to the treatment. Some of these factors are: the imaging element of the transducer, which was not included in the model, additional heating occurring due to absorption within the marker or viscous relative motion between the marker and surrounding tissue and the effect of multiple sonications during such treatments. In order to confirm the results observed from the acoustic simulations while taking into account the impact of many of the omitted factors above, our team is currently using experimental measurements on ex vivo tissue phantoms with implanted markers. The experimental work will also be extended to include measurements on ex vivo human prostate specimens.

It is also interesting to extend the results therein for salvage-HIFU treatment after failed (low-dose) brachytherapy. In this scenario, a large number of marker-like elements are introduced in the prostate. The material properties of these elements are similar to those of fiducial markers, and thus, expected to cause significant distortions to the HIFU beam.

References

[1] Ferlay, J., et. al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 International Agency for Research on Cancer, 2013

[2] Warmuth, M., et. al. Systematic Review of the Efficacy and Safety of High-Intensity Focused Ultrasound for the Primary and Salvage Treatment of Prostate Cancer European Urology, European Association of Urology, 2010, 58, 803-815. Doi: 10.1016/j.eururo.2010.11.042

[3] Fuller, C. D. & Scarbrough, T. J. Fiducial Markers in Image-guided Radiotherapy of the Prostate. Oncology & Hematology Review (US), 2006, 00, 75. Doi: 10.17925/OHR.2006.00.02.75

[4] Treeby, B. E. & Cox, B. T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. Journal of Biomedical Optics, 2010, 15, 021314. Doi: 10.1117/1.3360308

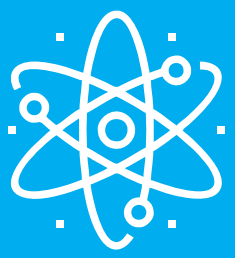
[5] Illing, R. and M. Emberton. Sonablate-500: transrectal high-intensity focused ultrasound for the treatment of prostate cancer. Expert Review of Medical Devices, 2006, 3, 717-729. Doi: 10.1586/17434440.3.6.717

Publication

[1] Georgiou, P. S.; Jaros, J., Payne, H., Allen, C., Shah, T.T., Ahmed, H.U., Gibson, E., Barratt, D. and B. E. Treeby. Beam Distortion Due to Gold Fiducial Markers During Salvage High-Intensity Focused Ultrasound in the Prostate. Medical Physics. 2016. Doi: 10.1002/mp.12044.

Project website: bug.medphys.ucl.ac.uk

07 | PHYSICS



QUANTUM COLLISIONS OF ELECTRONS WITH HYDROGEN ATOMS

Research institution:
Charles University in Prague

Principal investigators:
Jakub Benda, Karel Houfek

Project ID:
OPEN-6-16

Introduction

Quantum interaction of electrons with atomic hydrogen gas has been studied almost since the inception of quantum theory. One of the reasons is the exceptional apparent simplicity of the e-H scattering problem. There are three particles only: the proton, the atomic electron and the incident electron, and their motion is described very well by the non-relativistic Schrödinger equation with simple Coulombic potentials. As such, this model served as a testing ground for various physical theories. Still, this three-particle problem is not analytically solvable and, as it turned out, even the numerical solution is difficult; in some ways even more difficult than similar calculations for heavier atoms, because of a much weaker attractive force caused by the nucleus and due to a (related) huge polarisability of the "electronic cloud" surrounding the nucleus.

The other reason for the popularity of e-H scattering calculations is their real-world value. While the atomic hydrogen gas is unstable in terrestrial conditions, where it quickly recombines – and this is incidentally a reason for virtual absence of available laboratory data from these collisions – it is very abundant in outer space, where densities are so low that the equilibrium between the UV dissociation and recombination is well on the side of the former. The collisional transitions in hydrogen atom levels caused by electron impacts majorly contribute to the thermal properties of the interstellar gas [1], which makes the microscopic data an important ingredient for analysis of astrophysical observations. There is also a lot of hydrogen in stars, where the electronic collisions define the quantum level populations and thus also specific shapes of the emission spectral lines and their polarisation [2]. This way, the knowledge of the low-level electron-hydrogen processes makes it possible to translate the optical observations to the real particle dynamics in the surface layers of stars, providing another tool for understanding, and possibly predicting, the events taking place in the stellar matter, especially including our Sun. Yet another use case for electron-hydrogen scattering has been the thermonuclear fusion industry, where the isotopes of hydrogen in an ionised plasma mixture play the lead [3].

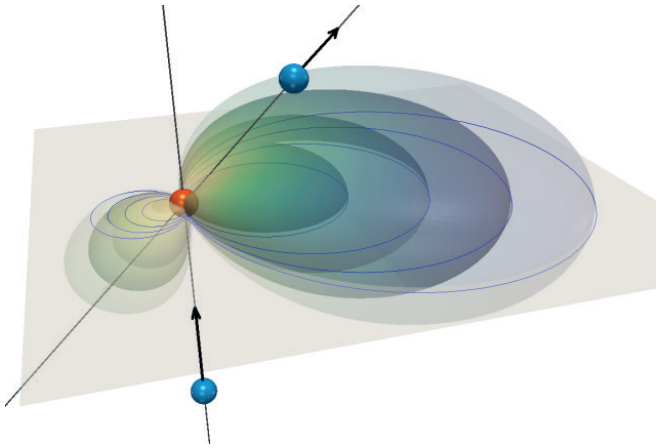


Figure 1.
Visualisation of the hydrogen ionisation amplitude for given impact and retreat directions of the projectile electron.

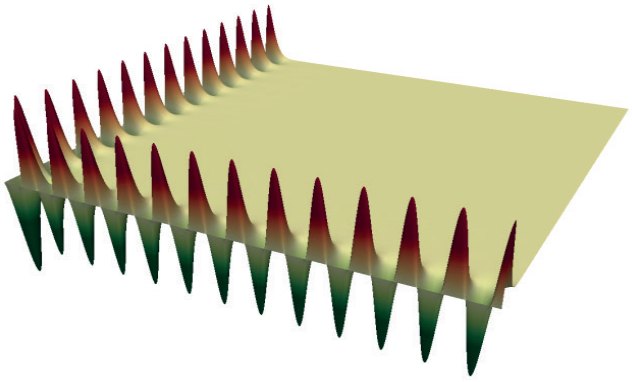


Figure 2.
A typical radial part of a two-electron wave function below the ionisation threshold. The two evanescent waves correspond, each, to a single electron escaping to the infinity.

The aim of this project is to produce a complete and consistent database with results for all possible (or astrophysically relevant) collision processes in the e-H system. This involves a lot of work as there is theoretically an unlimited amount of initial and final possible atomic states, impact energies, angular momenta and other physical parameters and conditions of the collision. A careful selection (and subsequent interpolation) of a representative sample is almost as important a task as the numerical calculations themselves. The resulting database will be freely accessible on the web of the Institute and connected to global specialised astrophysical data searching engines.

Results and Methods

Electron-hydrogen collisions have managed to resist many ingenious solution methods throughout the last century. While some qualitative results have been available for a long time, achieving a few-percent accuracy, which this problem deserves (and the astronomers require) only became possible quite recently with the rise of modern supercomputers.

The typical method of choice in this field of physics is the R-matrix method [4], which we, nevertheless, avoided, as it is not sufficiently flexible due to a necessity of a diagonalisation of huge matrices. The size of the matrices in our two-electron case is effectively quadratically dependent on the simulated radius around the atom, which needs to be very large because the excited hydrogen state becomes very diffuse. Moreover, a diagonalisation is notorious to loosing precision, and difficult to parallelise.

Instead, we use the Exterior Complex Scaling (ECS) method, where the Schrödinger equation – which in this setting is essentially a complex two-dimensional Helmholtz equation with a continuous source and feedback – is expanded into a product basis of B-splines to become a very sparse linear-algebraic problem, which is then solved by an iterative method for every initial target state and impact energy of interest [5]. The complicated boundary conditions are transformed to simple zero Dirichlet conditions using a clever complex rotation trick. The independence of calculations for individual states and energies allows us to run many smaller calculations concurrently with an angular and radial basis tailored to the impact energy, etc., rather then attempting a single vast diagonalisation, on which everything else depends.

The iterative method used, called conjugated orthogonal conjugate gradients, is a variant of the well-known conjugate gradients targeted on indefinite symmetric complex problems. Because the systems can become very large, even up to billions of unknowns, we use preconditioning, typically block incomplete LU (done by Pardiso, UMFPACK or MUMPS libraries) or a special Kronecker product approximation [6], where dense matrix algebra is used.

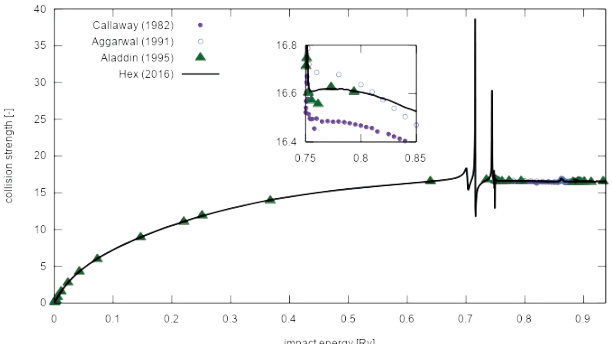


Figure 3.
Collision strength for the elastic scattering on the ground state at low impact energies compared with some already published data. This most basic process is easily calculable using virtually any quantal method. Our results are shown as a solid line.

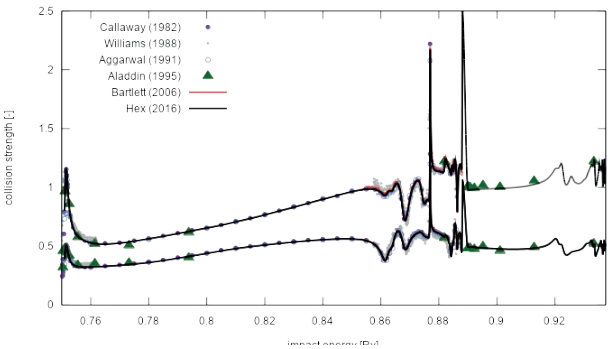


Figure 4.
Collision strengths for the two lowest excitations from the ground state: to H(2s) and H(2p), summed over final magnetic sub-levels. As above, the agreement with older results is good. This is no longer true for higher excitations, where ECS is able to reach larger radial grids and so achieves a better radial convergence.

During the OPEN-6-16 project we have calculated accurate collision cross sections for all transitions between the hydrogenic states 1s, 2s, 2p, 3s, 3p and 3d, including all magnetic sub-levels, at finely spaced energies below the n = 4 excitation threshold. These data are already relevant for astrophysical simulations of low-temperature processes, though they will become much more useful when supplemented with the higher-energy results to enable Maxwellian weighting and calculation of effective scattering rates.

On-going Research / Outlook

Recently we successfully enriched the ECS method with an efficient asymptotic extension of the real radial grid, which allows us to easily expand the solution domain to cover all long-range polarisation forces. This reduces the asymptotic computational complexity by a power for impact energies below the ionisation threshold [7].

In the next step we'd like to focus more on high-energy scattering, where a different approach will be necessary to maintain the embarrassingly parallel character of the calculations. A time-proven method for all variants of the Helmholtz equation is the domain decomposition [8], which will be used in the continuation of this project.

Conclusion

Collisions of electrons with hydrogen atoms are in a great position to be solved once and for all to a great benefit of the applied researchers from the astrophysics and fusion plasma industry. This is the purpose of our project, which intends to collect a comprehensive set of simulated data from e-H scattering events.

Using the computational resources from IT4Innovations we have advanced the state-of-the-art ECS method for perfect theoretical efficiency in the low-impact-energy regime and produced a major part of the desired data volume.

References

[1] Furlanetto, S.R. and M. R. Furlanetto. Spin-exchange rates in electron-hydrogen collisions. Monthly Notices of the Royal Astronomical Society. 2007, 374(2), 547-555. Doi: 10.1111/j.1365-2966.2006.11169.x

[2] Kašparová, J. and P. Heinzel. Diagnostics of electron bombardment in solar flares from hydrogen Balmer lines. Astronomy & Astrophysics. 2002, 382(2), 688-698. Doi: 10.1051/0004-6361:20011599

[3] Lisitsa, V.S. et al. Hydrogen spectral line shape formation in the SOL of fusion reactor plasmas. Atoms. 2014, 2(2), 195-206. Doi:10.3390/atoms2020195

[4] Descouvemont, P. and D. Baye. The R-matrix theory. Reports on Progress in Physics. 2010, 73(3), 036301-. Doi: 10.1088/0034-4885/73/3/036301

[5] Benda, J. and K. Houfek. Collisions of electrons with hydrogen atoms II. Low-energy program using the method of exterior complex scaling. Computer Physics Communications. 2014, 185(11), 2903-2912. Doi: 10.1016/j.cpc.2014.05.021

[6] Bar-On, I., Edlund, Å. and U. Peskin. Parallel solution of the multidimensional Helmholtz/Schrödinger equation using high order methods. Applied Numerical Mathematics. 2000, 33, 95-104. Doi: 10.1016/S0168-9274(99)00072-0

[7] See "Publications" [2] below.

[8] Stolk, C.S. A rapidly converging domain decomposition method for the Helmholtz equation. Journal of Computational Physics. 2013, 241, 240-252. Doi: 10.1016/j.jcp.2013.01.039

Publications

[1] Benda, J. and K. Houfek. New version of hex-ecs, the B-spline implementation of exterior complex scaling method for solution of electron-hydrogen scattering. Computer Physics Communications. 2016, 204, 216-217. Doi: 10.1016/j.cpc.2016.03.020

[2] Benda. J. and K. Houfek. Reducing the dimensionality of grid based methods for electron-atoms scattering calculations below the ionisation threshold. Computer Physics Communications, under review, 2016.

Project website: utf.mff.cuni.cz/data/hex

TEBDOL – PARALLEL SIMULATION OF ULTRACOLD ATOMS IN OPTICAL LATTICES

Research institution:
Charles University in Prague

Principal investigators:
Miroslav Urbánek, Pavel Soldán

Project ID:
OPEN-4-18

Introduction

Optical lattices are experimental devices that open huge possibilities of examination of fundamental microscopic laws. They have a potential to answer several unsolved questions in theoretical physics, which still resist all attempts to solve them. They have also been proposed as a possible realisation of a quantum computer. Experiments in optical lattices make use of ultracold atoms that are confined into a periodic structure created using laser light. Currently suitable atoms are alkali-metal atoms and alkaline-earth-metal atoms.

The aim of this project was to study the behaviour of such systems and to calculate their properties numerically. We found allowed states of selected systems and determined their time evolution as a result of changes in lattice parameters. Because the calculations were numerically challenging, we developed a parallel program to simulate these systems.

Results and Methods

Our method to simulate collective dynamics of a complex multi-particle quantum system is based on the Tensor Network family of algorithms. The first such algorithm, the Density Matrix Renormalisation Group (DMRG) [1] algorithm, became a standard tool for calculating the ground state of one-dimensional quantum systems. With the invention of the Time-Evolving Block Decimation (TEBD) [2] algorithm, it became possible to study the time evolution of these systems. More recently, these algorithms and several others were united in the framework of Matrix Product States and later in the framework of Tensor Network States [3].

In contrast to the standard approach to quantum mechanics, where a physical state is described by a complex vector, we describe the state as a set of interconnected multi-dimensional complex tensors. Tensor networks allow us to truncate the state vector components to those with relevant contributions only, thus reducing the total problem dimension.

The main objective of this project was to develop a parallel code for simulation of one-dimensional systems. The original algorithm is not scalable, because normalisation conditions require working on a single tensor at the time. It is necessary to keep track of the global system state to correctly cut off states with low probability. Therefore we slightly modified the method to relax the normalisation conditions. Our approach introduced another source of inaccuracy, because we do not require complete normalisation of all tensors. However, this inaccuracy is very small in the regime we are interested in, i.e., for simulations with small truncations. The speedup of parallel calculations makes it a reasonable compromise.

We wrote the program in Common Lisp, which is a non-typical language in high-performance computing. Common Lisp is a high-level language that can produce fast machine code. It allowed us to handle the complex tensor structures easily and to quickly test new ideas. The program depends on BLAS, LAPACK and MPI libraries.

The algorithm is bound by the amount of available memory. We use MPI parallelisation to distribute computing processes across nodes. Although Salomon includes nodes with powerful GPU and MIC accelerators, they do not help in our case. We tested the code using a MIC accelerator for Intel Math Kernel Library routines, but did not see any speedup. Distributed-memory calculation with MPI is therefore the preferred way.

We benchmarked the code in both strong and weak scalability tests (Figures 1 and 2). In the largest model we simulated 513 particles on a lattice with 513 sites. The results show that strong scaling is linear with a coefficient of ½. With 256 MPI processes we obtained about 128x speedup. For weak scaling we measured the performance with respect to the increasing system size and increasing number of particles. Again, we tested lattices sizes of up to 512 sites. We found that the weak scaling is constant with the system size. We also found that increasing the number of particles does not inhibit the parallel speedup.

To test the program we recalculated several published state-of-the-art results. We found excellent agreement with experimental and numerical results. Besides the actual simulation program, we also developed a tensor library, which can be used to model various problems. It supports sparse tensors of arbitrary rank

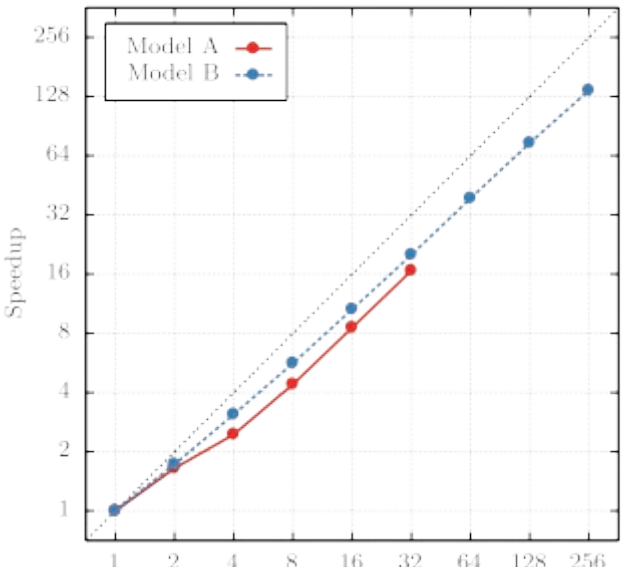


Figure 1.
Strong scaling benchmark for two models. Model A has 65 lattice sites and 65 particles, and model B has 513 lattice sites and 513 particles.

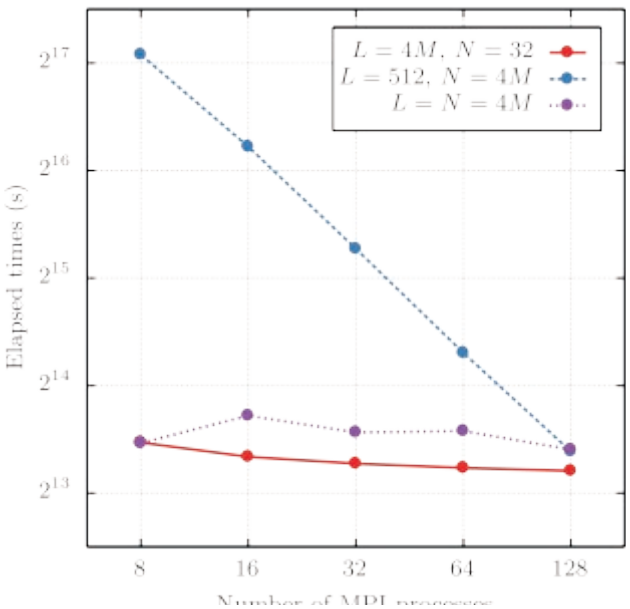


Figure 2.
Weak scaling for three definitions of a problem size. The number of lattice sizes L, the total number of particles N, and both parameters at the same time were increased with the number of MPI processes M. The results show approximately constant weak scaling with the number of lattice sites.

and several tensor operations. The operations include tensor contractions, tensor decompositions, and manipulations with tensor indices. It is also possible to transfer full tensors between compute nodes using MPI.

On-going Research / Outlook

There are many interesting one-dimensional problems that can be simulated with our code. One problem we studied in detail was the Kibble-Zurek mechanism in optical lattices. The Kibble-Zurek mechanism describes the formation of topological defects as the result of the non-equilibrium dynamics of a system that is driven through a continuous phase transition. This phenomenon can take place in one-dimensional systems and is accessible to the program.

Our code focuses on the Bose-Hubbard model which describes the behaviour of ultracold atoms in optical lattices. It is straightforward to extend it to other models, for example to the Heisenberg or Ising model, which describe the physics of spin systems.

Meanwhile, we have shifted our focus from one-dimensional systems to multi-dimensional systems, in particular to two-dimensional systems. Two-dimensional systems are interesting from several perspectives. Many physical phenomena are believed to originate from constraints that make a system effectively two-dimensional. An important example is high-temperature superconductivity. Materials exhibiting high-temperature superconductivity are composed of two-dimensional layers limiting the movement of electrons to these layers only. Similar models can be engineered in two-dimensional optical lattices. Experiments with atoms in two-dimensional optical lattices are very common, because the lattices sites can be addressed individually and measured easily. Two-dimensional models are hard to solve analytically. It is necessary to utilise numerical methods or to extract their properties from experimental data. Our experience with simulating one-dimensional models helps in this research.

The developed code is freely available for download and distributed under an open-source license. It will contain codes for both one-dimensional and two-dimensional simulations.

Conclusion

In the course of this project we developed TEBDOL, a parallel implementation of the TEBD algorithm to simulate the time evolution of ultracold atoms in optical lattices. Developing and running our code on Salomon gave us an opportunity to rapidly optimise the code, to benchmark its performance, and to obtain results quickly. Strong and weak scaling results show that it scales well and is ready for production use.

References

[1] White, Steven R. Density matrix formulation for quantum renormalisation groups. Physical Review Letters. 1992, 69(19), 2863-2866. Doi: 10.1103/PhysRevLett.69.2863.
[2] Vidal, Guifré. Efficient Classical Simulation of Slightly Entangled Quantum Computations. Physical Review Letters. 2003, 91(14). Doi: 10.1103/PhysRevLett.91.147902.
[3] Schollwöck, U. The density-matrix renormalisation group in the age of matrix product states. Annals of Physics. 2011, 326(1), 96-192. Doi: 10.1016/j.aop.2010.09.012.
[4] Dziarmaga, J. Dynamics of a quantum phase transition and relaxation to a steady state. Advances in Physics. 2010, 59(6), 1063-1189. Doi: 10.1080/00018732.2010.514702.

Publications

[1] Urbanek, Miroslav and Soldán, Pavel. Quantum physics simulations in Common Lisp. Conference, 8th European Lisp Symposium, London, United Kingdom, 2015.
[2] Urbanek, Miroslav and Soldán, Pavel. Large-scale simulations with tensor networks. Conference, Quantum Technologies Conference VI, Warsaw, Poland, 2015.
[3] Urbanek, Miroslav and Soldán, Pavel. Parallel algorithms with tensor networks. Conference, Non-Equilibrium Quantum Dynamics in Low Dimensions Mini-Conference, Durham, United Kingdom, 2015.
[4] Urbanek, Miroslav and Soldán, Pavel. Parallel implementation of the time-evolving block decimation algorithm for the Bose-Hubbard model. Computer Physics Communications. 2016, 199, 170-177. ISSN: 0010-4655.

ONE-DIMENSIONAL AND TWO-DIMENSIONAL MAGNETIC SYSTEMS

Research institution:
IT4Innovations National Supercomputing Center

Principal investigator:
Dominik Legut

Researchers:
Rudolf Sýkora, Urszula D. Wdowik

Project partner:
Pedagogical University of Cracow

Project ID:
IT4I-4-1

Introduction

Low dimensional magnetic systems, e.g. one-dimensional magnetic chains are prototypes of potential spintronic devices, magnetic superconductors, and materials with large magnetic heat transport. One of the challenges for the experimentalist is how to identify the one-dimensional vs. two-dimensional magnetic ordering in often magnetically frustrated systems. Using quantum mechanical (ab initio) calculations one can identify the size of magnetic interactions and hence the corresponding magnetic order. $Cu(en)(H_2O)_2SO_4$, $en=C_2H_8N_2$ (Cuen) was found to be magnetically one-dimensional. Here we would like to investigate the optical properties as well as the elastic and magnetic exchange-interactions under pressure, where preliminary measurements show that the quasi one-dimensional character is lost. Therefore, we do search for models of magnetic ordering based on the electronic structure calculations and its correspondence to measured physical quantities at low temperatures that could give us decisive information about the low-dimensional character of magnetically frustrated solids.

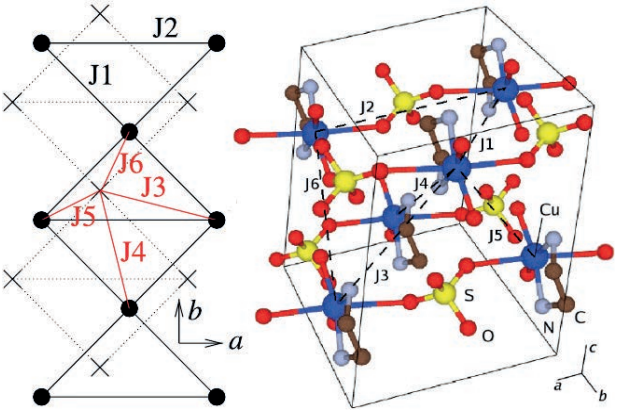


Figure 1.
The schematic of the exchange interactions (left) and the crystal structure of Cuen with them

Results and Methods

Investigation of materials that have anisotropic magnetic interactions in one or two dimensions are of high interest because of the material design for/home/legut/Public/vasp5.4.1-NEB-170 spintronics, for magnetic superconductors and materials with large magnetic heat transport [1-3]. However, for some of the experimentally investigated magnetically frustrated systems, like $Cu(en)(H_2O)_2SO_4$, $en=C_2H_8N_2$(Cuen), see Figure 1, the actual magnetic order is not understood. Recently, the equilibrium structure and atomic positions were determined [4], where positions of the H-atoms were previously not reported (not accessible from X-ray studies).

Based on our first-principle calculations employing density functional theory [5], we determined magnitudes of the exchange-interactions (J's) and hence predict that the Cuen is a one-dimensional antiferromagnet, with the strongest exchange-interaction within the bc-plane, in contrast with what was previously thought [6], see Figure 1. We calculated the J's for the optimal structure as well as the change of the magnitudes of the J's with pressure and surprisingly we found large changes of the crystal and corresponding change of J's along the covalent bonds [7]. We also calculated all dielectric functions tensor elements to understand optical and magneto-optical properties of Cuen with a monoclinic structure in collaboration of the experimentally recorded data at VŠB - Technical University of Ostrava on various Cuen samples. The calculated reflection coefficients and their orientation with respect to the crystal monoclinic symmetry agree very well with those measured [8,9] ones. However, only in the static limit of the dielectric tensor. The study proves that many-body effects alter the optical properties to a large extent and the single electron framework is not satisfactory, i.e. large simulations for the optical properties are required. Further, we calculated the elastic properties and magnetic exchange interactions under pressure as it seems that the one-dimensional character is changed towards a two-dimensional one under pressure as indicated by the preliminary measurements of our foreign collaborators, UPJŠ, Košice, Slovakia.

For all studies we have employed a VASP code [10] for the optimisation of the geometry of the lattice vectors and precise atomic positions need to be determined for those low symmetry structures. The advantage of this technique comes from the evaluation of the stress tensor and hence allowing us to optimise simultaneously many degrees of freedom. The code is well MPI parallelised utilising LAPACK, BLAS, SCALAPACK and BLACS libraries optimised by vendor compiler regarding CPU types. For the best performance the MPI version of FFTW3 is needed.

On-going Research / Outlook

In future we plan to simulate low dimensional magnets under the influence of an external magnetic field. Another task it to perform the lattice dynamics calculations. Together it will allow us to obtain the thermal conductivity as well as the possibility to model the magnetic behaviour for finite temperatures and to compare our results with the experimental records with newly established collaborations with the group at UPJŠ, Košice, Slovakia.

Conclusion

The structural parameters and the equilibrium positions of the H-atoms were determined for the $Cu(H_2O)_2(en)SO_4$ (Cuen) [4]. Next, we have to identify the strongest magnetic exchange interactions by means of first-principle calculations of $Cu(H_2O)_2(en)SO_4$ [4,5]. We confirm the magnetic ordering in this low-dimensional antiferromagnetic insulator is one-dimensional rather than two-dimensional as supposed previously in Ref. [6]. The magnetic moments on the Cu atoms point nearly along the crystal a axis and their magnitude stays essentially intact upon exposure of the system to an external pressure. At the highest considered pressures, the values of the two strongest couplings get closer, changing the system's magnetic character from quasi-one-dimensional to quasi-two-dimensional [5].

References

[1] Žutić, I.; J. Fabian and S. Das Sarma. Spintronics: Fundamentals and applications. Rev. Mod. Physics. 2004, 76, 323. Doi: 10.1103/RevModPhys.76.323

[2] Sirenko, V. A.; Magnetostrictive effects in superconductors, Superlattices and Microstructures. 1998, 23, 1155. Doi: 10.1006/spmi.1996.0436

[3] Sologubenko, A. V.; H. R. Ott. In Strong Interactions in Low Dimensions, Eds. D. Baeriswyl, L. Degiorgi, Kluwer Academic Publishers, Dordrecht 2004, p. 383. ISBN 978-1-4020-3463-3

[4] Sýkora, R.; D. Legut and U. D. Wdowik. Exchange Interactions in a Low-Dimensional Magnetic System $Cu(H_2O)_2(en)SO_4$, Acta. Phys. Pol. A. 2014, 126, 50. Doi: 10.12693/APhysPolA.126.50

[5] Hohenberg, P. and W. Kohn. Phys. Rev. 1964, 136, B864. Doi: 10.1103/PhysRev.136.B864

[6] Kajňáková M., et al., Phys. Rev. B. 2005, 71, 014435. Doi: 10.1103/PhysRevB.71.014435

[7] Sýkora R. and D. Legut, Magnetic Interactions in a Quasi-One-Dimensional Antiferromagnet $Cu(H_2O)_2(en)SO_4$, J. Appl. Phys. 2014, 115, 17B305. Doi: 10.1063/1.4865323

[8] Sýkora R.; K. Postava, D. Legut and R. Tarasenko: Optical properties of a monoclinic insulator $Cu(H_2O)_2(en)SO_4$, Acta Phys. Pol. A. 2015, 127, 469. Doi: 10.12693/APhysPolA.127.469

[9] Sýkora R.; K. Postava, D. Legut and R. Tarasenko, Calculated Reflection Coefficients of a Single Planar Interface with an Optically Biaxial $Cu(en)(H_2O)_2SO_4$ – Material Compared to Experiment, Journal of Nanoscience and Nanotechnology. 2016, 16(8), 7818-7921. Doi: 10.1166/jnn.2016.12560

[10] Kresse G. and J. Furthmuelluerr, Phys. Rev. B. 1996, 54, 11169. Doi: 10.1103/PhysRevB.54.11169

Publications

[1] Sýkora, R.; D. Legut and U. D. Wdowik. Exchange Interactions in a Low-Dimensional Magnetic System $Cu(H_2O)_2(en)SO_4$. Acta. Phys. Pol. A. 2014, 126, 50. Doi: 10.12693/APhysPolA.126.50

[2] Sýkora, R. and D. Legut. Magnetic Interactions in a Quasi-One-Dimensional Antiferromagnet $Cu(H_2O)_2(en)SO_4$. J. Appl. Phys. 2014, 115, 17B305. Doi: 10.1063/1.4865323

[3] Sýkora, R.; K. Postava, D. Legut and R. Tarasenko. Optical properties of a monoclinic insulator $Cu(H_2O)_2(en)SO_4$. Acta Phys. Pol. A. 2015, 127, 469. Doi: 10.12693/APhysPolA.127.469

[4] Sýkora, R.; K. Postava, D. Legut and R. Tarasenko. Calculated Reflection Coefficients of a Single Planar Interface with an Optically Biaxial $Cu(en)(H_2O)_2SO_4$ – Material Compared to Experiment. Journal of Nanoscience and Nanotechnology. 2016, 16(8), 7818-7821. Doi: 10.1166/jnn.2016.12560

COMPREHENSIVE MODELLING LONG TIME EVOLUTION OF SUDDENLY IONISED RARE GAS CLUSTERS THROUGH MULTISCALE MODEL AND FULL DYNAMICS CALCULATIONS

Research institutions:
Institute of Geonics of the Czech Academy of Sciences, Institute of Clean Technologies for Mining and Utilisation of Raw Materials for Energy Use in Ostrava

Principal investigator:
Ivan Janeček

Researchers:
René Kalus, Martin Stachoň

Project partners:
IT4Innovations National Supercomputing Center, Metacentrum – National Grid Infrastructure of the Czech Republic

Project IDs:
IT4I-1-17, OPEN-4-5, OPEN-5-18, OPEN-7-18

Introduction

The series of projects was focused on numerical modelling of the long time evolution of ionised rare gas clusters, and includes new results, which directly extend the huge calculations performed in the PRACE project through computational sources of the Hermit supercomputer in Stuttgart. The projects are particularly devoted to reliable description of microscopic processes during post-ionisation fragmentation of rare gas clusters of all sizes observed in molecular beam experiments. However, tested physical and numerical approaches, which were developed originally for the special problem, can be generalised for other molecular systems evolving through metastable states. Computational time obtained through the IT4Innovations projects was consumed, in addition to testing new methodologies, for treatment of PRACE data (particularly IT4I-1-17), and later (the continual series of OPEN Call projects) for huge full-dynamical calculations extending the short-time evolution over one hundred nanoseconds related to the time of finished fragmentation.

In the projects special simulation software MULTIDYN (prepared by our group), the use of parallel computations based on approaches of quantum mechanics (especially modified hemiqantal dynamics) was developed and tested on model systems represented by ionised rare gas clusters. In the framework of the mentioned projects, the reliability of used approaches was verified on the experimental data from cluster fragmentation after sudden ionisation through accelerated-electron impact to a size selected neutral parent cluster.

Basic research improving the precision of the dynamical simulation of molecular and particle systems is the dominant content of the comprehensive study. However, the proposed approaches can be used, e.g., for calculations of physical inputs necessary for modelling cold rare gas plasma, which can be applied for tissue sterilisation in medicine.

Many complicated molecules, especially organic, and their interaction and time evolution in chemical processes are important for the preparation of new materials, drugs, or for understanding many natural phenomena or processes, etc. Molecular systems can be successfully simulated through numerical models, which must reflect the quantum nature of building particles, nuclei and especially electrons. Unfortunately, the full quantum calculations are not possible, as a consequence of mathematical difficulties, or at least needs an extremely long run-time to model sufficiently long real times of system evolution.

Here are several paths to overcome this problem and to perform the modelling quicker and more efficiently. The first one lies in appropriate simplification of dynamical calculations. Very popular is an attempt called Molecular Dynamics, which is commonly adopted as the terminus for dynamical calculations utilising classical equations of motion, i.e. those non-quantal based on Newtonian principles, for description of nuclei, which interact through an effective molecular force potential obtained preferably by quantal treatment of electron system for static nuclei. Moreover, electrons in their ground state are the most frequently

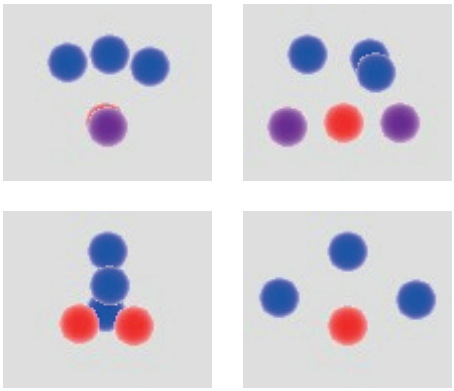


Figure 1.
Equilibrium configurations of Kr_6^+ cluster for ground (top two pictures) and selected excited (bottom two pictures) electronic states, the red color represents positive hole localized at "charged core" of the cluster.

considered. Another approach, which includes electronic excitation during the system evolution, calculates the force potential "on the fly" for current nuclei configuration. Such an attempt is typically named as Hemiquantal Dynamics. However, calculations through this approach can still be very time consuming. Therefore, besides further simplification of equations of motion or utilisation of alternative statistical approaches, a further way to attain the necessary final simulated time during reasonable simulating time-intervals can be considered, specifically technical improvement of dynamical calculation efficiency. The improvement can be achieved through parallel calculations, i.e. many simultaneous calculations performed on many computer cores of supercomputer clusters. However, the necessary condition for such acceleration of the calculations is splitting the code of the simulating program to independent simultaneous branches. The conditions are filled especially by the statistical tasks based on evolution of non-interacting atomic systems, molecules or atomic and molecular clusters.

Rare gas clusters, owing to their relative simplicity, can play the role of a useful "testing laboratory" for verification of special approaches. While neutral clusters of rare gas atoms only survive at very low temperatures, because the atomic complexes are only bound through weak residual forces, the ionised rare gas cluster has binding energy comparable with other strong chemical bonds, and so after conversion of kinetic energy excess can survive up to times from microseconds (metastable) to infinity (stable). Hemiquantal calculation of the necessary thousands of trajectories for clusters of sizes up to about ten atoms, realised on a single core, can in detail involve only processes up to several hundreds of a picosecond. Moreover, the total simulation time can still vary from several days to several months in dependence on cluster size. Owing to parallel computation, the runtimes can be significantly lowered, and the final time of simulated evolution can be prolonged up to microseconds, which are approximately the time of decay of metastable rare gas clusters in considered beam experiments.

Results and Methods

For dynamical computation we have used the Mean Field with Quenching (MFQ) method (see [1-3] and reference herein), which represents an extension of the original Ehrenfest MF approach [4]. The implementation of the method is prepared for the specific case of rare-gas cluster cations. The main idea of the improved methodology consists in periodic inclusion of quantum decoherence, missing in the original MF approach. The dynamical calculations are the necessary input for multiscale models (its basis was constructed in [2] and [3]), which represent combination of hemiquantal calculations and kinetic treatment. Details related to the implemented interaction model based on the diatomic in molecules approach [5-6] can be found in refs. [1-3]. The advantages of the used hemiquantal model lie in inclusion of electronic transitions and quantum decoherence. These effects are usually omitted in standard molecular dynamics (evolution on ground state only and classical equation of motion).

In a preceding study we have used simulations based on the method combining parallelised short-time hemiquantal dynamics with the kinetic model for long-time scales. While extensive short-time dynamics was performed for clusters Ar, Kr and Xe up to size 13 (this includes all experimental sizes) due to the PRACE project, the additional calculations, primarily including data treatment through the kinetic model, were performed in the framework of initial IT4Innovations projects and with the support of computational and storage sources of Metacentrum.

Results of the short time dynamics and their analysis through a multiscale approach are presented in prepared work [8]. For clusters of krypton and xenon the multiscale simulations predict the fragmentation with dominant production of charged monomers followed by charged dimers. Predicted charged-fragment fractions were in good agreement with the experiment in the case of small clusters, while for large clusters the obtained predictions are only qualitative. So, an additional calculation based on both diabatic and adiabatic model of initial electronic excitation was realised for a selected instance – cluster Kr_7. Putting the new Salomon computer cluster into operation allowed us to even obtain full dynamics data, which are compatible with experimental findings, see Figure 2 for illustration.

On-going Research / Outlook

In the next period we will focus on detailed analysis of long-time data obtained for cases of large as well as small sizes of krypton clusters. Particular attention will be devoted to the role of initial

electronic excitation after electron impact, because especially for heavy rare gas clusters Kr_N and Xe_N the spin-orbit gap in their electronic states plays a key role in the formation of metastable fragments. Calculations in the pending OPEN-7-18 project indicate that Xe clusters also show behaviour very similar to the Kr case. Results of this new analysis of long-time dynamical data will be published in a subsequent paper directly extending the findings presented in work [8]. Later, the analysis needs to be completed by a theoretical study of argon clusters, the behaviour of which is different (more charged dimers are observed) than for Kr_N and Xe_N, probably as a consequence of weaker spin-orbit splitting.

Conclusion

The theoretical modelling of fragmentation dynamics of ionised clusters certainly represents a special problem of the given physical system. However, there are several reasons why these studies are important for the advancement of chemical physics: the MFQ dynamics include electronic degrees of freedom to the simulations in contrast to popular MD considering only the electronic ground state; besides other methods considering the quantum description of electrons, the MFQ include the simple model of quantum decoherence, which is naturally presented only in full quantum dynamics of electrons and nuclei; the MFQ dynamics is a tool for modelling elementary scattering processes, which are important for description of cold rare gas plasma having potential for medical applications [7].

Multiscale modelling, which combines short-time hemiquantal dynamics with the kinetic approach represents an effective method for overcoming time limits in simulations of large molecules. In addition, permanent progressive growth of computational resources (now, mainly those of IT4Innovations) enabled a more detailed and more reliable study of rare gas clusters on scales of hundreds of nanoseconds, even through full Hemiquantal Non-adibatic Dynamics, taking into account the evolution of electronic states.

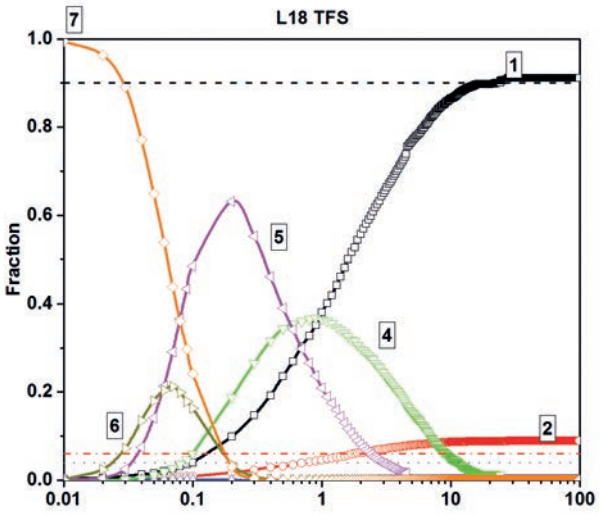


Figure 2.
Time evolution of fragment fractions obtained by TFS(1,1)/S hemiquantal model for Kr_7 ionised into the initial adiabatic level L18 (full lines and open points, number in curve labels are equivalent to fragment sizes). Horizontal lines represent known experimental values for charged monomers (black dashed), dimers (red dash-dot), and trimers (blue doted).

References / Publications

[1] Janeček I.; S. Cintavá, D. Hrivňák, R. Kalus, M. Fárník, and F. X. Gadea. J. Chem. Phys. 2009, 131, 114306. Doi: 10.1063/1.3224855
[2] Janeček I.; T. Janča, P. Naar, F. Renard, R. Kalus, and F. X. Gadea. EPL. 2012, 98, 33001. Doi: 10.1209/0295-5075/98/33001
[3] Janeček I.; T. Janča, P. Naar, R. Kalus, and F. X. Gadea, J. Chem. Phys. 2013, 138, 044303-. Doi: 10.1063/1.4775804
[4] Ehrenfest P. Z. Phys. 1927, 45, 455-. Doi: 10.1007/BF01329203
[5] Ellison, F. O. J. Am. Chem. Soc. 1963, 85, 3540-. Doi: 10.1021/ja00905a002
[6] Kuntz P. J. and J. Valldorf. Z. Phys. D. 1988, 8, 195-. Doi:10.1007/BF01384514
[7] Chicheportiche A.; M. Stachoň, M. Benhenni, F. X, Gadéa, R. Kalus, and M. Yousfi. J. Chem. Phys. 2014, 141, 134302-. Doi: 10.1063/1.4896613
[8] Janeček I.; P. Naar, M. Stachoň, R. Kalus, and F. X. Gadea, Submitted to Phys. Chem. Chem. Phys

ISBN 978-80-248-4037-6
www.it4i.cz